



ICQGG·6

Abstracts

2020

Abstract book sponsored by



ONE FAMILY.
ONE PURPOSE.

Talks	Page
Session 1	2
Session 2	3
Session 3	6
Session 4	9
Session 5	11
Session 6	15
Session 7	20
Session 8	23
Session 9	27
Session 10	33
Session 11	39
Session 12	43
Session 13	49
Session 14	53
Session 15	54
Session 16	58
Session 17	61
Session 18	64
Session 19	70
Session 20	76
Session 21	80
Session 22	85
Session 23	88
Session 24	91
Session 25	94
Poster Presentations.....	97-279

Quantitative Genetics Isn't Dead Yet

Professor Peter M. Visscher¹

¹*Institute for Molecular Bioscience, University of Queensland, ST LUCIA, Australia*

Session 1, November 3, 2020, 7:00 AM - 8:30 AM

Biography:

Peter Visscher FRS is a quantitative geneticist with research interests focussed on a better understanding of genetic variation for complex traits in human populations, including quantitative traits and disease, and on systems genomics. The first half of his research career to date was predominantly in livestock genetics (animal breeding is applied quantitative genetics), whereas the last 15 years he has contributed to methods, software and applications to quantify and dissect genetic variation in human traits.

The discipline of quantitative genetics (QG) was established more than a century ago, and most agree that RA Fisher's 1918 publication, which reconciled Mendelian genetics with the resemblance between relatives for quantitative traits, is its foundation paper. It was mostly theoretical for the first half of the 20th century but became much more empirical in the latter half, with empirical tests of theory using selection experiments and applications in plant and animal breeding. As eloquently summarised by Bill Hill in a 2010 Royal Society publication, QG as a discipline has been written off many times in the last 40 years, in particular when the molecular biology revolution started to flex its muscles in the 1980s. Surely it was just a matter of little time before the few relevant mutations for each trait would be found and quantitative genetics would become a mere sub-branch of molecular genetics? If only. I will show that not only is QG alive and kicking, it has revolutionised many fields of research in the last ~15 years, including plant and animal breeding, human genetics, and evolutionary and ecological genetics. Examples will be given of how the combination of whole-genome genetic data and trait information has led to new insights into (i) the polygenic basis of complex traits (including common disease), (ii) the partitioning of genetic variation between and within families and (iii) the effects of natural selection and non-random mating on trait variation.

The Genetic Architecture of Drosophila Lifespan

Professor Trudy Mackay¹

¹Center for Human Genetics and Self Family Endowed Chair, Clemson University, Greenwood, South Carolina, USA

Session 2, November 3, 2020, 12:00 PM - 1:25 PM

Biography:

Trudy Mackay is the Director of the Center for Human Genetics, the Self Family Endowed Chair of Human Genetics and Professor of Genetics and Biochemistry at Clemson University. Her laboratory focuses on understanding the genetic and environmental factors affecting variation in quantitative traits, using Drosophila as a translational model system. Her laboratory seeks to identify the genetic loci at which segregating and mutational variation occurs, allelic effects and environmental sensitivities, and the causal molecular variants. Her research utilizes mutagenesis to identify candidate genes and pathways, quantitative trait locus mapping of alleles segregating in nature, systems genetics analyses to provide biological context and identify transcriptional and genetic networks affecting complex traits; and germline gene editing to prove causal associations. She is a Fellow of the American Academy of Arts and Sciences and the Royal Society, a member of the US National Academy of Sciences, the 2016 Wolf Prize Laureate for Agriculture and the 2018 Dawson Prize recipient, Trinity College, Dublin.

The world population is rapidly growing older, and population aging will be one of the most important social and health problems in the coming half-century. Lifespan is a typical quantitative trait, with natural variation attributable to segregating variants at multiple interacting loci, the effects of which are sensitive to the environment to which the individuals are exposed. However, only a handful of candidate genes associated with natural variation in lifespan in human populations have been identified. *Drosophila melanogaster* is a powerful genetic model system for identifying evolutionarily conserved genes and genetic networks causally associated with variation in lifespan, due to the ability to accurately measure lifespan while precisely controlling both genetic background and environmental conditions, and publicly available genetic resources. I will describe the properties of allelic variants affecting *D. melanogaster* lifespan from analyses of new mutations; as well as naturally occurring variants in the Drosophila Genetic Reference Panel (DGRP), a population of 205 sequenced inbred strains; a large outbred, advanced intercross population derived from a subset of DGRP lines; and a classic laboratory evolution experiment. Lifespan is highly polygenic, and naturally segregating alleles have effects that are largely context-dependent, varying according to sex, environment, and genetic background. These context-dependent effects may explain why variation for lifespan is maintained in natural populations. Further, many of the implicated genes are evolutionarily conserved and have human orthologs, facilitating direct tests for effects on lifespan in human populations.

Estimation of realized rates of genetic gain for breeding program assessment: Insights from rice research at IRRI

Dr Jessica Rutkoski¹

¹University of Illinois, Urbana, United States

Session 2, November 3, 2020, 12:00 PM - 1:25 PM

Biography:

Jessica received a Bachelor of Science degree in genetics from the University of Wisconsin-Madison in 2009. During her time at Wisconsin she worked for Professor Bill Tracy's sweetcorn breeding program where she developed a love for plant breeding. Shortly after graduation, Jessica began her PhD work at Cornell University under the direction of wheat breeder and plant breeding professor Mark Sorrells. Jessica's PhD research focused on genomic selection for quantitative disease resistance in wheat, and it included one of the first empirical genomic selection experiments in plants. After receiving her PhD in 2014, Jessica took a position as an assistant professor at Cornell and an adjunct associate scientist at CIMMYT working on integrating genomic selection and high-throughput phenotyping to predict breeding values for yield in wheat. Jessica is now leading the quantitative genetics cluster at the international rice research institute (IRRI) where her research currently focuses on improving rice breeding efficiency and monitoring breeding program effectiveness.

Advancements in quantitative genetics have promised to revolutionize plant breeding, but most plant breeding programs continue to use traditional methods. More intensive monitoring and evaluation of breeding programs, especially those at CGIAR centers, based on realized rates of genetic gain (ΔGt) has been suggested to promote the adoption of improved breeding methods. However, methods to estimate realized ΔGt from plant breeding datasets have not been systematically evaluated. To develop recommendations for routine estimation of ΔGt in plant breeding, I conducted a stochastic simulation study of 80 rice breeding programs over 28 years and used the simulated data to compare five different methods for estimating realized ΔGt . At best, estimates of realized ΔGt were under or overestimated by 15% when using all 28 years of data and by 27% when using the most recent 15 years of data. On average, the best methods were the control population, estimated breeding value, and era trial methods. All methods led to estimates that were biased, and the direction of this bias was associated with the breeding program that was simulated. I conclude that estimates of realized ΔGt can be accurate in some cases, but they are not useful for comparing the effectiveness of different breeding programs. In light of these challenges, I will discuss other means of monitoring breeding program performance which I used at the International Rice Research Institute (IRRI), and I will share my perspective on how the quantitative plant breeding revolution will unfold.

The role of cross-sex genetic covariances in the evolution of sexual dimorphism among species

Dr Jacqueline Sztepanacz¹

¹University of Toronto, Canada

Session 2, November 3, 2020, 12:00 PM - 1:25 PM

Biography:

Dr. Jacqueline Sztepanacz is an Assistant Professor in the Department of Ecology and Evolutionary Biology at the University of Toronto. Her research focuses on the evolution of genetic variation, with a particular interest in how traits evolve (or don't) under stabilizing selection. Prior to her appointment as assistant professor, Dr. Sztepanacz was a postdoctoral researcher at Florida State University, and she completed her PhD at the University of Queensland. Dr. Sztepanacz is the proud recipient of the 2020 Dobzhansky prize from the SSE, and her research has been published in Genetics, Evolution, and The American Naturalist, among others.

The independent evolution of males and females is potentially constrained by their shared genome that generates genetic correlations between the sexes. These cross-sex genetic covariances determine whether and to what extent sexual dimorphism can evolve to resolve sexual conflict. *Drosophila* wing-shape has emerged as a model high-dimensional complex trait; wing-shape is consistently highly evolvable in contemporary populations, and yet perplexingly stable across phylogenetic timescales. The potential role of cross-sex covariances in dictating patterns of evolutionary stasis has received limited attention. Here we study the role of cross-sex covariances in a comparative context. We characterized sexual dimorphism in wing-shape in 83 species of *Drosophila* and estimated a multivariate cross-sex genetic covariance matrix in *D. melanogaster*. We show that sexual dimorphism in wing-shape is common, that it varies among species, and that it has a low phylogenetic heritability. These data all suggest that sexual dimorphism evolves rapidly. However, when we incorporate cross-sex genetic covariances in evolutionary predictions of sexual dimorphism, we find that its divergence both in contemporary populations and across the *Drosophila* phylogeny is constrained by over 90%. Therefore, cross-sex covariances may indeed contribute to evolutionary stasis across phylogenetic timescales. Our results highlight the importance of considering between-sex genetic covariances when making predictions about evolution on both macro- and micro-evolutionary timescales and may provide one more piece to the puzzle of stasis.

Crowdsourcing genetic data. A look from the outside and inside.

Dr Yaniv Erlich¹

¹My Heritage, Israel

Session 3, November 3, 2020, 7:00 PM - 8:25 PM

Biography:

Dr. Yaniv Erlich is the Chief Science Officer of MyHeritage.com and an Associate Professor of Computer Science and Computational Biology at Columbia University (leave of absence). Prior to these positions, he was a Fellow at the Whitehead Institute, MIT. Dr. Erlich received his bachelor's degree from Tel-Aviv University, Israel (2006) and a PhD from the Watson School of Biological Sciences at Cold Spring Harbor Laboratory (2010). Dr. Erlich's research interests are computational human genetics. Dr. Erlich is a TEDMED speaker (2018), the recipient of DARPA's Young Faculty Award (2017), the Burroughs Wellcome Career Award (2013), Harold M. Weintraub award (2010), the IEEE/ACM-CS HPC award (2008), and he was selected as one of 2010 Tomorrow's PIs team of Genome Technology. He is currently working on statistical genetics

Precision medicine is a data-hungry endeavor. However, traditional cohort ascertainment strategies poorly scale and necessitate substantial investments to obtain genomics data, conduct physical exams and lab tests, and assess familial history. But are these really required in today's world? In the last decade, the human population has produced zettabytes (10^{21}) of digital data. Here, I will present our successes in repurposing participants' data for ultra-large scale genetic studies. First, I will describe our long-term project to build a 13-million family tree by a mining genealogy-driven social. Second, I will present DNA.Land, our website to crowd-source genetic data of Direct-To-Consumer participants. Third, I will describe MyHeritage Health, a novel product to empower participants to learn about their genetic predispositions and the challenges in building this product. Last, I will talk about genetic privacy implications of this brave new world.

Adaptation and quantitative trait evolution in an urban context

Professor Anne Charmantier¹, Dr Charles Perrier², Aude Caizergues¹, Dr Arnaud Grégoire³

¹CNRS, Montpellier, France, ²INRAE, Montpellier, France, ³University of Montpellier, Montpellier, France

Session 3, November 3, 2020, 7:00 PM - 8:25 PM

Biography:

Anne Charmantier is an evolutionary ecologist interested in between-individual variation of life-history, morphological, and behavioural traits in natural populations. She holds a senior CNRS position (eq. Prof) in the Centre d'Ecologie Fonctionnelle et Evolutive (CNRS, UMR 5175, Montpellier, France). Her main research interests are focused on understanding the mechanisms involved in the evolution of adaptive traits, especially in a context of climate change and urbanisation. Since 2007 she has been managing a blue tit/great tit project dating back from 1976, which offers unique opportunity to study adaptation in heterogeneous and rapidly changing environments. In 2014, she published the OUP book 'Quantitative Genetics in the Wild' with Profs. Dany Garant and Loeske Kruuk.

With its conspicuously altered ecological dynamics, the urban environment stands in stark contrast to the natural environment that has been used as research ground for virtually all long-term studies used as cornerstones in evolutionary ecology. Because of this ecological contrast, it is often assumed that new phenotypes found in cities are the result of an evolutionary response to novel selection shaping genomic variation. In this talk, I will argue that urban environments offer exciting perspectives to investigate processes of rapid adaptation, using combinations of approaches, in particular quantitative genetics and genomics. Results from a project on great tits *Parus major* along an urbanisation gradient will illustrate the potential of studies on urban local adaptation. This study revealed phenotypic divergence for a large set of avian morphological, behavioural and life history traits: e.g. birds in the city are smaller, more aggressive, and breed earlier and smaller clutches. First, the genetic architecture and evolutionary potential of these traits was explored using animal models based on long-term field pedigrees and genomewide relatedness matrices. Second, comparing natural selection in forest versus urban habitats failed to support the hypothesis that contemporary selection explains phenotypic differences between urban- and forest-breeding great tits. Third, (epi)genomic investigations using RAD- and RRBS- sequencing revealed an effect of urbanisation on (epi)genetic diversity with numerous (epi)genetic regions showing elevated divergence, suggestive of a polygenic adaptation. These findings open exciting perspectives for broader investigations of genomic bases related to adaptation in urban environments, notably in relation to avian personality and metabolism.

PolyOrigin: Haplotype Reconstruction in Tetraploid Multiparent Populations

Dr Chaozhi Zheng¹

¹Wageningen University, The Netherlands

Session 3, November 3, 2020, 7:00 PM - 8:25 PM

Biography:

Chaozhi Zheng is a researcher at the Biometris group at the Wageningen University and Research. Zheng obtained his Ph.D. in statistical ecology from the University of Helsinki in 2009. He was a postdoctoral researcher in statistical genetics at the University of Washington, Seattle from 2010 to 2012. He nurtures a keen interest in developing statistical and computational methods for biological data analysis. At the Biometris group since 2013, he has been working on the methodological development for genetic analysis in diploid and polyploid multiparental populations.

Many diploid multiparent populations have been developed to increase genetic diversity and quantitative trait loci (QTL) mapping resolution. Haplotype reconstruction for these populations has been used to increase QTL detection power compared to genetic analysis with the marker alleles per se. To realize similar benefits in autotetraploids (and eventually higher ploidies), a statistical framework for haplotype reconstruction in autotetraploid multiparent populations has been developed and implemented in the software PolyOrigin. Haplotype reconstruction proceeds in two steps: first, parental genotypes are phased; second, genotype probabilities for the parental alleles are inferred in the progeny. PolyOrigin can utilize genetic marker data from SNP arrays or from sequence-based genotyping; in the latter case, bi-allelic read counts can be used (and are preferred) as input data to minimize the influence of genotype call errors at low depth. To account for errors in the input map, PolyOrigin includes functionality for filtering markers, inferring inter-marker distances, and refining local marker ordering. Simulation studies were used to investigate the effect of several variables—including the mating design, number of parents, population size, and sequencing depth—on the accuracy of haplotype reconstruction. We further evaluated PolyOrigin using an autotetraploid potato dataset with a 3x3 half-diallel mating design.

Impact of omics on the genetic improvement of Zebu cattle in the tropics

Professor Lucia Albuquerque¹

¹Sao Paulo State University, Brazil

Session 4, November 4, 2020, 7:00 AM - 8:25 AM

Biography:

Lucia Galvão de Albuquerque is Professor of Animal Breeding and Genetics at the Department of Animal Science at São Paulo State University and a researcher of the National Council for Scientific and Technological Development (CNPq). She has experience in beef cattle breeding, working closely with the beef industry in Brazil. Lately, she has been leading research projects, based in complete DNA and RNA sequences, with the objective of developing technologies to increase biologic knowledge and to promote genetic improvement of quality, efficiency and sustainability of Nelore beef cattle production.

Meat production in the tropics is mainly based on Zebu (*Bos indicus*) cattle and their crosses with European (*Bos taurus*) breeds. In general, Zebu breeds are well adapted to warm and humid climate in the tropics and also resistant to a variety of endo and ecto parasites common in these regions. However, improvement of sexual precocity, carcass and meat quality traits as well as feed efficiency are still necessary in Zebu breeds, in order to be more competitive with other activities. These traits present low heritability estimates and/or are expensive and difficult to measure. In this presentation, we are going to show applications of genomics in Zebu cattle in Brazil, more specifically in Nelore breed, focusing mainly in new or difficult to measure traits. In the tropics, animals are subjected to huge environmental differences, even in the same herd and, consequently, genotype by environmental interaction plays a significant role in the expression of complex traits. Studies about genotype by environmental interactions in Zebu cattle, using genomic information, will also be discussed. Moreover, results from GWAS and RNA_Seq studies and from combination of omics data will be presented. Finally, we will show our recent research project focused on enteric methane emission and studies using the whole DNA sequence, aiming to better understand the genetic components affecting the quality, efficiency and sustainability of Nelore beef cattle production. Acknowledgments: São Paulo Research Foundation - FAPESP (#2009/16118-5 and #2017/10630-2).

Predicting the Accuracy of Genomic Predictions

Dr Jack Dekkers¹, Dr Hailin Su¹

¹Iowa State University, Ames, United States

Session 4, November 4, 2020, 7:00 AM - 8:25 AM

Biography:

Jack grew up in the Netherlands and received B.Sc. and M.Sc. degrees from the Wageningen Agricultural University and a Ph.D. from the University of Wisconsin in Animal Breeding. From 1989 to 1997 he was on faculty at the University of Guelph, working closely with the Canadian industry on genetic improvement of dairy cattle. He moved to Iowa State University in 1997, where he currently is a C.F. Curtiss Distinguished Professor and Leader of the Animal Breeding and Genetics group. Current research focuses on the genetic basis of feed efficiency and health in pigs and poultry, and on the integration of quantitative genetics and genomics in breeding programs.

Design of breeding programs using genomic prediction requires methods to model such programs. While deterministic methods are available to model breeding programs with selection on pedigree-based estimates of breeding values (PEBV), such methods have not been fully developed for breeding programs that include genomic prediction, with a key missing component being the accuracy of genomic EBV (GEBV) of selection candidates. Here, a deterministic method was developed for the prediction of this accuracy based on the accuracy of GEBV and PEBV in the reference population and the distance of selection candidates from their closest ancestors in the reference population. Two approaches were used to model the accuracy of GEBV as a combination of the accuracy of PEBV and of EBV that are based on genomic relationships deviated from pedigree, with independent sampling errors; selection index theory and Fisher information theory. Loss of the accuracy of genomic information from the reference to the target population was modeled based on the effective number of loci in the reference population (M_e), which determines the size of independent segments whose effects are estimated in the reference population and the probability that a random segment is broken up by recombination when moving from the reference to the target population. M_e in the reference population was estimated based on accuracies of GEBV and of PEBV in the reference population, using either the Fisher information or the selection index approach. Using simulation, both the Fisher and the selection index approach correctly predicted accuracy in the target population over time, both with and without selection. The Fisher and index approaches, however, resulted in different estimates of M_e , with the index approach resulting in estimates that were less affected by heritability, reference size, and selection, and which are, therefore, more appropriate as a population parameter. A leave-one-out cross-validation approach was proposed to estimate required accuracies of EBV in the reference population. A deterministic method was developed to predict the accuracy of genomic predictions in selection candidates. This method can be used to evaluate the benefit of genomic prediction and optimize genomic selection breeding programs. Funding by USDA-NIFA grant # 2016-10148

Polygenic Risk Scores in Humans – What are they Good For?

Dr Greg Gibson¹

¹*Georgia Tech, Atlanta, United States*

Session 5, November 4, 2020, 12:00 PM - 1:25 PM

With the recognition that polygenic risk scores (PRS) now routinely identify the top few percentile of individuals who have at least 3-fold elevated genetic risk of complex disease has come a push for clinical implementation. However, numerous barriers need to be overcome, such as low repeatability of similar scores for the same trait; heavily biased estimation across mixed ancestries and in admixed individuals; and low positive predictive values. I will report results addressing these issues in a variety of contexts. Regarding cardiovascular disease, reanalysis of the UK Biobank cohort integrating genome-wide PRS and known clinical risk factors with cross-validated logistic regression shows that after the age of 50, the genetic component barely improves prediction since the risk is mostly associated with development of clinical risk that is usually evident in middle age. Regarding inflammatory bowel disease, PRS for both prediction of disease onset and disease progression are outperformed by transcriptional risk scores (TRS) that sum over observed gene expression of GWAS-associated loci. Furthermore, predicted gene expression in the TWAS framework leads to development of a predicted polygenic PRS that provides significant discrimination of progression to colectomy in both the UK Biobank and IBD Genetics Consortium datasets. Finally, I will argue that negative prediction, namely defining that portion of the population unlikely to benefit from medication, has enormous potential in the clinical setting to reduce costs and avoid unnecessary medication. Discriminate application of PRS requires careful consideration of the rationale in each specific situation.

Genome-wide association study identifies tissue-specific regulation of human protein N-glycosylation

Mr Sodbo Sharapov¹

¹*Novosibirsk State University, Russia*

Session 5, November 4, 2020, 12:00 PM - 1:25 PM

Biography:

Sodbo Sharapov is a Ph.D. candidate at the Institute of Cytology and Genetics, Novosibirsk, Russia, and a part-time faculty member at the Department of Cytology and Genetics at the Novosibirsk State University. His main research interest is the application of statistical genetics and multi-omics approaches to study the etiology of human diseases and traits. He is on a way to defend his Ph.D. theses, where he aimed to expand the knowledge about the genetic control of human protein glycosylation.

Glycosylation is a common and structurally diverse posttranslational modification of proteins that influences the physical properties of proteins as well as their biological functions. In recent years, changes in different glycan abundances were associated with various diseases; however, interpretation of observed disease-associated glycan changes is difficult due to limited knowledge about tissue-specific regulation of glycosylation, although its biochemical pathways are known. Here we present the most powered genome-wide association study of the human blood plasma protein N-glycosylation measured by Ultra Performance Liquid Chromatography technology in up to 3811 people. We studied the association between 8.5 million genetic polymorphisms on human autosomes and 113 relative abundances of N-glycan structures. We discovered and replicated twelve associated loci, seven of which were novel. To prioritize potentially causal genes in the found loci, we performed an extensive in silico functional study. Eight loci contained genes encoding enzymes directly involved in glycosylation and a known regulator of plasma protein fucosylation (HNF1A). However, we also found loci that could reflect other, more complex aspects of the glycosylation process. Functional genomic annotation suggested the role of several genes, including SMARCB1 (acts in chromatin remodeling in B-cells), DERL3 (plays a role in the degradation of misfolded glycoproteins), CHCHD10, TMEM121, IGH (encodes immunoglobulin heavy chains), and IKZF1 (regulator of lymphocyte differentiation). Using a network-based multi-omics approach, we demonstrated a clear overlap in genetic control of glycosylation of plasma proteins and immunoglobulin G, leading to future studies that will distinguish global, cell-, tissue-, and protein-specific pathways of protein glycosylation.

Genomic mate selection for clonal crops: improving the chance of breeding top ranking clones by predicted variance in total merit

Dr Marnin Wolfe¹, Dr Jean-luc Jannink^{1,2}

¹Cornell University CALS Section on Plant Breeding And Genetics, Ithaca, United States, ²USDA ARS, Ithaca, United States

Session 5, November 4, 2020, 12:00 PM - 1:25 PM

Biography:

Marnin Wolfe is a Research Associate in Plant Breeding and Genetics at Cornell University and part of an international, multidisciplinary team empowering small-holder farmers through innovative, genomics-enabled cassava breeding. Since 2013, in Jean-Luc Jannink's lab group, Marnin has addressed fundamental questions about the inheritance and prediction of complex traits and leads its application to cassava breeding project-wide. During his PhD at the University of Pittsburgh with Stephen Tonsor, Marnin studied the pervasive phenomenon of local climatic adaptation in plants. Marnin is an integrative ecologist-turned-plant-breeder committed to applying advances in genomics, phenomics and computation to deliver sustainable, climate-adaptation solutions for agroecosystems.

Diverse crops ranging from staples (e.g., cassava) to cash crops (e.g., cacao) are both outbred and clonally propagated. In these crops, exceptional genotypes can be immortalized and commercialized as clonal varieties, which is a distinct advantage relative to many other crops. Genomic prediction can incorporate both additive and non-additive effects in clonal crops to select candidates with high breeding value as parents for crossing and candidates with high total merit as varieties. Improvements over genomic truncation selection have recently been made by predicting genetic variance and selecting crosses instead of parents in plant and animal breeding. In this study, we extend these developments to the breeding of clonal organisms. First we predict both additive and non-additive variances in outbred crosses using genome-wide phased parental haplotypes, marker effects estimates, and recombination maps. We use the predicted distribution of progeny merit to calculate the probability of observing an elite clone from a cross with a given number of matings. We use simulations to assess the accuracy of variance prediction in a scenario modelled after cassava breeding. Finally, we compare our predictions to the realized means, variances and family sizes of >200 real crosses from a cassava genomic selection program and consider where selection and field testing efforts are predicted to have been over or underspent. We discuss strategies for optimizing mate selection and clone testing efforts to serve simultaneous population improvement and variety development in clonally propagated crops.

A mate allocation program for Mass Spawning Species

Dr Cecile Massault¹

¹James Cook University, Australia

Session 5, November 4, 2020, 12:00 PM - 1:25 PM

Biography:

Cecile completed her PhD in animal breeding and genetics in 2010 from the Roslin Institute (Edinburgh, Scotland) and Wageningen University (The Netherlands) looking for QTL for stress, disease and morphology in European seabass and gilthead seabream. Cecile moved on to the University of New England (Armidale, Australia) where her work focussed on the development of phenotyping strategies. She returned to aquaculture in 2018 at James Cook University working on the various aspects of breeding programs for pearl oysters and others important Australian aquaculture species such as barramundi and abalone closely such as pedigree reconstruction, BLUP and genomic selection and GWAS.

While there are several software tools currently available to select individuals which result in limited increase of inbreeding, no solution have been provided for mass-spawning species such as pearl oysters, where mating occurs simultaneously between groups of males and females. To provide mating designs for these species, we adapted the mate selection algorithm to optimize selection of individuals with high breeding values and limiting relatedness between members of each group.

We used our optimized solution (OS) to obtain mating designs with group size from 5 to 80 oysters from each sex, using 1 to 17 tanks. We used a population of 176 oysters with known relatedness and allocated high EBVs to highly related individuals. Mating designs were compared with the selection of individuals based solely on their EBVs (BS). On average, OS mating designs selected individuals with EBVs 13% lower than BS and average relatedness between male and female was below the set threshold of 0.15 for both methods. However, on average, 16% of matings with BS solutions were between individuals over the threshold while no problematic mating were observed in OS solutions.

In conclusion, mass-spawning mate allocation provides a mating design useful in breeding programs with no mating between related individuals and high average breeding values for each group. Additionally, this method is very versatile as the relatedness matrix can be easily replaced by pedigree or genomic relationship matrices, the number of groups is user determined, as well as the number of males and females for each group.

Improved heritability models improve our ability to analyze complex traits.

Dr Doug Speed¹

¹Aarhus University, Denmark

Session 6, November 4, 2020, 7:00 PM - 8:35 PM

Biography:

*Doug Speed develops statistical methods for better analyzing complex trait genetic data. He releases these method through his software package *ldak* (www.ldak.org). Currently, *LDAK* includes the best tools both for understanding genetic architecture and for creating prediction models (polygenic risk scores).*

When analyzing a complex trait, the three main aims are to detect causal variants, construct prediction models and understand genetic architecture. Except for the most basic analyses, it is necessary to assume a heritability model, which specifies how we expect heritability to be distributed across the genome. Over the last few years, there has been much debate over the most appropriate heritability model. The authors of the software GCTA recommend the GCTA-LDMS Model, the authors of LD Score Regression recommend the Baseline LD Model, while we recommended the LDAK Model.

We provide a method for evaluating heritability models using summary statistics from genome-wide association studies (GWAS). By analyzing results from 30 GWAS (average sample size 120,000), we show that of the existing models, the Baseline LD Model fits real data best, although it can be improved by modifying how it assumes heritability varies with minor allele frequency. Our new method improves our understanding of the genetic architecture of complex traits. In particular, it can be used to estimate the selection parameter from summary statistics; we find strong evidence of negative selection for a range of traits, including height, college education and blood pressure. We also demonstrate how by incorporating the best-performance heritability model in existing tools (e.g., Bolt-LMM, BayesR, LDPred), we can improve power to detect causal variants and construct more accurate prediction models.

Genomic prediction using a multi-breed reference data of purebred and admixed populations

Dr Emre Karaman¹

¹ Aarhus University, Denmark

Session 6, November 4, 2020, 7:00 PM - 8:35 PM

Biography:

Emre Karaman is a post-doc at the Centre for Quantitative Genetics and Genomics at Aarhus University, Denmark. His research interests are mainly involved in the development and test of statistical methods and models for the analysis of livestock data. His recent research activities have concentrated on the use of genomic data for breeding value estimation of purebred and crossbred animals. These efforts include the development of in-house simulation and analysis tools to apply the quantitative and statistical genetics theory to medium-scale data sets.

Genomic prediction utilizes the linkage disequilibrium (LD) between genome-wide single nucleotide polymorphisms (SNPs) and quantitative trait loci (QTL). When predictions for multiple breeds are performed jointly, accuracy of predictions relies on the consistency of SNP-QTL LD across breeds. However, LD and persistence of phase may vary among the breeds. In rotational crossbreeding, genome of crossbred animals becomes an admixture of the breeds in the rotation. Modelling of data from admixed individuals may have a large impact on the prediction accuracy. In this study, we propose a methodology suitable for genomic prediction using a reference population of purebred and admixed animals. An admixed population and three purebred populations were simulated for nine generations, using animals with genomic data from three dairy cattle breeds as base populations. We investigated the impact of correlation of QTL effects among the breeds, and the heritability of the trait on accuracy of genomic predictions. Three approaches were compared: (i) treating the data of different populations as of a single homogeneous population, (ii) considering breed-specific SNP effects without accounting for correlations of SNP effects among the breeds, and (iii) using priors which lead to region specific correlations among the breeds. Assuming breed-specific SNP effects without accounting for correlations among the breeds generally improved accuracies over a traditional SNP-BLUP model, where the data were treated as of a single homogeneous population. Accuracies were also improved when the SNP effects were assumed to be breed-specific but correlated, particularly when the predictions relied on the region-wise, rather than genome-wide, correlations.

Genomic and temporal analysis of genetic covariation between traits

Dr Gregor Gorjanc¹

¹University of Edinburgh, United Kingdom

Session 6, November 4, 2020, 7:00 PM - 8:35 PM

Biography:

I lead the HilghladerLab at The Roslin Institute (University of Edinburgh, UK). We use data science, genetics, and breeding to manage and improve populations. We are particularly interested in: (i) quantitative methods for genetics and breeding, (ii) design and optimisation of breeding programs, and (iii) analysis of phenotypic and genetic data to unravel biology and to inform new ways of improving populations. We work with a range of species, spanning animals, plants and insects.

Understanding genetic variation is a fundamental task in quantitative genetics. Traditionally we used the pedigree-based model to estimate genetic (co-)variances between founders of a pedigree. We can estimate temporal changes in genetic (co-)variances by summarising sampled realisations of genetic values from a fitted model. Here we extend this approach to a marker-based model for temporal and genomic analysis. We demonstrate the approach by analysing dairy and beef traits in ~9K progeny-tested dual purpose Fleckvieh bulls genotyped with ~50K markers. We fitted a multivariate marker-model and used sampled realisations of marker effects to calculate realisations of genetic values for the whole genome or genome regions. We then summarised these genetic values to analyse genomic and temporal trends in genetic covariation within and between the traits. Results show differences by genome regions and temporal trends (decrease of genetic variation over time and build-up of negative genetic covariances). Genomic analysis of the trends shows that most of the changes are driven by the build-up of negative linkage-disequilibrium (the Bulmer effect) for individual traits as well as for pairs of traits, while allele frequencies and corresponding genetic covariances changed very little. These observations are consistent with the directional selection process in the breeding programme and a near infinitesimal genetic architecture of the analysed traits.

FstSeg: an efficient multiple changepoint procedure/algorithm for the detection of local signatures of selection

Dr Tristan Mary-Huard^{1,4}, Dr Guillem Rigai^{2,3}

¹GQE - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France,

²Laboratoire de Mathématiques et Modélisation d'Evry (LaMME), Université d'Evry Val d'Essonne, UMR CNRS 8071,

ENSIIE, USC INRA, Evry, France, ³Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud, Université

Evry, Université Paris-Saclay, Orsay, France, ⁴UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, , France

Session 6, November 4, 2020, 7:00 PM - 8:35 PM

The detection of genomic region involved in local adaptation is, arguably, one of the main challenges of modern population genetic. Many statistical procedures have been developed to detect them. All are computationally slow. This hamper their application to high-throughput SNP genotypic datasets. Indeed, on such datasets their runtime is in hours if not days. Most procedures rely on a divergence score based on population allelic frequencies measured at each marker (e.g Fst). This score is then analyzed either (i) through multivariate procedures that perform a joint analysis of the score but do not account for the spatial arrangement of the markers (e.g. Bayesian Factor models) or (ii) through genome scan procedures that account for the spatial distribution of the signal (e.g. BayeScan). We follow this last line of research and propose a new genome scan approach based on the robust Fst estimator recently proposed by Bhatia. We demonstrate that the problem of segmenting the genome into Fst-homogeneous regions can be recast as a weighted changepoints detection problem. When comparing two populations we propose to use the quasi-linear, yet exact fpop-algorithm, to recover the penalized maximum likelihood estimator. It runs in a matter of minutes for a 1KG dataset. This drastically reduces the computational burden while retaining strong statistical guarantees. We illustrate our approach on the 1KG dataset.

Multi-trait analysis with Bayesian models instructed by functional priors identify informative markers that predict traits across populations and countries

Dr Ruidong Xiang¹

¹Agriculture Victoria, Australia

Session 6, November 4, 2020, 7:00 PM - 8:35 PM

Biography:

Ruidong Xiang obtained his Ph.D in animal genetics at the University of Adelaide, South Australia. Transitioned from wet-lab molecular biology to statistical genetics, his research focuses on understanding the mechanisms via which genetics shape phenotype in agriculture species. Currently, he works as a Research Fellow at the University of Melbourne, on how to use multi-trait and multi-omics data to improve genomic selection of cattle complex traits. He works closely with scientists from Agriculture Victoria to integrate biological and functional information into breeding programs to more accurately select productive and healthy animals.

Accurate genomic prediction of phenotype, especially to predict across populations/breeds, is of high importance for precision medicine and agriculture. One way to achieve this is to obtain genotypes of causal variants but they are difficult to pinpoint. Alternatively, a set of informative markers that reasonably covers the genome and includes causal variants or ones in perfect LD them will allow accurate prediction. Directly genotyping informative markers on a customized array will avoid errors in imputation. We aim to identify a set of informative markers from 17.7 million sequence variants in 44,000+ Australian dairy cattle. We first reduce these variants to a set of 165K markers using their effects on 34 Cholesky-decorrelated traits and the external information of the Functional-And-Evolutionary Trait Heritability (FAETH) score on each variant. Then, a Bayesian method was applied to these 165K markers which 1) fitted all markers simultaneously 2) accounted for biological priors provided by the FAETH score and 3) avoided bias of variant pre-selection, and 4) estimated local gEBV with variants from each 50Kb chromosome-segment. We prioritised a set of 80K markers that can best explain the local gEBV variance and selected 47K to customise a bovine 50K genotyping array. This new array contained many pleiotropic variants that are also associated with RNA splicing (sQTLs), metabolic profiles (mQTLs), sites under selection (within-species) and conservation (cross-species). In independent 88,000+ cattle from the USA, New Zealand and Australia, our customised array outperformed the existing 50K array in genomic prediction of multiple traits across breeds and countries

Quantitative genetics of the genotype-phenotype Map

Professor David Houle¹

¹*Dept. Of Biological Science, Florida State, Tallahassee, United States*

Session 7, November 5, 2020, 7:00 AM - 8:25 AM

Biography:

*David is an evolutionary geneticist based at Florida State University who studies the relationship between the genomic, genetic and functional underpinnings of life and the process of adaptation at the phenotypic level. His lab uses the appendages of *Drosophila melanogaster* as a model. His current research goal is to understand how the processes of mutation and the forces that maintain quantitative genetic variation within populations are connected to long-term evolutionary trends.*

Evolution over long time scales is paradoxically predictable from the genetic variation within populations. To explain this observation, either the genotype-phenotype (GP) map must constrain responses to selection or the map itself must have been shaped by the pattern of evolution. To distinguish these possibilities we need convert the GP map from a heuristic metaphor to an object of experimental study. We are attempting to do this for the wings of fruit flies (*Drosophila melanogaster*). We have been using a high-throughput technique to measure wings to characterize variation within and among species, to perform artificial selection on wing phenotype, and to perform a genome-wide association study. We are estimating a Dictionary of Genetic Effects by quantitatively manipulating gene expression and then measuring the effects on wing phenotype. Remarkably the variation generated by the Dictionary recapitulates the mutation and genetic covariance matrices for the wing. To complement these results, we are developing a cell-based model of wing development that can predict the consequences of changes in gene function. Combining these approaches allows us to generate hypotheses about the form of the GP map for wings, and then test them.

Turbocharging Fruit Tree Breeding

Dr Satish Kumar¹

¹*Plant and Food Research, Havelock North, New Zealand*

Session 7, November 5, 2020, 7:00 AM - 8:25 AM

Biography:

Satish Kumar is a scientist at Plant & Food Research, New Zealand. He earned his PhD degree in Genetics and Breeding from Massey University, and then worked as a scientist (Tree Improvement) at the New Zealand Forest Research Institute, Rotorua for 10 years before joining Plant & Food Research in 2009. His area of research includes breeding and genetic evaluation strategies, germplasm conservation and improvement, and genetic architecture of fruit traits. Dr Kumar led the first application of genomic selection and GWAS in apple breeding. His current research focuses on the integration of genetics and genomics technologies into breeding programmes of horticultural crops.

The genetic diversity of germplasm of fruit crops serves as a useful resource for introgressing novel genes to meet consumer preferences and biotic/abiotic challenges. Most fruit tree crops have a long juvenility period, so that traditional fruit breeding often takes up to 25 years to develop commercial cultivars. The timeline could be even longer if novel attributes (e.g., disease resistance) are introgressed from wild germplasm to first develop parents for cultivar breeding. Genomics-assisted breeding provide more direct and quicker solutions to allow faster accumulation of favourable alleles from donor and recipient parents, resulting in faster development of high-value cultivars. Sequencing of genomes of some major fruit species during the last 10 years has facilitated the development of high-throughput cost-effective genotyping platforms, which provided opportunities for application of genomic selection (GS). The most attractive feature of GS is that it has the potential to dramatically shorten the breeding cycle length by obviating the need to phenotype selection populations. Fruit characters that constitute a superior cultivar are often polygenic in their inheritance, and GS is best suited for selection of cultivar candidates for such traits. GS studies in various fruit tree crops have provided promising selection accuracies. Application of GS, in combination with the use of artificial growth conditions to facilitate early transition from the juvenile to the adult phase, has accelerated the delivery of novel cultivars. Latest results from GS studies on apple and pear, and future applications of GS in fruit tree crops will be discussed.

Tracking the tempo and target of natural/artificial selection in a thousand mice and butterflies by haplotype tagging and sequencing

Dr Frank Chan¹

¹Max Planck Society, Germany

Session 7, November 5, 2020, 7:00 AM - 8:25 AM

Biography:

Frank Chan is a Group Leader at the Friedrich Miescher Laboratory, Max Planck Campus Tübingen in Germany. With a passion for genetics, his group focuses on understanding the molecular basis and evolution of complex traits, by establishing the mechanisms connecting adaptive mutations to their phenotypic and evolutionary impact. His work leverages unique study systems including F1 hybrid stem cells and rapidly evolving mouse populations (e.g., "Longshanks" mouse selection experiment and large Faroese Island mice). These projects draw on techniques from single-cell sequencing, tissue engineering and image analysis, under the unifying theme of using data-driven approaches to study genetics and development.

Identifying and quantifying the target and strength of selection is an important task in quantitative genetics. With short-read sequencing, many population genomics analyses focus on single nucleotide polymorphisms (SNPs) but lack data on the underlying haplotype block - valuable information for natural selection on a genomic region. We introduce here "haplotagging", a simple, one-step technique for linked-read sequencing using standard Illumina reagents with no instrumentation or extra costs. Haplotagging preserves linkage information among short reads enabling reconstruction of molecules and haplotypes hundreds of kilobase in size. It is affordable and scalable, allowing multiplexing of 96 or 384 samples into single Illumina lanes, and enables haplotype discovery at low sequencing coverage.

We demonstrate haplotagging with a mouse selection experiment for longer legs ("Longshanks" mice). By sequencing 1400 breeders in a fully pedigreed time series, we directly track the frequency and segregation of haplotypes through 20 generations of selection and estimate selection coefficients across the genome. Next, using 500 *Heliconius* butterflies from a hybrid zone, we show how direct haplotyping of selected variants reveals candidate SNPs unique to the selected haplotype. In both cases, using haplotypes enhances the genetic signature and avoids problems common in SNP-based analyses.

We anticipate that haplotagging to be broadly useful in quantitative genetics, including (meta-)genome assembly and large-scale mapping studies with thousands of samples.

Genetic correlation between two populations can be predicted from the F_{ST} between the populations and the non-additive genetic variance

Professor Michael Goddard¹

¹*Agriculture Victoria, Australia*

Session 8, November 5, 2020, 12:00 PM - 1:25 PM

Biography:

Prof. Michael Goddard holds a joint appointment at the faculty of Veterinary and Agricultural Sciences, University of Melbourne and at Agriculture Victoria as a professorial fellow in animal genetics. His research is on quantitative genetics in livestock and humans.

Marker genotypes allow us to calculate the relationship between individuals even if they belong to different populations. This allows us to calculate the genetic correlation (r_g) between the same trait measured in different populations. This has important implications in human medicine (e.g. predicting disease risk in different populations) and agriculture (e.g. selecting animals for crossbreeding vs purebreeding). If there is non-additive genetic variance for the trait, then the allele substitution effect (b) depends on the allele frequencies in the population and hence b may vary between populations and so r_g may be <1 . In fact, $1-r_g = 4 F_{ST} (VD + VAA)/VA$ where VD = dominance variance, VAA is additive by additive epistatic variance and VA = additive variance. For most traits $(VAA + VD)/VA < 0.33$ and typically $F_{ST} < 0.1$, so r_g should be > 0.87 . However, most estimates of r_g are < 0.87 . r_g can only be estimated using SNP data but a closely related parameter, the correlation between the breeding value of an individual in purebreds and crossbreds (r_{pc}) can be estimated directly from phenotypes. Most estimates are > 0.8 . r_{pc} can be predicted by a very similar formula to r_g . So r_g should be only a little smaller than r_{pc} . If $r_{pc} < 1$, it implies that heterosis in crosses will decline as selection improves the purebreds, but there is no evidence that this occurs. It is difficult to reconcile the low estimates of r_g with the theory or the estimates of r_{pc} or observations of heterosis.

Cross-population fine-mapping of complex traits and diseases in ~675,000 individuals across three global biobanks.

Mr Masahiro Kanai¹

¹Broad Institute of MIT & Harvard University, USA

Session 8, November 5, 2020, 12:00 PM - 1:25 PM

Biography:

Masahiro Kanai is a PhD candidate in the Bioinformatics and Integrative Genomics PhD Program, Harvard Medical School. Co-advised by Drs. Mark Daly and Hilary Finucane at the Analytic and Translational Genetics Unit, Massachusetts General Hospital and Broad Institute, his research focuses on cross-population analysis of complex diseases and traits to better understand their genetic architecture across diverse populations. Prior to joining ATGU, Masahiro completed his B.S. degree in Japan, where he worked closely with Dr. Yukinori Okada to study genetics of complex traits in the Japanese population using the Biobank Japan.

Identifying causal variants for complex traits is one of the major challenges in human genetics. The causal variants in most GWAS loci remain unknown due to lack of power and to high linkage disequilibrium (LD) in a locus. Cross-population fine-mapping studies may improve fine-mapping resolution by leveraging differences in LD and minor allele frequency (MAF) across populations. However, lack of availability of non-European GWAS and proper statistical methods have limited the success of previous studies. Here, we fine-mapped hundreds of complex traits and diseases from three large-scale biobanks, UK Biobank (n = 361,194), FinnGen (n = 135,638) and Biobank Japan (n = 179,066). We first conducted single-population fine-mapping using FINEMAP (Benner et al., 2016) and SuSiE (Wang et al., 2018). In total, 54,403 putative causal variants were identified (posterior inclusion probability [PIP] > 0.1). Our results suggested that single-population fine-mapped variants are rarely shared across populations (1.8%), whereas their posterior effect sizes in allelic scale are largely consistent if shared (PIP-weighted regression slope: 1.02). This inconsistency of PIP is partly due to difference in MAF and imputation quality which resulted in missing causal variants in one population. Allelic series analysis highlights different population-specific coding variants were fine-mapped together in specific exons, which also differentiates putative causal variants across populations. To maximize our power of cross-population analysis, we developed a novel cross-population fine-mapping method that extends SuSiE model into cross-population setting to allow shared/population-specific causal effects. Our preliminary results from simulations suggest that our method successfully improves fine-mapping power and resolution.

Partitioning risk for complex diseases by race and genetic ancestry

Ms Ky'Era Actkins¹

¹Meharry Medical College Tennessee, USA

Session 8, November 5, 2020, 12:00 PM - 1:25 PM

Biography:

Ky'Era Actkins is a doctoral candidate at Meharry Medical College in Nashville, TN. Under the mentorship of Dr. Lea Davis at Vanderbilt University Medical Center, her research focuses on elucidating the genetic architecture of health disparities with an emphasis on women's reproductive health. To achieve this, she is using the clinical and genetic information captured in electronic health records to understand the underlying determinants of health outcomes. Ky'Era plans to pursue a postdoctoral fellowship to continue applying computational approaches to study conditions that disproportionately affect women's health.

Health disparities have complex biological, environmental, social, and geopolitical origins. Although human genetic studies have identified population specific associations with multiple complex diseases, such as kidney failure and hypertension, it remains unclear whether these genetic associations can fully explain these disparities. Current genetic models often lack the data and statistical methods necessary to capture and account for non-biological effects that increase risk for disease and exacerbate health disparities. Disentangling these factors is crucial to understanding the etiology of health disparities. We developed an approach to analyze the independent effects of genetic ancestry and the social construct of race, recorded in the electronic health record by third parties. Using BioVU, a repository of DNA linked to de-identified electronic health records, we observed that genetic ancestry, measured by the first principal component, is associated with approximately 20% of the medical phenome. Of those associations, 96% are in the direction of risk for increased African ancestry. We then partitioned the independent effects of genetic ancestry from race by calculating regression residuals for one after removing variance due to the other. African genetic ancestry (adjusted for race) is significantly associated with "Uterine leiomyoma" (OR=1.32, p=5.62e-07), "Renal dialysis" (OR=1.70, p=6.31e-06), and "Sickle cell anemia" (OR = 1.85, p = 9.55e-11). Race (adjusted for genetic ancestry) remained significantly associated with "Hypertension" (OR=1.16, p=7.57e-05). This method allows us to examine the effects of genetic ancestry and the social construct of race separately, enabling us to parse out our resources and efforts to each contributing cause more effectively.

Family-Based GWAS of Educational Attainment

Dr Alexander Young¹, Dr Patrick Turley², Mr Sean Lee³, Professor Peter Visscher⁴, Professor Augustine Kong⁵, Professor David Cesarini⁶, Professor Daniel Benjamin¹

¹University of Southern California, Los Angeles, USA, ²Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA, ³Harvard University, Boston, USA, ⁴The University of Queensland, Brisbane, Australia, ⁵University of Oxford, Oxford, United Kingdom, ⁶New York University, New York, USA

Session 8, November 5, 2020, 12:00 PM - 1:25 PM

Biography:

Alexander Young is a Research Scientist with the Social Science Genetic Association Consortium (SSGAC) at USC. Following degrees in Mathematics and Statistics and in Computational Biology at the University of Cambridge, Alex pursued a doctorate in Genomic Medicine and Statistics with Peter Donnelly at the Wellcome Trust Centre for Human Genetics, University of Oxford. During his PhD, he visited deCODE Genetics in Reykjavik and began to work out how to use large amounts of genetic data on families to tease apart genetic and environmental effects on traits. He pursued postdoctoral research with Augustine Kong at deCODE Genetics in Iceland and at the Big Data Institute at the University of Oxford. While working with the deCODE data, Alex developed a novel method for estimating heritability using genetic data on parents and offspring. He continues to pursue development of novel methodology to better analyse and understand the role of genetic variation in human trait variation, with a particular focus on social and behavioural traits.

Genetic analyses of educational attainment (EA) have shown that EA is heritable and affected by thousands of common genetic variations. Furthermore, genetic variants present in the parents affect offspring through the environment, an example of indirect genetic effects. The degree to which genetic relationships between educational attainment and other traits reflect direct versus indirect genetic effects is currently unknown. We conducted a genome-wide association study (GWAS) of educational attainment (EA) in individuals of European ancestry, using a sample of ~3 million to estimate additive effects and a sample of ~2.3 million to estimate dominance effects. We found 3899 loci with genome-wide significant additive effects and 2 loci with genome-wide significant dominance effects. Our analysis shows that dominance effects play only a limited role in the effects of common genetic variants on EA. We constructed a polygenic score (PGS) from the additive effect estimates, which explained 16.3% of the variation in EA. We examined the degree to which the EA PGS predicts various medical and behavioural traits due to direct versus indirect genetic effects. We estimated direct and indirect genetic effects of genome-wide SNPs using both observed parental genotypes and parental genotypes imputed from sibling genotypes. Using these results, we estimated that the correlation between SNP effects from GWAS on unrelated individuals have correlation 0.61 (S.E. 0.09) with direct genetic effects. This implies that GWAS on unrelated individuals may give biased estimates of direct genetic effects on EA due to some combination of population stratification and indirect genetic effects.

How fast are wild animals currently adapting?

Dr Timothée Bonnet¹

¹*Australian National University, Australia*

Session 9, November 5, 2020, 7:00 PM - 8:30 PM

Biography:

Post-doc at the Research School of Biology and the Biological Data Science Institute at the Australian National University. Interested in contemporary selection and evolution in wild animal populations.

The additive genetic variance in relative fitness determines the rate of adaptation; this is the Fisher's fundamental theorem of natural selection. Additive genetic variance in fitness is therefore a central parameter describing evolutionary and demographic processes in a population. Unfortunately, inference of genetic variance in fitness requires enormous quantities of high quality data, but also statistical methods that are only just reaching maturity. As a consequence existing estimates are very sparse and uncertain in wild populations. We assembled data from 20 long-term monitored wild populations of birds and mammals, with high-quality pedigrees and records of individuals lifetime fitness, and analyzed them with quantitative genetic "animal models" using recent conceptual and methodological improvements to better deal with non-Gaussian traits. We estimated additive genetic variance and narrow-sense heritability in relative fitness, therefore almost doubling the number of estimates available in wild vertebrates to date. Our analyses indicate widespread evidence for additive genetic variance in fitness. We discuss their interpretation in term of rates of adaptive evolution across vertebrates and the relevance of these rates for the fates of populations.

Estimating maternal and paternal genetic effects on offspring phenotypes in large scale cohorts when parental genotypes are unavailable

Professor David Evans^{1,2}, Doctor Daniel Hwang¹, Doctor Justin Tubbs³, Mr Justin Luong¹, Doctor Nicole M Warrington¹, Doctor Gunn-Helen Moen^{1,2,4}, Professor Pak Sham³, Doctor Gabriel Cuellar-Partida¹
¹University of Queensland, BRISBANE, Australia, ²Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, ³Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China, ⁴Institute of Clinical Medicine, Oslo, Norway

Session 9, November 5, 2020, 7:00 PM - 8:30 PM

Biography:

David Evans is Professor of Statistical Genetics at the University of Queensland Diamantina Institute. He completed his PhD in Statistical Genetics at University of Queensland in 2003, before undertaking a four year post-doctoral fellowship at the Wellcome Trust Centre for Human Genetics, University of Oxford. In 2007 he moved to take up a Senior Lecturer position at the University of Bristol before returning in 2013 to Australia to take up his current position at the University of Queensland. He has co-led several large GWAS consortia and is interested in the development of statistical methods for gene mapping and causal inference.

Parental genetic effects may be defined as the influence of parental genotypes on offspring phenotypes over and above that which results from the transmission of genes from parents to children. However, given the relative paucity of large-scale family based cohorts around the world, it is difficult to demonstrate parental genetic effects on human traits, particularly at individual loci. In this presentation, we illustrate how parental genetic effects on offspring phenotypes, including late onset diseases, can be estimated at individual loci using large scale GWAS data, even in the absence of parental genotypes. Our strategy involves creating “virtual” mothers and fathers by estimating the genotypic dosages of parental genotypes using physically genotyped data from relative pairs. We then utilize the expected dosages of the parents, and the actual genotypes of the relative pairs, to perform conditional genetic association analyses and obtain asymptotically unbiased estimates of maternal, paternal and offspring genetic effects. We develop a freely available R shiny web application that quantifies the power of our approach using closed form asymptotic solutions. We implement our methods in a user-friendly software package which allows users to quickly and efficiently impute parental genotypes across the genome in large genome-wide datasets, and then use these estimated dosages in downstream linear mixed model association analyses. Finally, we illustrate our methods by using the UK Biobank data to estimate parental genetic effects on offspring birthweight.

Genomic architecture of 184 plasma proteins in 18,884 individuals: the SCALLOP Consortium

Erin Macdonald-Dunlop¹, Peter K Joshi¹, James E Peters^{3,4}, Lasse Folkersen⁵, Erik Ingelsson^{6,7,8}, Paul RHJ Timmers¹, Karl Michaelsson⁹, Stefan Gustafsson¹⁰, Stefan Enroth¹¹, Åsa Johansson¹¹, Gustav Smith¹², Daria Zhernakova¹³, Agneta Siegbahn¹⁴, Anette Kalnapenkis¹⁵, Niclas Eriksson¹⁶, Eleanor Wheeler¹⁷, Jingyuan Fu¹³, Lude Franke¹³, Caroline Hayward², Lars Wallentin¹⁴, Tonu Esko^{15,18}, Eleftheria Zeggini¹⁹, Charlotte Teunissen²⁰, Claudia Langenberg¹⁷, Oskar Hansson^{21,22}, Per Eriksson²³, Ulf Gyllenstein¹¹, Adam S Butterworth³, Anders Mälarstig^{23,24}, James F Wilson^{1,2}, SCALLOP Consortium

¹Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom, ²MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom, ³Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Worts Causeway, Cambridge, United Kingdom, ⁴Health Data Research UK, , United Kingdom, ⁵Institute of Biological Psychiatry, Copenhagen, Denmark, ⁶Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA, ⁷Stanford Cardiovascular Institute, Stanford University, , Stanford, USA, ⁸Stanford Diabetes Research Center, Stanford University, Stanford, USA, ⁹Department of Surgical Sciences, Uppsala University, Uppsala, Sweden, ¹⁰: Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden, ¹¹Department of Immunology, Genetics, and Pathology, Biomedical Center, Science for Life Laboratory (SciLifeLab) Uppsala, Uppsala University, Uppsala, Sweden, ¹²Department of Cardiology, Clinical Sciences, Lund University and Skåne University Hospital, Lund, Sweden, ¹³Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, ¹⁴Department of Medical Sciences and Uppsala Clinical Research Center, Uppsala University, , Sweden, ¹⁵Estonian Genome Center, University of Tartu, Tartu, Estonia , ¹⁶Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden, ¹⁷MRC Epidemiology Unit, University of Cambridge, Cambridge, UK, ¹⁸Broad Institute of MIT and Harvard , Boston, USA, ¹⁹Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), , Germany, ²⁰Department of Clinical Chemistry, Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, ²¹Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden, ²²Memory Clinic, Skåne University Hospital, Malmö, Sweden, ²³Department of Medicine, Karolinska Institutet, Stockholm, Karolinska University Hospital, Solna, Sweden, ²⁴Pfizer Inc, , USA

Session 9, November 5, 2020, 7:00 PM - 8:30 PM

Proteins are the fundamental building blocks of life and participate in all biological processes. Plasma proteins show great promise as novel disease biomarkers, and their genetic determinants also help unravel underlying networks and causal pathways, pointing to new drug targets. However, to date, discoveries have been limited due to small sample sizes. The SCALLOP consortium of 20 cohorts was established to discover protein quantitative trait loci (pQTLs) in a large combined sample. We used Olink proximity extension assays to quantitate the abundances of 184 plasma proteins (from the CVD II & III panels) in up to 18,884 subjects. Genome-wide association meta-analysis revealed 17,506 genome-wide significant SNPs ($P < 2.72 \times 10^{-10}$) associated with at least one protein, including both cis- and trans-pQTLs. Contrary to findings from previous smaller individual studies, our increased sample size reveals the majority of proteins to have at least one pQTL. Utilising a variety of state-of-the-art methodologies, we show how a subset of these pQTL co-localise with eQTLs and disease-related traits. We also elucidate novel regulatory mechanisms and protein-protein interaction networks from our greatly expanded catalogue of trans-pQTL. Finally, we apply two-sample Mendelian randomisation to explore causal mechanisms in a broad set of complex diseases and risk factors.

Genotype-by-environment interactions among Scots pine vitality and growth traits, in harsh and mild environments in Northern Scandinavia

Dr Ainhoa Calleja-Rodriguez¹

¹Skogforsk (The Forestry Research Institute of Sweden), Sweden

Session 9, November 5, 2020, 7:00 PM - 8:30 PM

Biography:

Ainhoa Calleja-Rodriguez (Doctor of Forestry) is researcher in Tree Breeding at Skogforsk (The Forestry Research Institute of Sweden), which is the Institution responsible of the Operational Tree Breeding Programs in Sweden. She has received her graduate education at the Swedish University of Agricultural Sciences, on the topic "Quantitative Genetics and Genomic Selection of Scots pine". Her work focuses on quantitative genetics and quantitative genomics, but she is also interesting in tree resistance to forest pests and diseases.

In the harsh climates of Northern Sweden, tree vitality and growth are two of the most important traits to maximize the stand volume production, within the Scots pine breeding program. However, in such areas mortality is considerable high and normally caused by accumulated damage, such that only vigorous individuals can express their growth potential full, influencing the genetic association between tree vitality and growth. In the current study, multivariate multi-environment analyses, with factor analytic variance-covariance structures, were used to estimate the degree and pattern of genotype-by-environment interactions ($G \times E$), as well as to evaluate the genetic relationship between tree vitality and growth, in two different populations each consisting on five trials located in sites with heterogeneous temperatures. 16 to 19 variables were possible to fit and notable $G \times E$, based on type-B genetic correlations, were observed for growth and vitality traits, among trials with larger differences in temperature. Type-A genetic correlations between vitality and growth were higher as the mortality and harshness of the sites increased; type-AB genetic correlations between both traits changed from negative to positive, as the differences in temperature between sites decreased. Additive and environmental coefficients of variation for height increased with the age in harsh sites, indicating that those trees are still vulnerable to environmental disturbances at older ages. The positive genetic association detected between tree vitality and height in harsh sites, weakened as the temperature increased, suggesting that tree growth on harsh and mild sites should be treated as separate traits.

Phenotypic covariance across the entire spectrum of relatedness for 88 billion pairs of individuals

Dr Kathryn E Kemper¹, Dr Loic Yengo¹, Dr Zhili Zheng¹, Dr Abdel Abdellaoui², Prof Michael E Goddard^{3,4}, Prof Naomi Wray¹, Prof Jian Yang¹, Prof Peter Visscher¹

¹University of Queensland, Brisbane, Australia, ²University of Amsterdam, Amsterdam, the Netherlands, ³University of Melbourne, Melbourne, Australia, ⁴Department of Economic Development, Jobs, Transport and Resources, Bundoora, Australia

Session 9, November 5, 2020, 7:00 PM - 8:30 PM

Biography:

Kathryn is a postdoctoral fellow at the University of Queensland. She is fascinated in all things inherited, particularly the use and estimation of genetic variation within families. Her experience includes both human and livestock complex trait genetics, and population genetics. Kathryn obtained her PhD at the University of Melbourne (2011).

Attributing the similarity between individuals to genetic and non-genetic factors is a central question in quantitative genetics and genetic epidemiology. Here, we consider the phenotypic covariance between pairs of individuals across the entire spectrum of relatedness, from unrelated pairs through to identical twins, using a genomic relationship (π) matrix among 420,000 individuals. From single dataset and approach, we show the relationship between π and phenotypic covariance has two approximately linear sections, consistent with quite different estimates of heritability for conventionally unrelated pairs ($\pi < 0.02$) and close relatives ($\pi > 0.05$). For the first time, we show the phenotypic covariance increases faster than π in distant relatives ($0.02 < \pi < 0.05$). We use simulation to show that imperfect linkage disequilibrium between causal variants and common SNP can drive the observed relationship in distant relatives, and also that non-additive sources of genetic variation are inconsistent with observed phenotypic covariance in close relatives. To further quantify the influence of genetic and non-genetic factors, we performed a within-family analysis on the $\sim 20K$ full-sib pairs in our data and meta-analysed the results with published estimates. We discovered that non-random mating for height and educational attainment influences the underlying parameters under different experimental designs, and that common environmental effects can be confounded with these effects in some types of analyses. Our work highlights the importance understanding and appreciating the full spectrum of genetic and non-genetic influences on complex trait variation as genetic parameters are estimated with ever increasing precision.

Sex difference in genetic correlations and its implication on causality

Dr Xia Shen¹

¹University of Edinburgh, United Kingdom

Session 9, November 5, 2020, 7:00 PM - 8:30 PM

Biography:

Xia is a statistical geneticist specialized in statistical modeling of high-throughput omics big data. His interdisciplinary research consists of genetics, statistical inference, and computational biology. He currently leads an international collaborative research team across institutions in the UK, Sweden, and China.

Sex differences exist for nearly all human complex traits, where genetics contributes a part of such differences. Although gene-by-sex interaction has been discussed in recent literature, most underlying genetic mechanisms have been unexplored. We develop a new method for the estimation of genetic correlations using genome-wide association study (GWAS) summary statistics, greatly improving the estimation accuracy. This allows us to detect a significant amount of sex differences in many genetic correlations across human complex phenotypes. Thereafter, we focus our investigation on the emerging area of causal inference based on the GWAS discoveries for complex traits, thereby incorporating multi-trait genetic architecture when inferring causal relationships between the traits. We show that the sex heterogeneity in genetic correlations provides a robust test for the existence and direction of a causal relationship between phenotypes. By contrasting the GWAS results of males and females in the UK Biobank, testing thousands of pairs of human complex traits, we suggest the causal relationships between various human behaviors and complex diseases. The causal inference results are significantly correlated with those from existing causal inference methods, while the results suggested by genetic correlation heterogeneity between sexes appear to have higher specificity. This provides a distinct method with distinct assumptions to triangularize evidence for inferring causal relationships using GWAS results. Our results suggest social policy and guidance as tools for disease prevention and support diversification of genomic studies to a broader range of populations.

Predicting Phenotype: Modeling Approaches Within The GxE Space

Dr Jacob Washburn¹

¹USDA Missouri, USA

Session 10, November 6, 2020, 7:00 AM - 8:30 AM

Biography:

Dr. Jacob D. Washburn is a Research Geneticist at the United States Department of Agriculture – Agricultural Research Service in Columbia, Missouri, USA. He studies the genetic and physiological basis of plant phenotypes and adaptations to environmental and human imposed conditions. He uses growth chamber and field experiments, as well as quantitative genetics, machine learning, and process-based modeling. His previous appointments include a B.S. in Biology from Brigham Young University, an M.S. in plant breeding from Texas A&M University, a Ph.D. in Biological Sciences from the University of Missouri, and postdoctoral research at Cornell University and the University of Queensland.

Predicting an organism's phenotype from the combination of its genotype and environment has long been a goal of both basic and applied research. However, this goal is complicated by the fact that both the genomes of organisms, and the environmental conditions in which they grow and develop are complex. Drawing upon analytical tools from quantitative genetics, physiological growth modeling, and machine learning, a series of methods were developed for predicting plant phenotypes across more than 70 environments (year and location combinations) and 1000 genotypes. These methods include optimized physiological growth models and convolutional neural networks, and use inputs like weather, soil, and genetic data to predict crop grain yield/plant fitness. The accuracy of these predictions varies substantially (0-90%) by environment, genotype, and modeling method. Further analysis provides insights into genetic, environmental, and sampling considerations that impact the likelihood of predictive success. These methods will afford researchers and practitioners a greater ability to predict and understand an organism's phenotype under various environmental and genetic conditions. Potential applications of the approaches include agriculture, breeding, climate mitigation, ecosystem management, and others.

Population genetics of transposable elements in the maize Wisconsin diversity panel

Dr Christine O'Connor¹

¹University of Minnesota, USA

Session 10, November 6, 2020, 7:00 AM - 8:30 AM

Biography:

Christine is a bioinformatics analyst in the Minnesota Supercomputer Institute at the University of Minnesota. She supports research projects and develops custom clinical genomics pipelines. During her postdoc in Dr. Candice Hirsch's lab at the University of Minnesota Christine worked on a number of different projects that all focused on better understanding genome content variation in a 500-genotype maize diversity panel. In particular, she focused on understanding what characteristics of transposable elements, which make up 65% of the maize genome, affect their population frequency in maize.

Intact transposable elements (TEs) make up around 65% of the maize genome and can impact gene function and regulation. However, genome-wide patterns of TE polymorphisms in maize have only been studied in a handful of maize genotypes, due in part to the challenging nature of assessing highly repetitive sequences. We developed a method to use short read sequencing data to score the presence/absence of annotated TEs and validated the accuracy of this method using genome-wide TE annotation and polymorphism data from four maize genotypes for which we have reference genomes and genotype specific TE annotations. This method is able to identify a TE as present or absent with $\geq 85\%$ accuracy, and was implemented to score TEs in 500 genotypes. TE insertion frequency varies with location in the genome, age and size of the TE family. We also find that TE presence/absence variation generally mirrors maize breeding history. However, a subset of TE polymorphisms are not in linkage disequilibrium with nearby SNPs and are not reflective of pedigree relationships. Our results illustrate that TEs are variable between genotypes, characteristics of TEs can predict their frequency in a population, and have identified TEs that are associated with population structure, as well as TEs that may contain information not captured by SNP-based surveys. Accurately characterizing TE presence and absence with short read data has allowed the first population scale genome-wide assessment of TE variation and allowed us to better understand the roles TEs may play in functional phenotypic variation in maize.

Modelling GxE in High Throughput Phenotyping data from crop variety trials

Dr Joanne De Faveri¹

¹CSIRO Data61, Australia

Session 10, November 6, 2020, 7:00 AM - 8:30 AM

Biography:

Dr Joanne De Faveri is an applied statistician with interest in the application of statistical methods to agricultural research, especially plant breeding. She has spent most of her career as a Biometrician in Far North Queensland working with Horticulture breeding programs. More recently she joined CSIRO as part of SAGI-North, the Statistics for the Australian Grains Industry project, where she develops and applies statistical methods to grains research projects. Currently she is working on statistical methodology to best integrate High Throughput Phenomics (including aerial image, sensor and hyperspectral) data into crop breeding programs for better variety predictions.

Modern plant breeding trials are employing drones, helicopters, sensors and buggies to collect High Throughput Phenotyping (HTP) data at multiple timepoints over the crop growing season at multiple sites. Often the data is in the form of images or spectra and in the case of hyperspectral data may consist of thousands of wavelengths measured over time and environments. Whilst modelling the genetic effects of HTP traits is of primary interest there is also a need to account for non-genetic effects (design, spatial and temporal) to obtain accurate and unbiased estimates of the genetic effects. The interaction between genotype, environment, time and wavelength also needs to be modelled in a suitable way. Often HTP traits are measured with the aim of informing other traits of primary interest, for example yield. In many cases HTP data is used together with genomic information for genomic prediction of these primary traits. Functional Data Analysis (FDA) models provide one approach for the analysis of this type of data. In this talk new functional regression models in ASReml are presented, allowing for spatio-temporal modelling and efficient modelling of Genotype by environment (GxE) effects. The models are implemented in the analysis of multi-site, multi-time, hyperspectral image data from crop variety trials.

QTL mapping in autotetraploid multi-parent populations

Mr Rodrigo Amadeu¹

¹University of Florida, USA

Session 10, November 6, 2020, 7:00 AM - 8:30 AM

Biography:

Rodrigo Amadeu is a PhD student at University of Florida interested in quantitative genetics applied for plant breeding. He has a bachelor degree in Agriculture and a MS in Plant Genetics and Breeding both earned from University of São Paulo. His doctorate research is on the development of new quantitative genetics based-tools and software to guide genomic-assisted selection of autopolyploid crops. He is also the developer of different R packages such as AGHmatrix (relationship matrices computation) and diaQTL (multiparental QTL mapping).

Designed populations are a powerful tool for genetic mapping of the chromosomal regions, or quantitative trait loci (QTL), that influence traits of interest in plants and animals. However, genetic background effects often limit the applicability of results based on populations with only two founders. Multi-parent populations have been used to overcome this limitation in diploids, and our objective was to develop this capacity for autotetraploid crops, such as potato, blueberry, alfalfa, and rose. To cope with the high dimensionality of the problem (each founder can transmit 10 different diploid gametes), genetic effects were modeled as random, using Bayesian methods and software developed for genomic prediction. The new software allows for prior distributions that induce a combination of variable selection and shrinkage, with posterior means estimated based on Gibbs sampling. Utilizing a half-diallel mating design, and keeping the total population size constant, the influence of several variables on the receiver-operating characteristic curve was explored: (1) the number of parents, (2) the presence of selfed populations, and (3) the relative importance of QTL x sub-population interactions. Genotyping was simulated based on next generation sequencing platform, and the influence of marker density and total sequencing depth (as a proxy for genotyping cost) were also investigated.

Improving genomic prediction of target hybrids in unobserved environments using geospatial assessment of predictive analytics derived from machine learning techniques

Dr Diego Jarquin¹

¹University of Nebraska-Lincoln, USA

Session 10, November 6, 2020, 7:00 AM - 8:30 AM

Biography:

Dr. Jarquin is a Research Assistant Professor in the Department of Agronomy and Horticulture at the University of Nebraska-Lincoln (UNL). Dr. Jarquin is a statistician who merges statistical methodology, computer algorithm development, data science and collaborative work with plant sciences (plant breeding, biometrics, biostatistics, quantitative genetics, etc.). Dr. Jarquin is actively collaborating on several projects (Genomes to Fields, SoyNAM, etc.) with public (University of Tokyo, ICRISAT, CIMMYT, IRRI, EMBRAPA, CENICANA, etc.) and private (Advanta Seeds) sectors. Recently, Dr. Jarquin has earned the 2020 National Association of Plant Breeders Early Career Scientist Award.

Sustained increases of crop yield in the Midwest USA are produced by the development of better and more stable genotypes coupled with the introduction of improved technological developments, management practices, and resources availability. However, the occurrence of extreme hydrometeorological and climate events (EHCEs) may affect negatively these trends. So far, the geospatial extent of these effects on genotype responses are unclearly identified and consequently such effects across scales are poorly understood. The Genomes to Fields (G2F) project have monitored environmental variables through more than 100 weather stations since 2014 across the USA. In such locations a wealth of multidimensional, discontinuous and heterogenous data are also collected from multiple sources (i.e., remote and proximal sensing and various monitoring networks). We explored the potential of machine learning techniques to integrate multi-dimensional databases across scales using a conceptual model to geospatially identify the areas where the intensity of water deficits challenges our ability to predict phenotypes. We integrate Artificial Neural Networks and kriging interpolations to take advantage of the spatiotemporal proximity between G2F data and publicly available data. For describing environmental differences/similarities among environments we focused on environmental/hydroclimate variables (i.e., precipitation, temperature, wind, etc.). Results evidence an improvement on precipitation and temperature estimates of 20% in some locations when data from other sources (same location) was used as input. Also, the improvements in predictive ability of tested and untested genotypes in unobserved G2F environments varied between 200% and 250% compared with commonly used genomic selection models.

A statistical framework to incorporate high-throughput proxy phenotypes in genomic predictions for wheat breeding

Dr Lee Hickey¹

¹University of Queensland, Australia

Session 10, November 6, 2020, 7:00 AM - 8:30 AM

Biography:

Dr Hickey is a plant breeder and crop geneticist within the Queensland Alliance for Agriculture and Food Innovation at The University of Queensland, Australia. His lab conducts genetic studies on key biotic (foliar and soil-borne diseases) and abiotic (drought and heat) factors that limit the production and productivity of wheat and barley. The group also works to develop novel breeding tools and methodologies that improve agricultural efficiencies. He has a strong interest in combining leading-edge plant breeding technologies, including speed breeding, high-throughput phenotyping, genome editing and genomic selection, to accelerate genetic gain.

Phenotyping wheat plants, seeds and flour using image analysis, near infra-red and other technologies is attractive given the scale at which these technologies can be deployed and their low-cost per phenotype. These phenotypes are usually “proxies” for the real traits of interest (breeding goal traits), such as yield and end use quality. As such they are correlated with, but not the same as these breeding goal traits, and breeding directly for the proxy traits may have unintended consequences. To correctly incorporate proxies in prediction of genomic breeding values for wheat breeding, we suggest a multi-trait genomic restricted maximum likelihood approach. This approach has the advantages that, provided at least some lines are measured for the breeding goal traits, 1) the genetic correlations between the high-throughput proxy traits and the breeding goal traits are estimated, 2) genomic breeding values are predicted for all lines for the breeding goal traits, appropriately using and weighting the information from the proxy traits given the genetic correlations, 3) information from different populations can be included to increase accuracy of the genomic predictions, even if some of these populations are measured for different proxy traits, provided there is some genomic relationship amongst the lines in the different populations, 4) the approach can be extended to multiple environments and accounts for GxE if the different environments are treated as different traits. We demonstrate the approach using examples of a speed breeding genomic selection program to improve yield, where proxy traits were measured in the greenhouse and field.

Fast Parallelized Sampling of Bayesian Linear Mixed Models for Whole-genome Prediction

Professor Hao Cheng¹

¹University of California Davis, USA

Session 11, November 6, 2020, 12:00 PM - 1:30 PM

Biography:

Dr. Hao Cheng is an assistant professor of Quantitative Genetics in the Department of Animal Science at University of California, Davis. He got his Ph.D. in Genetics and Statistics at Iowa State University. His research interests are broadly involved in the development of statistical and computational methods for the genetic improvement of populations in agriculture through more accurate and efficient genetic analysis. He has focused on the use of phenomics, genomics, pedigree, and other sources of big data in various species to better predict a wide variety of traits.

Inferences from most Bayesian linear mixed models (BLMM) for genomic prediction are based on Markov chain Monte Carlo methods, where statistics are computed from a Markov chain constructed to have a stationary distribution equal to the posterior distribution of the unknown parameters. In practice, chains of tens of thousands steps are typically used in whole-genome Bayesian analyses, which is computationally intensive. Here we propose a fast parallelized algorithm for BLMM called independent intensive BLMM (II-BLMM, "II" stands for "parallel") and shown how the sampling of marker effects can be made independent within each step of the chain. This is done by augmenting the marker covariate matrix with the number p of markers new rows such that columns of the augmented marker covariate matrix are orthogonal. Ideally, the computations at each step of the MCMC chain can be accelerated by the number k of computer processors up to the number p of markers. We demonstrate II-BLMM algorithm using the prior for BayesCpi, a Bayesian variable selection regression method, applied to simulated data with 50,000 individuals and a medium-density marker panel (about 50,000 markers). To reach about the same accuracy as the conventional samplers for BayesCpi required less than 30 minutes using II-BLMM algorithm on 24 nodes with 24 cores on each node. In this case, II-BLMM algorithm required one tenth of the computation time of conventional samplers for BayesCpi. Addressing the heavy computational burden associated with Bayesian methods by parallel computing will lead to greater use of these methods.

Population scale single cell eQTL mapping identifies cell type specific control of disease

A/Professor Joseph Powell¹

¹Sydney, Australia

Session 11, November 6, 2020, 12:00 PM - 1:30 PM

Biography:

Joseph Powell is the Director of the Garvan-Weizmann Centre for Cellular Genomics and Deputy Director of the UNSW Cellular Genomics Futures Institute at the University of New South Wales in Sydney. He currently holds a National Health and Medical Research Council (NHMRC) Leadership Investigator Fellowship. He was awarded an NHMRC Research Excellence Award in 2016 and the 2017 Commonwealth Health Minister's Medal for Excellence in Medical Research. His research is focused on understanding the functional mechanisms by which genetic variants contribute to disease susceptibility at a cellular level, and ultimately achieve therapeutic and diagnostic outcomes.

The human immune system displays remarkable variation between individuals, which itself leads to differences in how individuals respond to pathogens, and their susceptibility to disease. However, knowledge of how genetic differences contribute to this variation at the level of individual immune cell types has been limited. This is due to the challenges of generating data from a large number of cells in large numbers of individuals, that is required to address these questions.

This talk will cover both conceptual and practical challenges in generating and analysing population scale single cell data. And present results from the OneK1K project, identifying single cell eQTL and their role in contributing to disease risk.

A novel linear mixed model approach for multi-omics data.

Dr S. Hong Lee¹

¹University of South Australia, Australia

Session 11, November 6, 2020, 12:00 PM - 1:30 PM

Biography:

Hong is an associate professor and the leader of the statistical genetics group at the Australian Centre for Precision Health at University of South Australia. He is an ARC Future Fellow. He graduated from Dong-A University in S. Korea, and did his Master and PhD at the University of New England. He has extensive experience in developing advanced statistical methods to estimate genetic variance and individual genetic effects based on phenotype-genotype association analyses. Currently, Hong is focusing on understanding the biological architecture of complex traits by tackling interaction and correlation problems using novel statistical models.

The multi-omics data provide unprecedented opportunity to dissect the genomic, transcriptomic and exposomic architecture of complex traits and diseases. We introduce methods how to estimate the proportion of phenotypic variance and covariance explained by the genome, transcriptome and exposome using a novel linear mixed model (LMM) fitting their effects jointly. Existing conventional LMMs assume independence between random effects, which can cause biased estimates of variance and covariance due to multi-omics' effects on phenotypes. Here, we introduce a new approach relaxing the conventional assumption of independence between random effects, which is particularly useful to dissect multi-omics architecture of complex traits. We also apply a novel genotype-by-environment interaction to estimate the multi-omics' effects modulated by lifestyle or diet. This approach can adjust for confounding from genotype-environment correlation and environment-by-environment interaction. We show that the proposed approach outperforms existing methods using various simulations and real data analyses. We implemented our approach in publicly available software, MTG2 (<https://sites.google.com/site/honglee0707/mtg2>). Our approach and software will be useful in complex trait analyses.

Metagenomics for selection of environmentally friendly ruminants

Dr Melanie Hess¹

¹AgResearch Otago, New Zealand

Session 11, November 6, 2020, 12:00 PM - 1:30 PM

Biography:

Melanie Hess is a postdoc in the Animal Genomics team at AgResearch in New Zealand. She completed her PhD in Animal Breeding and Genetics with a specialization in Quantitative Genetics and minor in Statistics at Iowa State University in 2016 with Dr Dorian Garrick. Melanie's postdoc has been focused on establishing a workflow for rumen microbial profiling using restricted enzyme-reduced representation sequencing and incorporating rumen microbial profiles into predictions of environmentally and economically important traits in livestock. Her interests are focused on modelling complex data in a way that can be practically used in industry.

The rumen microbiome plays an important role in feed digestion and is associated with environmentally and economically important traits such as methane production and feed efficiency. We have profiled over 4,500 sheep and cattle rumen samples using high-throughput, restriction enzyme-reduced representation sequencing. Two bioinformatic pipelines were used to generate rumen metagenome profiles: a reference-based (RB) and a reference-free (RF) approach. The RB compared reads against the Hungate1000 Collection of bacterial and archaeal genomes using BLAST, followed by taxonomic assignment at the genus level. The RF approach counts the occurrence of a set of tags (65 bp reads that are present in at least 25% of samples) within each sample. On average, 20-25% of reads were assigned at the genus level using the RB approach while 40-65% of the reads for each sample were captured in the metagenome profile using the RF approach. Network analyses of rumen metagenomic profiles taken from individuals with extreme phenotypes for methane yield (sheep) or residual feed intake (RFI, cattle) showed segregation of samples by cohort and breed. However, after adjusting for these effects, samples clustered by high and low methane yield (sheep) or RFI (cattle). We also evaluated the impact of species, age, diet and host genetics on microbial profiles. We have used these findings to appropriately incorporate rumen metagenomic profiles into models for animal-level predictions of environmentally and economically important traits. This approach shows the potential to use rumen metagenomic profiles for selection purposes in a practical, agricultural setting.

Can dominance genetic variance be ignored in evolutionary quantitative genetic analyses of wild populations?

Dr Barbara Class¹

¹University of the Sunshine Coast, Australia

Session 12, November 6, 2020, 7:00 PM - 8:30 PM

Biography:

My research situates at the intersection between behavioural ecology and evolutionary biology: I have been studying the behaviour of wild animals from an evolutionary quantitative genetics perspective, using statistical tools such as mixed models and simulations. I am particularly interested in consistent among-individual variation (i.e. 'animal personality') and sexual selection, which I studied in birds during my PhD and first postdoc (University of Turku, Finland), but also social behaviour, disease, and their interaction, which I am now investigating in urban Eastern water dragons.

A substantial part of genetic variance under controlled conditions is due to dominance. Ignoring dominance inflates estimates of additive genetic variance. Whether this issue is pervasive in natural systems is unknown, because we lack estimates of dominance variance in wild populations obtained in situ. Using data on 8 traits measured in over 9000 blue tit nestlings and a genetically resolved pedigree, we estimate additive and dominance genetic variance, common-environment and maternal variance and other sources of non-genetic variance. We find that dominance variance, when estimable, represents only 2-36% of genetic variance, which is a modest amount compared to point estimates from controlled environments. Simulations show that (1) inferences of the above-listed variance components in an average trait measured in this population are unbiased, although power to detect dominance variance is low; (2) ignoring dominance mildly inflates additive genetic variance and heritability estimates. Inflation become substantial when also maternal effects are ignored, which we show to be common. These first estimates of dominance variance in a wild population obtained in situ therefore suggest that dominance is a small source of phenotypic variance, but requires proper model construction to avoid biasing estimates of evolutionary potential in the wild.

Genetic constraints persist through metamorphosis: RNA seq reveals major pleiotropy within and between life stages.

Dr Julie Collet¹, Dr Simon Fellous²

¹Cefe, CNRS, Montpellier, France, ²Cbcp, INRA, Montpellier, France

Session 12, November 6, 2020, 7:00 PM - 8:30 PM

Genetic correlations for traits differentially selected during a life cycle can create antagonistic pleiotropy, where alleles have beneficial effects at a specific time of life and correlated detrimental effects at a different life stage. In most animals, juvenile and adult stages are separated by a major development event: metamorphosis. According to the Adaptive Decoupling Hypothesis - the prevailing adaptive explanation for the pervasiveness of complex life cycles - metamorphosis facilitates the independent evolution of phenotypes expressed at different times in life and therefore maximizes adaptation to their respective environments. Despite the ubiquity of complex life cycles, we know little of the evolutionary constraints exerted or relieved by metamorphosis. Here we measured genetic covariance between life-stages using transcriptomics data on 41 inbred lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP). Almost a quarter of the 6000 gene expression traits that showed genetic variance in larvae and in adults, was significantly genetically correlated between larval and adult stages (22% at the *fdr* level), suggesting strong genetic constraints between life stages. However, an alternative approach was to test whether those 6000 expression traits showed some level of independence. We found that 51% of the correlations between life stages were significantly different from 1 (*fdr* level), suggesting some level of independent expression. Additional functional and multivariate analyses of pleiotropic modules will, for the first time, reveal how genetic constraints apply on life stages separated by metamorphosis in a model species.

The interaction of quantitative genetics and changing environment on a wild bird population

Professor Loeske Kruuk¹

¹Australian National University, Australia

Session 12, November 6, 2020, 7:00 PM - 8:30 PM

Biography:

I am an evolutionary ecologist interested in how evolution works in natural environments, especially considering the role of genetics and the environment in shaping life history variation within populations. I did a PhD in population genetics at the University of Edinburgh, and then a postdoc at the Universities of Cambridge and Edinburgh. I use long-term studies of animal populations (typically vertebrates) to explore quantitative genetics, life history evolution and effects of climate change in wild populations. I took up an Australian Research Council Future Fellowship at the Australian National University, Canberra, in 2012, and an ARC Laureate Fellowship in 2020.

Quantitative genetic analyses can provide valuable insights into the effect of environmental change on wild populations, as well as allowing tests of fundamental evolutionary theory in the wild. I present here results from a 30-year study of a population of a cooperatively-breeding passerine bird, the superb fairy-wren, which has shown a marked decline in population size – possibly driven by warming temperatures. I explore genetic variances and covariances with individual fitness, and show how a quantitative genetic analysis provides insights into causal associations underlying the appearance of phenotypic selection. I also assess the impact of different aspects of the environment – in particular, the effects of changing weather and changing social environment – on quantitative genetic parameters.

Genome-wide chromatin accessibility and transcriptome profiling during onset of maturation in Atlantic salmon

Dr Amin Mohamed¹

¹GEOMAR Helmholtz Centre for Ocean Research, Germany

Session 12, November 6, 2020, 7:00 PM - 8:30 PM

Biography:

Dr Amin Mohamed holds a PhD degree in Coral Reef genomics (2016) from James Cook University (Townsville, Australia). During his PhD, he looked at the molecular basis of coral-algal interactions using transcriptomics. Then he joined CSIRO (Brisbane, Australia) as an OCE postdoctoral fellow to understand epigenetic regulation during onset of sexual maturation in Atlantic salmon. He analyzed multitissue multiomic (transcriptomic and epigenomic) datasets to identify key regulators of salmon maturation. He is interested in genomics, epigenomics and gene regulatory networks in a variety of non-model organisms.

There is currently a great deal of interest in understanding the biological mechanisms driving early maturation in Atlantic salmon due to its negative consequences on both growth and product quality. Hence, we performed an animal trial specifically designed to elucidate the molecular mechanisms associated with onset of sexual maturity. We initiated maturation through photoperiod manipulation and sampled 3 tissues (pituitary gland, ovary and liver) before and after maturation onset. RNA-seq and ATAC-seq were performed to study gene expression and regulation. Multi-tissue transcriptome analysis revealed key players as the earliest triggers in the pituitary gland as well as genes implicated in follicular development and energy homeostasis in both ovary and liver. ATAC-seq data were obtained from ovary and liver to identify differentially accessible chromatin (DAC) regions. Intersecting these DACs with protein coding genes revealed enrichment at intronic and intergenic regions. DACs at genic and proximal regions were significantly positively correlated with gene expression data. In order to assess the regulatory potential in the gained accessible peaks, we identified DNA motifs for transcription factors in the set of proximal and distal regions. Gene regulatory networks were constructed by combining the results from these multi-omics data along with results from genome-wide CpG methylation. These networks revealed key genes and regulators that will potentially pave the way to new strategies to control the timing of maturation in Atlantic salmon.

Does mutation explain standing genetic variation in complex phenotypes?

Dr Robert Dugand¹

¹University of Queensland, Australia

Session 12, November 6, 2020, 7:00 PM - 8:30 PM

Biography:

I am a postdoctoral research fellow at The University of Queensland. I am currently studying how mutation contributes to phenotypic variation in an outbred population.

The distribution of genetic variation across multivariate trait space, described by the genetic (co)variance matrix, G , is uneven as a consequence of pleiotropy generating covariance among traits. While data have accumulated that describe how this unevenness of genetic variation affects phenotypic evolution, we know relatively little about the causes of this bias in G . Comparing the orientation of G and M , the mutational (co)variance matrix, can provide insight both into the relative contributions of mutation and selection to shaping G , and also to long-term evolution. Here, using a large, middle-class neighbourhood breeding design, we accumulated mutations in an outbred population of *Drosophila serrata* for 14 generations. We then utilised Wray's (1990) numerator relationship matrix in the animal model framework to simultaneously estimate G and M for a set of wing traits measured on approximately 37,000 individuals. We found that G and M were largely aligned; trait combinations with high mutational variance also had high standing genetic variance, suggesting that M would stabilise G over evolutionary timescales. However, despite the general alignment of G and M , there was almost no mutational variance in the last eigenvector of M , while G was full rank, suggesting that new mutations on average have stronger pleiotropic effects than standing genetic variants. Our results align with recent analyses of the genetic variation underlying wing shape across drosophilids, and support the hypothesis that genetic variation in these traits is maintained by mutation-selection balance. Wray, N. R. 1990. *Biometrics*, 46, 177-186.

Triangulation of analysis strategies establishes relevant tissues and cell types for complex traits

Mr Zhijian Yang¹

¹*Sun Yat-sen University, China*

Session 12, November 6, 2020, 7:00 PM - 8:30 PM

Biography:

Zhijian is a Ph.D student of biostatistics group in School of Life Sciences in Sun Yat-sen University. His major interests are statistical genetics and quantitative genetics. He is trying to develop novel method to analyze big omics data, such as GWAS summary statistics, RNA-seq data.

Linking complex traits and diseases to their relevant tissues and cell types of humans is of great importance, providing useful etiological and functional insights for understanding the regulatory mechanisms of the traits and diseases. In recent years, different methods have been developed to detect such links by integrating GWAS results and gene expression data of tissues and cells. However, distinct results from these methods generate confusion while no gold standard is currently accepted, making it difficult to evaluate the discoveries. Here, we estimate the operating characteristics of four methods via maximum likelihood in the absence of a gold standard and scored the discoveries using the estimated specificities to provide more specific associations. By integrating GWAS summary statistics with gene expression data and eQTL data from the Genotype-Tissue Expression (GTEx) project, we distinguish good methods from limited power methods and establish tissue-trait associations with higher specificity. Furthermore, three of the four methods are able to take advantage of single-cell RNA sequencing data so that we can establish more specific associations between complex traits and cell types. Our results provide higher confidence in comprehending the underlying tissues and cell types for complex traits, assisting future experimental designs and clinical researches.

Does epistasis matter?

Dr Nicholas Barton¹

¹IST Austria, Klosterneuburg, Austria

Session 13, November 9, 2020, 7:00 AM - 8:30 AM

Biography:

*Nick Barton is a Professor of Evolutionary Genetics at the Institute of Science and Technology, Austria. Nick and his research group focus on the evolution of populations that are distributed through space, and that experience natural selection on many genes. Their research covers a wide range, including a long-term study of a flower colour hybrid zone in snapdragons (*Antirrhinum*), statistical analysis of the effects of selection on sequence variation, and theoretical work on speciation and quantitative genetics. Recent papers include analyses of "islands of divergence" in *Antirrhinum*, of a mouse selection experiment, and of the infinitesimal model for the inheritance of quantitative traits.*

Organisms are not simply the sum of their parts: the effect of an allele depends on its genetic background. However, interactions between genes play little direct role in quantitative genetics: most trait variance can be attributed to the marginal effects of alleles, through the additive genetic variance. Moreover, if traits depend on very many loci, so that alleles are weakly selected, the infinitesimal model applies: then, the genetic background changes slowly, so that the response to selection can be predicted from the initial genetic variance. Arguably, natural selection is most efficient in this regime.

Phantom Epistasis in Genomic Selection: On the predictive ability of epistatic models

Mr Matias F Schrauf¹

¹University of Buenos Aires, Argentina

Session 13, November 9, 2020, 7:00 AM - 8:30 AM

Biography:

Matías Schrauf is a doctoral student at the School of Agriculture, University of Buenos Aires. He is doing research on the statistical properties of the genomic prediction models used in genomic selection for his doctoral thesis, directed by Sebastián Munilla from the school's Animal Breeding Group. He is also teaching assistant at the school's Department of Quantitative Methods.

Genomic selection uses whole-genome marker models to predict phenotypes or genetic values for complex traits. Some of these models fit interaction terms between markers, and are therefore called epistatic. The biological interpretation of the corresponding fitted effects is not straightforward and there is the threat of over-interpreting their functional meaning. Here we show that the predictive ability of epistatic models relative to additive models can change with the density of the marker panel. In more detail, we show that for publicly available Arabidopsis and rice datasets, an initial superiority of epistatic models over additive models, which can be observed at a lower marker density, vanishes when the number of markers increases. We relate these observations to earlier results reported in the context of association studies which showed that detecting statistical epistatic effects may not only be related to interactions in the underlying genetic architecture, but also to incomplete linkage disequilibrium at low marker density (“phantom epistasis”). Finally, we illustrate in a simulation study that due to phantom epistasis, epistatic models may also predict the genetic value of an underlying purely additive genetic architecture better than additive models, when the marker density is low. Our observations can encourage the use of genomic epistatic models with low density panels, and discourages from their biological over-interpretation.

The impact of physiological non-additivity on variance components for complex traits

Dr Kai Voss-Fels¹

¹University of Queensland, Australia

Session 13, November 9, 2020, 7:00 AM - 8:30 AM

Biography:

Dr Kai Voss-Fels is a Senior Research Fellow at The University of Queensland in Brisbane focusing on plant breeding and quantitative genetics. Kai is interested in developing and implementing new genomics-assisted breeding approaches to improve yield, quality and resistances in major crops, such as wheat, barley, sugarcane, rapeseed and chickpea, with a particular focus on the integration of quantitative genetics, genomics and computational approaches. This also includes the study of GxE interaction and non-additive gene action, which represent a source of non-linearity observed in crop breeding, and thereby a key challenge for increasing genetic gain in crops.

For over a century the quantitative genetics community has extended and successfully applied Fisher's infinitesimal model to study many properties of quantitative traits for diverse species. New empirical research findings promote the importance of regulatory networks and gene-gene interactions in determining phenotypes of traits for individuals. Simultaneously at the population level, a large body of empirical data shows that when partitioning total genetic variance into its major components, the additive fraction is larger than the non-additive components. The foundational theoretical work of Cheverud and Routman demonstrated in a two-locus model context that even under strong physiological epistasis, a large fraction of the statistical total genetic variance can be detected as additive. Hill and colleagues made similar observations for outcrossing populations when allele frequencies follow extreme U-shaped distributions. We revisit some of the early work on the connections between physiological and statistical epistasis through the perspective of a crop growth model (CGM) that provides a multi-trait CGM-G2P map. Inspired by recent empirical work on regulatory networks determining plant development we are using simulation to investigate how physiological epistasis is affecting classical variance component partitioning. Our approach is using the hierarchical structure of the CGM-G2P map to investigate the opportunities to improve genomic prediction in crop breeding for complex traits.

Differential complex trait architecture across humans: epistasis identified in non-European populations at multiple genomic scales

Dr Michael Turchin¹

¹Brown University, USA

Session 13, November 9, 2020, 7:00 AM - 8:30 AM

Biography:

Dr. Michael Turchin is a postdoctoral research associate in the Ramachandran Lab at the Center for Computational Molecular Biology at Brown University. Dr. Turchin is a computational geneticist and works on the genetic architecture and evolution of complex traits. He currently works on projects including investigating epistasis on the pathway level in complex traits and multi-ancestry considerations in GWAS approaches. Dr. Turchin received his PhD in Human Genetics from the University of Chicago while working in the lab of Dr. Matthew Stephens and was a NIH NRSA Pre-Doctoral Fellow.

Genome-wide association (GWA) studies have identified thousands of significant genetic associations in humans across many complex traits. However, the vast majority of these studies use datasets of predominantly European ancestry. It has generally been thought that complex trait genetic architecture should be transferable across human ancestries, but recent work has shown a number of differences in trait architecture between populations, including heterogeneity in both identified causal variants and estimated effect sizes. Here, we report further evidence that complex trait genetic architecture is fundamentally different among human ancestries by jointly leveraging pathway and epistasis analysis. Under the assumption that a given complex trait may have differential polygenic architectures across human ancestries, we hypothesize that human populations may also be enriched for differences in epistatic effects. However, since polygenic traits tend to have smaller GWA effect sizes, combining variants via pathway analysis may allow us to better reveal these signals. To accomplish this, we extend the concept of identifying marginal epistasis, moving from testing single variants to testing groups of variants for nonlinear association with a trait of interest. We apply our new method to multiple ancestries present in the UKBioBank. Using morphometric traits we find evidence for genome-wide epistasis in African and other non-European populations. We also find evidence that these trends exist on the SNP and gene levels as well. Results also indicate this may be due to increased heterozygosity in non-European populations. This suggests that non-European populations may be well-suited for identifying non-additive effects in human complex trait architecture.

Dependencies within and among forensic match probabilities.

Professor Bruce Weir¹

¹*University Of Washington, Seattle, United States*

Session 14, November 9, 2020, 12:00 PM - 1:30 PM

Biography:

Bruce Weir is a Professor of Biostatistics at the University of Washington in Seattle. His research focuses on statistical methodology for genetic data, with an emphasis on allelic dependencies, population structure, disease associations and relationships, including applications in forensic genetics.

DNA profiling has become an integral tool in forensic science, with widespread public acceptance of the power of matching between an evidence sample and a person of interest. The rise of direct to consumer genetic profiling has extended this acceptance to findings of matches to distant relatives of the perpetrator of a crime. Along with the greater discriminating power of profiles as forensic scientists have moved from Alec Jeffreys' multi-locus "DNA fingerprinting" to next-generation sequencing, has come the need to re-examine the usual assumptions of independence among the components of forensic profiles. It may still be appropriate to regard variants at a single marker as being in Hardy-Weinberg equilibrium but it is doubtful that all components in a 20-locus STR profile are independent, let alone neighboring sites in an NGS profile. The very basis for forensic genealogy is that human populations contain many pairs of distant relatives, whose profile probabilities are not independent. The expansion of forensic typing to include lineage markers and protein variants raises even further questions of independence. These issues will be discussed and illustrated with forensic and other genetic data, all within a re-examination of the concept of identity by descent and current work to estimate measures of identity within and between individuals, and within and between populations.

Quantitative genetics of environmental variance, uniformity and resilience in livestock animals

Dr Han A. Mulder¹

¹Wageningen University & Research Animal Breeding And Genomics, Wageningen, Netherlands

Session 15, November 9, 2020, 7:00 PM - 8:20 PM

Biography:

Herman (Han) Mulder is Associate Professor at Wageningen University & Research Animal Breeding and Genomics, Netherlands. His main research interests are quantitative genetics of genotype by environment interaction and genetic control of environmental variation in livestock and aquaculture. Han's research focuses on developing statistical methods to estimate genetic variance in environmental sensitivity, either due to known environmental factors, such as temperature, or unknown factors that could be animal specific and appear as differences in within-individual variance or within-family variance. Currently, Han's research is focused at how genetic variance in environmental variance is a measure of genetic variance in resilience when using longitudinal profiles of animals such as daily milk yield in cattle and daily feed intake in various livestock species.

Genetic control of environmental variance (V_e) has been studied for over 75 years, but with the advent of big data and novel statistical approaches our quantitative genetic understanding of heritable variation in V_e has dramatically improved. From a fundamental point of view, heritable variation V_e plays a role in maintenance and evolution of phenotypic variation among individuals. From a livestock perspective, heritable variation in V_e can be exploited to breed for improved uniformity and resilience. The aims of my research are to develop methods to study heritable variation in V_e using longitudinal profiles, to assess V_e as a resilience indicator, to investigate the impact of genotype by environment interaction (GxE) on V_e , and to study the genetic architecture of V_e . Longitudinal data with many records per animal makes genetic analysis of V_e simpler than when having single observations on individuals and increases the heritability from 0-5% towards 10-25%. Genetic coefficients of variation range between 0.2 and 0.6, showing great potential for genetic improvement, although part of the genetic variation in V_e is due to scaling effects. Both experiments and field data point towards that V_e is a promising resilience indicator: selection for low V_e improves health and longevity traits. First results show that V_e is more affected by GxE than levels of traits. GWAS studies show that V_e is a polygenic trait. Big data combined with genomics offer exciting opportunities to study the genetic architecture of V_e and to exploit heritable variation in V_e to improve resilience by breeding.

Increased developmental density decreases the magnitude of indirect genetic effects expressed during agonistic interactions

Prof Chang Seok Han¹

¹*Kyunghee University, South Korea*

Session 15, November 9, 2020, 7:00 PM - 8:20 PM

Biography:

I am an assistant professor at the department of biology at Kyung Hee University, Korea. My main research interests are causes and consequences of individual differences in behaviour.

The expression of aggression depends not only on the direct genetic effects (DGEs) of an individual's genes on its own behaviour, but also on indirect genetic effects (IGEs) caused by heritable phenotypes expressed by social partners. IGEs can affect the amount of heritable variance on which selection can act. Despite the important roles of IGEs in the evolutionary process, it remains largely unknown whether the strength of IGEs varies across life stages or competitive regimes. Based on manipulations of nymphal densities and >3000 pair-wise aggression tests across multiple life stages, we experimentally demonstrate that IGEs on aggression are stronger in field crickets (*Gryllus bimaculatus*) that develop at lower densities than in those that develop at higher densities, and that these effects persist with age. The existence of density-dependent IGEs implies that social interactions strongly determine the plastic expression of aggression when competition for resources is relaxed. A more competitive (higher-density) rearing environment may fail to provide crickets with sufficient resources to develop social cognition required for strong IGEs. The contribution of IGEs to evolutionary responses was greater at lower densities. Our study thereby demonstrates the importance of considering IGEs in density-dependent ecological and evolutionary processes.

Dynamics of secondary traits offer new possibilities for modelling genotype by environment interactions in focal traits

Fred Van Eeuwijk¹, Willem Kruijer¹, Pariya Behrouzi¹, Daniela Bustos-Korts¹, Emilie Millet¹, Jip Ramakers¹, Diana Marcela Pérez Valencia², Coté Rodríguez-Álvarez², Martin Boer¹

¹Wageningen University, Wageningen, Netherlands, ²Basque Centre for Applied Mathematics, Bilbao, Spain

Session 15, November 9, 2020, 7:00 PM - 8:20 PM

Genotype by environment interaction (GxE) is a phenomenon receiving continued interest in genetics and evolutionary biology. In plant biology, popular statistical models for GxE have the form of reaction norms with linear and bilinear terms expressing differential genotypic sensitivity to environmental gradients. The trait for which GxE is modelled is commonly a complex focal trait like yield. New devices for phenotyping of secondary traits together with new sensors for detailed envirotyping produce longitudinal information that has the potential to revolutionize the modelling of GxE in focal traits. GxE for the focal trait follows from interactions over time of secondary traits whose dynamic behaviour is a function of genetic and environmental inputs. We present models for GxE in focal traits that are based on functional and causal modelling of secondary traits. Functional models for secondary traits are defined in terms of P-splines, while causal models for multiple secondary traits were developed as a genetic generalization of the popular PC algorithm for network reconstruction. We demonstrate our approach to modelling GxE on real and simulated data, where the simulated data were generated within an APSIM framework.

Exploiting methylation to measure genome-by-smoking interactions

Dr Carmen Amador¹

¹University of Edinburgh, United Kingdom

Session 15, November 9, 2020, 7:00 PM - 8:20 PM

Biography:

Dr Carmen Amador is a postdoctoral researcher at the MRC Human Genetics Unit. Carmen studied Biology in Universidad Complutense of Madrid and undertook her PhD working in conservation genetics in Dr. Jesús Fernández's group at INIA (Madrid). In 2013 she moved to Edinburgh to work in Chris Haley's group. Her current research interest focuses in complex traits, particularly in the interaction between genetic and environmental causes of variation.

Most health-related outcomes and phenotypes are complex traits influenced by both genetic and environmental variation. For obesity-related traits, hundreds of genetic variants associated with body mass index (BMI) have been identified, and obesogenic environments have been defined. Correctly modelling genetics and environments would allow us to determine whether their contributions to trait variation are the sum of their independent effects or if one modulates the effects of the other (gene-by-environment interactions, GxE). Variation in DNA methylation is correlated with several different environmental variables and has, therefore, the potential to be used as a proxy for these environments without the limitations of answers provided by questionnaires. Here we use methylation variation as a proxy for tobacco usage. We used genotypes of ~500K SNPs and a subset of ~600 methylation CpG sites associated with smoking status, to measure the amount of BMI variation explained by genomic and environmental sources. We analysed differences between self-reported smoking and smoking-associated methylation variation and calculate the amount of variation in BMI explained by genome-by-smoking interactions in the Generation Scotland cohort. We estimated the heritability of BMI to be ~50%. While self-reported smoking status explains 2% of BMI variation, this increases to 22% if measured using smoking-associated methylation. Genome-by-smoking interactions explained an extra 10% of BMI variation (in addition to the independent direct effects of genetics and the environment). This work shows the potential of using omics to measure environmental variation to expand our knowledge on trait architecture and potentially improve trait prediction.

From pedigree to genomics based management of genetic diversity: how to measure and control genomic inbreeding in genomic selection schemes

Dr Theo Meuwissen¹

¹Norw. Univ. Life Sci., As, Norway

Session 16, November 10, 2020, 7:00 AM - 8:25 AM

Biography:

Theo Meuwissen is a Professor at the Norwegian University of Life Sciences. His work focuses on uniting quantitative genetic theory with genomics technology, revolutionizing the genetic improvement of livestock and crops.

Livestock and crop breeding schemes are currently accelerating their rates of genetic improvement by genomic selection. Large rates of genetic change in the genome may however jeopardize the genetic diversity of the population, and hence future opportunities for genetic change. Moreover, large rates of genetic drift may result in large changes in trait values, even for traits that are currently near optimal. Here, we explore genomic prediction methods combined with the management of genomic diversity in breeding schemes that make optimal use of genomic data. The definition of pedigree based inbreeding, i.e. the probability of identity by descent at neutral loci not linked to any loci under selection, has lost its practical relevance in modern genomic selection schemes since such loci do not exist when selection is based on genome-wide dense marker panels. Hence, the goals of diversity management will be revisited and casted in molecular genetic terms, in order to find alternative definitions of inbreeding with (more) practical relevance. Alternative molecular genetic definitions of inbreeding will be investigated w.r.t. how they are measured, their properties, whether they attain the goals of diversity management, and for their consequences on the rates of genetic improvement. This results in general recommendations on the optimal combination of genetic gain and diversity management in the era of genomics.

From Statistical Models to Biological Mechanisms of Human Mutation

Professor Shamil Sunyaev¹

¹Brigham and Womens Hospital/ Harvard Medical School, United States

Session 16, November 10, 2020, 7:00 AM - 8:25 AM

Biography:

Shamil Sunyaev is a Professor of Computational Genomics Medicine at Harvard Medical School. He obtained a PhD in molecular biophysics from the Moscow Institute of Physics and Technology and completed his postdoctoral training in bioinformatics at the European Molecular Biology Laboratory (EMBL). The primary focus of research in his lab is genetic variation, including the biology and evolution of mutation, the effect of variation on molecular function and structure, population genetics as a lens on evolution, and the maintenance and allelic architecture of complex traits.

Mechanistic processes underlying human germline mutations remain largely unknown. Variation in mutation rate and spectra along the genome is informative about the biological mechanisms. We statistically decompose this variation into separate processes using a blind source separation technique. The analysis of a large-scale whole genome sequencing dataset (TOPMed) reveals nine processes that explain the variation in mutation properties between loci. Seven of these processes lend themselves to a biological interpretation. One process is driven by bulky DNA lesions that resolve asymmetrically with respect to transcription and replication. Two processes independently track direction of replication fork and replication timing. We identify a mutagenic effect of active demethylation primarily acting in regulatory regions. We also demonstrate that a recently discovered mutagenic process specific to oocytes can be localized solely from population sequencing data. This process is spread across all chromosomes and is highly asymmetric with respect to the direction of transcription, suggesting a major role of DNA damage.

Leveraging mathematical optimization to drive short-term gains while maintaining long-term genetic variability in a plant breeding program

Dr Nicholas Santantonio¹

¹Cornell University, USA

Session 16, November 10, 2020, 7:00 AM - 8:25 AM

Biography:

I am the new PI of a team of research scientists, staff and students that comprise the small grains breeding program at Virginia Tech in Blacksburg VA. My interests include the quantitative genetics of agronomic traits and the integration of genomic selection and high-throughput phenotyping into a working breeding program. I am currently looking for PhD students and Postdocs, so please reach out if interested!

Genomic selection in plant breeding programs must be adapted to an already complex system. Inbred or hybrid plant breeding programs must make crosses, produce inbred individuals, and phenotype the inbred lines or their hybrid testcrosses to select and validate superior material that may be released as products. These products are few, and it is not clear that improvement of the population per se is of interest as it is in animal breeding applications. Rapid cycle recurrent genomic truncation selection has been proposed to increase genetic gain by reducing generation time. This strategy has been shown to increase short-term gains, but can quickly lead to loss of genetic variance through inbreeding as relationships drive prediction. Mathematical programming can determine the optimal contribution of each individual to the next generation to maximize gain while holding inbreeding at a predefined level. While optimal contribution strategies do well at maintaining genetic variance in later generations, they suffer from a lack of short-term gains. We present a hybrid approach that branches out yearly to push means for material destined to be phenotyped, while maintaining genetic variance in the recurrent population, such that a breeding program can achieve short-term success without exhausting long-term potential. Because branching increases the genetic distance between the phenotyping pipeline and the recurrent population, this method requires sacrificing some trial plots to phenotype materials directly out of the recurrent population. We envision the phenotypic pipeline not only for selection and validation, but as an information generator to maximize the output of products.

Understanding the polygenic architecture and regulatory mechanisms of human complex traits

Prof Jian Yang¹

¹Westlake University, Hangzhou, China

Session 17, November 10, 2020, 12:00 PM - 1:25 PM

Biography:

Jian Yang is a Professor of Statistical Genomics at the Institute for Molecular Bioscience, The University of Queensland (UQ). He received his PhD in 2008 from Zhejiang University, China, before undertaking postdoctoral research at the QIMR Berghofer Medical Research Institute in Brisbane. He joined UQ in 2012. His primary research interests are in developing novel statistical methods to better understand the genetic architecture of complex traits and diseases, to identify putative target genes, and to improve the accuracy of genomic risk prediction using high-throughput genetic and genomic data. He has led the development of QG analysis tools such as GCTA, SMR and OSCA. His recent papers have been on causal inference using GWAS summary data, and an extremely resource-efficient GWAS tool for large-scale data.

Most common traits (including diseases) in humans are influenced by many genetic variants with small effects. This pattern of genetic architecture is often referred to as the polygenic model. Understanding the causes of polygenic architecture and the molecular mechanisms underlying the polygenic effects have important implications in human health and evolution. In this talk, I will discuss the use of mixed linear models that fit the effects of all genetic variants available in a genome-wide association study (GWAS) as random effects to estimate genetic architecture parameters for a complex trait (e.g., the overall contribution of the polygenic effects to the trait, and the proportion of variants with non-zero effects), and the use of genetic architecture parameters to infer the evolutionary causes of polygenic architecture. I will then demonstrate methods and analyses that integrate data from GWASs of complex traits and other molecular phenotypes (e.g. gene expression and DNA methylation) to predict putative causal genes and regulatory mechanisms underpinning the polygenic effects on complex traits.

Implementation of Genomic Selection in the CIMMYT Global Wheat Program, learnings from the past 10 years

Dr Susanne Dreisigacker¹

¹CIMMYT, Mexico

Session 17, November 10, 2020, 12:00 PM - 1:25 PM

Biography:

Over the past ten years, Dr Susanne Dreisigacker has been leading the Wheat Molecular Breeding Laboratory at the International Wheat and Maize Improvement Center (CIMMYT) in Mexico. Dr Dreisigacker's main interest is to bring promising genomics tools into use in the CIMMYT Global Wheat Program. Her lab conducts genetics studies on key biotic (foliar diseases), abiotic (yield and yield stability) and quality factors that limit the production and value of wheat, especially in the developing world. She provides the Global Wheat Program access to state-of-the-art molecular technologies and together with the CIMMYT's wheat breeders she applies marker-assisted selection strategies such as MAS, MABC and genomic selection targeting to accelerate genetic gains in bread and durum wheat. Dr Dreisigacker's research team collaborates with scientists around the world, including teams in India, China, Middle East, UK, USA, Germany, and South Africa. She trains and supervises students coming from local universities or internationally through CIMMYT's intensive wheat training program.

In the last decade, genomic selection (GS) has been implemented in a variety of species, with particular success in animal breeding. The International Center for Maize and Wheat improvement (CIMMYT) leads the Global Wheat Program, whose main objective is to increase the productivity of wheat cropping systems to reduce poverty in developing countries. Multi-trait, multi-environment GS has been applied in the Global Wheat Program for several years to reduce phenotyping costs and accelerate genetic gains. Important genetic-statistical complexities to increase GS accuracy have been explored, e.g. incorporating pedigree information, G x E interactions that create trait and environmental structures or involving high-throughput phenotyping data. Predicted values have shown to be particularly useful to assist in the selection for quality traits, disease resistance but also grain yield in preliminary and elite yield trials. An additional critical challenge for the wheat breeders in the program is to identify suitable parents for creating the genetic variation to maximize selection response in subsequent breeding cycles. Parental-selection strategies have been initially tested and will be the focus in up-coming years when combined with rapid generation advance.

Indirect genetic effects on biomedical phenotypes measured in laboratory mice and rats

Dr Amelie Baud¹

¹European Bioinformatics Institute, USA

Session 17, November 10, 2020, 12:00 PM - 1:25 PM

Biography:

Dr. Amelie Baud did her PhD at the University of Oxford where she studied the genetic basis of complex traits in laboratory rats. She is now a Sir Henry Wellcome postdoctoral fellow at the European Bioinformatics Institute, where she leads an independent research program into indirect genetic effects - genetic effects on an individual's phenotype that arise from genotypes of social partners - focusing on biomedical phenotypes measured in laboratory mice and rats. She was a Visiting Scholar at the University of California San Diego for three years, where she carried out mouse experiments and started working on host-microbiome interactions

Indirect genetic effects (IGE) are genetic effects on the phenotype of an individual that arise from genotypes of social partners. IGE contribute to phenotypic variation in a broad range of species, yet there has been relatively little research into IGE in the biomedical field. As a result, the impact of IGE on human health and on biomedical phenotypes measured in rodent models is largely unknown, as are the underlying mechanisms. To address these questions, I studied behavioural, physiological and morphological phenotypes measured in genetically heterogeneous laboratory mice and rats, focusing on IGE arising between cage mates. I developed computational methods to model "classical" direct genetic effects and IGE jointly, which I used for variance decomposition and for the genome-wide association study of IGE. My results show that IGE affect a broad range of phenotypes in laboratory mice and rats, in some cases substantially, including anxiety, immune status, body weight, LDL levels, and wound healing. They also provided insights into the mechanisms of IGE: first, I demonstrated that phenotypic "contagion", whereby a phenotype of interest is affected by the *same trait* of social partners, is not generally sufficient to explain IGE. Secondly, I demonstrated that the genome-wide association study of IGE can identify genes and pathways underlying IGE, and validated the role of the gene *Epha4* in giving rise to IGE on stress-coping strategies and wound healing.

Extracellular vesicles with specific surface proteins are associated with waist circumference and visceral fat

Miss Ranran Zhai¹

¹*Sun Yat-sen University, China*

Session 18, November 10, 2020, 7:00 PM - 8:30 PM

Biography:

Ranran Zhai is a Ph.D. student in the Biostatistics lab of School of Life Sciences, Sun Yat-sen University where she got her Bachelor's degree in Biology in 2019. She is interested in quantitative genetics and human complex traits. She is currently working on the discovery of extracellular vesicles as biomarkers by integrating omics data.

The field of genomics has spent tremendous efforts trying to discover the genetic regulation and mechanisms underlying complex traits and diseases. While many associations have been found, our understanding remains far from satisfactory, due to the high complexity of genetic architecture. Omics techniques provide an opportunity to look at the problem with better resolution, e.g. integration of genome-wide association results and single-cell omics information can sometimes give us a clue in which cells our diseases may develop. As a proxy of cell-level biology, extracellular vesicles (EV) have become popular candidate biological complexes to study the source of cell regulation of complex diseases. EVs carry a lot of biological information and are largely enriched in human plasma. Here, we utilized a novel technology to detect the presence of 120 candidate proteins across millions of single EVs. By integration with GWAS summary statistics, we identified combinations of coding genes for the EV surface proteins being associated with obesity-related traits such as waist circumference. We subsequently verified such associations by quantifying these EVs with the particular protein profiles and testing their associations with body fat measured by DEXA scans in 96 individuals from Orkney. We found that the lower abundance of EVs that carry both ITGB6 and ITGB8 indicates larger waist circumference, as well as more waist visceral fat. Our findings provide the first evidence that EVs with specific surface proteins are associated with obesity, suggesting visceral fat can be tested using plasma and shedding light on future EV biomarker discovery.

Dissecting genetic variation in gene expression at the single-cell resolution in *Arabidopsis thaliana* root tips.

Dr Adam Reddiex¹

¹*Australian National University, Australia*

Session 18, November 10, 2020, 7:00 PM - 8:30 PM

Biography:

*Adam is a postdoctoral fellow at The Australian National University. Adam received his PhD in genetics and evolutionary biology from The University of Queensland in 2019. During his PhD, Adam studied the constraints on the evolution of sexually selected traits in *Drosophila* using whole-genome sequence data. Currently, Adam works on understanding genetic variance in gene expression of *Arabidopsis* root tips using single-cell technologies.*

Expression quantitative trait loci (eQTL) mapping studies are concerned with finding variants in the genome that underlie patterns of gene expression. However, the tissue sampled in such studies often contain a heterogeneous mix of different cell-types and it is unclear to what degree the effects of genetic variants are consistent across cell-types within tissues. To address this, we used single-cell RNA sequencing to measure transcriptome-wide gene expression of root tips in 150 accessions of *Arabidopsis thaliana*. In this presentation we present the results of eQTL style analyses conducted in a cell-type specific manner.

A genomic meta-analysis of 184 neuroproteins and their implied causality on psychiatric disorders

Miss Linda Repetto¹

¹University of Edinburgh, United Kingdom

Session 18, November 10, 2020, 7:00 PM - 8:30 PM

Biography:

Linda Repetto is currently a fourth year PhD student at the Usher Institute, University of Edinburgh. She studied at University College Maastricht, the Netherlands, where she took courses ranging from genetics, to psychology and neuroscience. She completed her Master's by Research degree in Integrative Neuroscience at the University of Edinburgh in 2016. Her research interests include - but are not limited to - psychiatric disorders, mental health, and cardiovascular conditions.

Protein quantitative trait loci (pQTL) are essential to study the molecular basis of complex diseases, as they provide insights on the role of genetic variation in determining protein levels that modulate an individual's metabolic state. We quantified 184 proteins involved in neurological processes using the Olink Neurology and Neuro-exploratory panels in 1070 individuals of the Orkney Complex Disease Study (ORCADES) with genotypic information. For each protein, we performed a genome-wide association analysis (GWAS) looking for loci associated with protein levels, both in the proximity of the protein-coding gene and distantly. We discovered 48 cis- and 59 trans-pQTL for 95 of these 184 neuroproteins. Enrichment analyses on the levels of expression of these proteins in different tissues show that the neuroproteins with cis-pQTL are expressed in the brain. We subsequently investigated the potential causal effect of protein level variation on psychiatric disorders for the proteins with cis-pQTL. Using our pQTL study and summary-level data from a large 2018 GWAS for major depressive disorder (Wray et al., 2018), we discovered 12 additional loci associated with the disease. Two of these 12 loci were also discovered in the latest 2019 GWAS meta-analysis for major depression (Howard et al., 2019), strengthening our novel results, that highlighted loci through neurological pQTL with potentially relevant roles in psychiatric disorders. With these promising results, we set up a meta-analysis of levels of the same proteins in seven cohorts with a maximum sample size of 12,000 individuals to reveal more biology underlying neuropsychiatric disorders.

Chromatin and epigenomic variation reveals the gene regulatory landscape of adaptive divergence in sticklebacks

Felicity Jones¹

¹Max Planck Society, Germany

Session 18, November 10, 2020, 7:00 PM - 8:30 PM

Biography:

Felicity Jones is a Research Group Leader at the Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany. Her research investigates genome function in naturally evolving populations with a focus on molecular mechanisms underlying adaptation and speciation. Her work has contributed to our understanding that adaptation in vertebrates proceeds predominantly through mutations in the non-coding genome (regulatory mutations, rather than protein-coding changes). Her team are functionally dissecting adaptively divergent regulatory elements using comparative epigenomics, transcriptomics and transgenic approaches. They leverage the unique advantages of stickleback fish – a classical model of evolution at the molecular and functional genetics level.

The adaptation of natural populations to changing environments is often driven by numerous genetic loci predominantly found in non-coding regions with likely gene regulatory roles. Using adaptively diverging marine and freshwater stickleback fish ecotypes as a model, we performed comparative epigenomics, chromatin profiling, transcriptomics and genetics to identify thousands of regulatory elements with divergent epigenomic profiles between the ecotypes. Divergent elements are enriched at the promoter and 5'UTR of genes, are proximal to genes showing differential expression, and vary across tissues, with the liver showing considerably higher regulatory divergence than kidney or gills. Allele-specific analyses in F1 hybrids reveals that divergence in chromatin accessibility is mostly cis-regulated and these elements show molecular signatures of natural selection. Additionally, divergent epigenomic marks cluster into 'islands' of genetic differentiation and low recombination, including chromosomal inversions. We show through functional transgenic assays how these cassettes act as hubs to cause concerted changes in gene expression between adaptively diverging populations. The high resolution maps of the chromatin and epigenomic landscape in diverging stickleback ecotypes provides functional annotation of regulatory elements within adaptive loci. Our study shows how cis-regulated chromatin variation and epigenomic marks at regulatory elements is associated with adaptive divergence and the early stages of speciation, and links their co-inheritance as adaptive regulatory cassettes to the fast and repeated adaptive radiation of sticklebacks.

Whole exome sequences reveal rare and common variants associated with 1102 plasma proteins

Dr Lucija Klaric¹, Dr Thibaud S. Boutin¹, Dr Cristopher V. Van Hout², Dr Nehal Gosalia², Dr Alan R. Shuldiner², Professor Chris Haley¹, Professor James F. Wilson^{1,3}

¹MRC Human Genetics Unit, IGMM, University Of Edinburgh, Edinburgh, United Kingdom, ²Regeneron Genetics Center, Tarrytown, United States of America, ³Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

Session 18, November 10, 2020, 7:00 PM - 8:30 PM

Biography:

Lucija earned her PhD in human genetics at the University of Edinburgh, where she studied genetic regulation of protein glycosylation. Working as a research fellow at the University of Edinburgh and data analyst in Genos Ltd on the project "Methods for Integrated Analysis of Multiple Omics datasets", she gained insights both in bottom-up pre-processing and analyses of glycomics data, but also in integrating these with genomics and clinical data. She has a general interest in data integration of different omics and has recently been appointed to a UKRI innovation fellowship in data science to work on linking whole genome and exome sequencing data with different omics in the context of complex traits and diseases.

Proteins are major drivers of all biological processes. Blood plasma proteins reflect the proteomes of a wide variety of cells, providing an insight into physiological processes occurring in diverse biological systems. Therefore, they are potentially useful biomarkers of various diseases. Combining proteomics with protein-altering variation captured by whole exome sequencing provides an opportunity for a more direct insight into mechanisms behind complex traits and diseases. We performed exome-wide association analysis of 1,102 proteins measured using OLINK proximity extension assays in plasma from 1059 individuals from the genetically isolated ORCADES cohort (Scotland) and whole exome sequencing (WES) generated by the Regeneron Genetics Center. We found 4,483 significant ($p \leq 4.5 \times 10^{-11}$) associations between 426 proteins and 358 genes. 22% of associations were with low frequency variants (MAF < 0.05). Of the 426 proteins, 57% had cis associations only, 43% trans only and 94 (22%) were associated with two or more genomic loci. Overall, 179 sentinel variants are missense, with 6 resulting in gained stop codon. Association of very rare variants was assessed using aggregate tests and revealed further loci influencing the proteome. Given the direct role of proteins in various biological processes and more interpretable functional consequence of WES variants, genetic associations of a large number of proteins and protein-altering variants provide not only a basis for understanding mechanisms behind complex traits and diseases, but also provide a framework for assessment of causality of individual biomarkers.

Genomic prediction of complex phenotypes in livestock – potential applications and lessons to be learned from crops and model species

Professor Henner Simianer¹

¹University of Goettingen, Germany

Session 18, November 10, 2020, 7:00 PM - 8:30 PM

Biography:

Henner Simianer is Professor of Animal Breeding and Genetics at the University of Goettingen, Germany. Being a quantitative geneticist by training, he has a special interest in the inheritance of complex traits and works on tools to design and optimize breeding programs, for which his group has recently provided a novel concept and software for modular breeding program simulation (www.mobps.de). Over the last years, he has expanded his research interest towards plant breeding and has been the founding Director of the Center of Integrated Plant and Animal Breeding (CiBreed), which aims at benefiting from synergies between crop and livestock breeding.

While genomic prediction of breeding values has revolutionized livestock breeding, little attention has been given so far to the prediction of complex phenotypes. Accuracy of phenotype prediction, defined as correlation between predicted and true phenotypes, is limited to the square root of the broad sense heritability, if only genomic information is used. However, it can exceed this level if information on environment, other omics layers or early phenotypes is taken into account as well. We showed in a series of empirical studies with crop and model species data that (i) using prior biological knowledge for feature selection improves phenotype prediction in various model species, (ii) accounting for the top 1 to 5 percent of all biallelic SNP interactions can improve phenotype prediction substantially, and (iii) using expression profiles can improve phenotype prediction accuracy, provided that the tissue and developmental stage chosen for RNA profiling is relevant for the target phenotype. While these results demonstrate some potential to improve accuracy of complex phenotype prediction in livestock, its usefulness appears to be limited to prediction of hybrid performance and production optimization by informed replacement and mating decisions. Long term genetic progress in complex traits remains largely driven by selection based on genomic prediction of breeding values.

Matching Genetics to Environment Using Genomics: Synthesis of results from USDA-NIFA Food Security Grant on Local Adaptation in Beef Cattle

Dr Jared E Decker¹

¹University of Missouri, USA

Session 19, November 11, 2020, 7:00 AM - 8:30 AM

Biography:

Jared E. Decker is an associate professor in the University of Missouri Division of Animal Science, Genetics Area Program, and Institute for Data Science and Informatics. Decker earned his Ph.D. at the University of Missouri in Genetics, with a Ph.D. minor in Statistics. His research combines evolutionary biology with quantitative genomics to better understand the history of populations and create new selection tools for cattle producers. He is active on social media and his blog A Steak in Genomics.

Cattle poorly adapted to their environment lose revenue and jeopardize sustainability of the food supply. Genomic data allows us to rigorously analyze local adaptation and genotype-by-environment interactions. We blend quantitative genetics, population genetics and genomics to assess environmental stressors. We use local adaptation selection scans, genotype-by-environment genome-wide association analyses, creation of hair shedding genomic predictions and environmental region-specific genomic predictions of growth traits to predict local adaptation in beef cattle. Analyzing ~40,000 cattle from three breed associations with ~850,000 high-accuracy imputed SNPs, we used novel selection mapping methods to identify genomic loci responsible for adaptation. Among the three data sets, we identify 114 genes responding to local adaptation selection. In cooperation with 74 producers, over 12,000 cattle were scored on a scale of 1-5 for the hair shedding phenotype over four years. Genomic breeding values were generated with a repeated records model using these phenotypes. Further, we identified loci with large allele substitution effects for hair shedding. When genotype-by-environment interactions exist, ranking animals using a regional genetic evaluation will be different from national cattle evaluations. We developed region-specific genomic predictions using a multivariate model in which phenotypes from different regions were fit as separate dependent variables. Genetic correlations between regions were moderate, indicating substantial re-ranking between environmental regions. Genomic loci with a large effect in one region may have little effect in a different region. Across all analyses, genes related to immunity and the circulatory system were enriched. Prototype region-specific genomic predictions will identify cattle best suited to an environment.

Lifting the lid of the rumen: genomic solutions for sustainable livestock production

Dr Suzanne J Rowe¹

¹Agresearch, New Zealand

Session 19, November 11, 2020, 7:00 AM - 8:30 AM

Biography:

Suzanne Rowe is a quantitative geneticist working on the development of genomic tools for improvement of productivity, profitability and sustainability in livestock production systems. She has a particular research interest in the use of host genomic information and rumen microbial sequence data for the prediction of enteric methane emissions, feed efficiency and productivity in pastoral based systems.

Divergent sheep selection lines for enteric methane emissions have also been shown to have been bred for divergent rumen microbiomes. These changes are postulated to have resulted in a cascade and chain of physiological outcomes that are informative for many aspects of ruminant production. These include smaller rumens, longer and denser papillae, and ultimately a change in the volatile fatty acids or energy source to the animal. Further research has shown that the fatty acid profile of meat and milk may also be affected, but it is unknown whether this is an indirect effect of changes in precursors from the rumen or the direct result of differences in host genetics. This change in fatty acid profiles suggests that not only rumen microbiomes but milk and meat profiles are potential predictors of methane. Furthermore, finding major host genes that control methane has been extremely difficult, whereas finding associations with interim phenotypes such as rumen, muscle, meat and milk physiology has a much greater chance of success. Here we look at major host effects associated with changes in physiology associated with breeding for divergent microbiomes. We compare the accuracy of prediction of methane and other energy related traits using direct genomic predictions, and genomic predictions using intermediate phenotypes including host microbial profiles. Finally, we explore what this may mean for inclusion of methane in genomic selection indices if genetic effects on yield and product quality are included as well as a “carbon” cost for methane emissions.

Scanning the Angus genome for homozygosity deficient haplotypes

Dr Duc Lu¹

¹Angus Genetics Inc, USA

Session 19, November 11, 2020, 7:00 AM - 8:30 AM

Biography:

Dr Duc Lu is a research geneticist at Angus Genetics Inc., American Angus Association. He does research in beef cattle genomics.

A lethal allele in homozygosity causes an animal to die before reproduction age, while homozygosity of semi-lethal alleles might allow some animals to reproduce or perform at a reduced level. Detecting such alleles might be beyond the capability of low/medium density SNP arrays. Nevertheless, a variant could be in linkage disequilibrium with surrounding genes, making it possible for detecting lethal or semi-lethal haplotypes. The exercise herein attempted to use 44,818 SNP that are common among various SNP arrays (AngusGS, HD50K, whereas GGPLD, GGPHD, and i50K were imputed to 50K) to identify potential lethal and semi-lethal haplotypes from Trio families (both parents and progeny genotyped) or PGP families (parent, grandparent and progeny genotyped). A group of 567,164 genotyped Angus cattle were used, including 202,381 Trios, and 202,584 PGP. Haplotypes of interest were detected by using sliding windows (1 SNP slide) of sizes 3 – 72 SNP (maximum 3.7Mb) and haplotype frequency tests. There were 108,281 overlapping haplotypes with frequency in the Trio and PGP progeny significantly less than expectation ($P < 0.01$), and the number of homozygous progeny less than expectation. They belonged to 184 regions across 29 chromosomes. Of those haplotypes, 59,542 were found to exist only in heterozygosity while the expected number of homozygous individuals ranged 5 – 243. A ~5.5Mb region on chromosome 11 harbored 990 of such haplotypes with frequencies 1.83-2.07%, and expected homozygosity in 189-243 progeny though no homozygosity was found. The impact of these haplotypes needs investigation for association with phenotypes, e.g. heifer pregnancy and calf survival.

Non-additive effects in dairy cattle

Mr Edwardo Reynolds¹

¹Massey University, New Zealand

Session 19, November 11, 2020, 7:00 AM - 8:30 AM

Biography:

Edwardo Reynolds is a PhD student at the AL Rae Centre of Genetics and Breeding at Massey University. He works in molecular and quantitative genetics, and his research focuses on using genome-wide association studies to improve our understanding of non-additive genetic mechanisms in cattle.

Additive genetic effects have been extensively explored and research findings have been used in selection programs. However, intensive selection has led to inbreeding depression in many livestock populations, likely due to non-additive genetic effects. This work is focused on identifying genetic variants underlying dominance and recessive mechanisms in dairy of cattle. Recessive mechanisms in livestock and humans have primarily been explored under a forward targeted-mapping paradigm such that a phenotype, often a qualitative syndrome, has been identified and used to map the responsible genomic locus. Hypothesis-free methods such as genome scans have been used to detect embryonic lethal variants, often caused by loss-of-function mutations. The existence of extreme non-additive genetic effects raises the hypothesis that non-additive genetic effects might cause variation in quantitative phenotypes such as body weight. We undertook genome-wide association studies based on conventional linear mixed model approaches extended to make inference on non-additive effects. We utilised a Bayesian Gibbs sampler framework to account for population structure. We assessed some 17 million variants throughout the genome for non-additive effects on over 80,000 genotyped Holstein-Friesian, Jersey, and admixed dairy cattle for production, and developmental traits. We identified variants with non-additive effects on cattle characteristics.

Evaluating the accuracy of imputed whole-genome sequence data in admixed dairy cattle

Dr Yu Wang¹

¹Massey University, New Zealand

Session 19, November 11, 2020, 7:00 AM - 8:30 AM

Biography:

Yu Wang is a postdoctoral scientist in quantitative genetics and genomics at Livestock Improvement Cooperation (LIC), New Zealand. She holds a Ph.D. in animal breeding and genetics from the University of Hohenheim, Germany. Before she joined LIC, she worked as a post-doc researcher at Massey University on the "Better Breeding Values" project under Genomics Aotearoa, which involved further development and implementation of a pipeline that comprises whole genome sequence imputation, sequence level genome-wide association studies, followed by single-step prediction using sequence imputed variants. Her main research interests are quantitative genetics, genomic selection, genetic diversity and conservation genetics.

The imputation accuracy was evaluated when high-density SNP genotypes (~644K) from an admixed population of 165,364 New Zealand dairy cattle were imputed to whole-genome sequence (WGS). WGS data were available for 336 Holstein-Friesian (H), 174 Jersey (J) and 535 H×J crossbred animals, among which 603 were sequenced with read depth >10x. The raw WGS data were aligned to the ARS-UCD1.2 bovine reference genome and variant calling was performed using GATK. VQSR filtering and standard quality control processes were conducted, resulting in ~20 million variants across the genome. Either high-depth sequenced animals or all sequenced animals were used as the reference population. The following software were singly or jointly used for phasing: Beagle 4.1, LinkPhase 3 and Beagle 5.0. Imputation was then performed from the phased data using Beagle 5.0. The quality of the imputation on chromosome 5 was evaluated by comparing the average dosage R², or based on 248 imputed animals sequenced for validation. Our study demonstrated that using Beagle 5.0 for phasing and imputation achieved high accuracy (average dosage R²=0.905) via using all sequenced animals and the reliable variants determined by high-depth sequenced animals. The sequence data from 248 validation animals exhibited an error rate of 1.11%, and correlation between imputed and called variants of 98.8%. Beagle 4.1 pre-phasing using genotype likelihood as input brought marginal benefit, however, it may be beneficial when low-depth sequenced animals were included in the reference. The imputed dataset will be used for future genome-wide association studies for casual variant detection and genomic selection.

Trait correlation and accuracy weighted selection index emphasis

Dr Luna Zhang¹

¹AbacusBio Ltd, New Zealand

Session 19, November 11, 2020, 7:00 AM - 8:30 AM

Biography:

Xinyue (Luna) Zhang is a consultant at AbacusBio Ltd, New Zealand. Her work mainly involves selection index, quantitative genetics&genomics, app and pipeline development with a focus on industrial application. Previously, Luna worked in BioNano Genomics Inc., San Diego as a biostatistician and bioinformatician, developing algorithms for their structural variation detection pipeline on optical mapping platform. Luna also interned in Cobb-Vantress Inc. developing EBV evaluation and GWAS pipeline for broilers. Luna graduated from University of Georgia with a PhD degree in quantitative genetics and breeding in 2015. Her advisor was Dr Ignacy Misztal. Her most cited paper is weighted ssGBLUP.

The current international standard methodology to quantify trait emphasis in selection indexes is a simple multiplication of the relative contribution of each trait's economic value and its genetic standard deviation. This method does not reflect the actual selection emphasis applied with the use of the index considered, because the economic value is not a direct reflection of selection operation when traits differ considerably in their accuracy of evaluation, and the summation of genetic standard deviations does not account for the correlations among traits. We propose a new emphasis method adjusted by both accuracy and genetic correlation. Briefly, we group genetically highly correlated traits, calculate each trait's sub-index emphasis within its sub-group, then weight each sub-group by their total emphasis. The sub-group could be clustered by applying hierarchical clustering method on genetic correlation matrix. When applied to New Zealand sheep breeding selection index, the new method shrank the emphasis on survival traits from 50% to 10%; and expanded that on growth traits from 30% to 55%, better reflecting the trait responses achieved in reality. The accuracy affected traits' within-sub-group ranking whereas the clustering affected each sub-group's emphasis. The new method also alleviates extreme weighting towards a single trait. When genetic correlations across traits were consistent, the emphases calculated by the new and the old methods were similar.

Complex-Trait Prediction using Sparse Selection Indices

Dr Gustavo de los Campos¹, Mr Marco Lopez-Cruz¹

¹Michigan State University, East Lansing, United States

Session 20, November 11, 2020, 12:00 PM - 1:25 PM

Biography:

Gustavo de los Campos was born in Uruguay where he received an Agronomy degree from Universidad de la República, and later on, worked in research at the Uruguayan National Institute for Agricultural Research (INIA-Uruguay). In 2003 he migrated to the US to pursue graduate studies at the University of Wisconsin-Madison where he completed two MS degrees and a PhD, under the supervision of Prof. Daniel Gianola. In 2010 he joined the Biostatistics department of the University of Alabama at Birmingham (UAB) as a postdoctoral fellow first (2010, working under the supervision of Prof. David Allison) and then as a Faculty (2011-2015). He obtained tenure at UAB in 2014 and in 2015 he joined Michigan State University where he is now Professor in the Departments of Epidemiology & Biostatistics, and Statistics & Probability.

Genomic prediction (GP) has been adopted in plant and animal breeding and has also gained ground in human genetics. GP models were initially developed with reference to homogeneous populations. However, many genomic data sets originate from genetically heterogeneous populations where SNP effects cannot be assumed to be homogeneous across subjects. Multivariate Gaussian models and SNP-by-group interactions methods can be used to model group-specific SNP-effects. However, these approaches are not adequate for data sets involving complex forms of genetic heterogeneity (e.g., varying degrees of admixture) where individuals do not cluster into clearly-defined disjoint genetic groups. To address the limitations of existing methods, we developed a model that identifies, for each subject in a prediction set, an optimal support set in the training data which is used to derive predictions. Our approach integrates selection index methodology with sparsity-inducing L1-penalization. The resulting method is a sparse 'selection' index (SSI), which can also be viewed as a sparse version of GBLUP. The methodology can be used to derive genetically accurate phenotypes from high-dimensional secondary traits (e.g., images, spectrometry) as well as for genomic prediction with data from genetically heterogeneous backgrounds. We will present a diverse array of applications of the SSI ranging from the use of SSIs for image processing to genomic prediction with heterogeneous genetic backgrounds with single and multi-trait data. Empirical results across a wide range of applications suggests that the proposed methodology can lead to increases in prediction accuracy (relative to GBLUP) between 5-20%.

Using information across tissues and genes to predict gene expression in Transcriptome-wide Association Studies

Dr Fabio Morgante¹

¹*Clemson University, South Carolina, USA*

Session 20, November 11, 2020, 12:00 PM - 1:25 PM

Biography:

Fabio Morgante is an Assistant Professor at the Clemson University Center for Human Genetics (US). He holds a BS and an MS in Agricultural Sciences from the University of Florence (IT), an MSc in Animal Breeding and Genetics from The University of Edinburgh (UK), an MR in Statistics and a PhD in Genetics from North Carolina State University (US), working under the mentorship of Dr. Trudy Mackay. He did postdoctoral research at The University of Chicago (US) with Drs. Matthew Stephens and Yang Li. He is interested in understanding the genetics of complex traits and improving polygenic prediction.

Several methods have been proposed to combine GWAS data with expression reference panels to discover genes associated with complex traits. One class of such methods goes by the name “Transcriptome-wide Association Study” (TWAS). Briefly, TWAS imputes the genetically regulated part of gene expression in GWAS samples, using weights calculated in reference panels. These imputed gene expression measurements are tested for association with phenotypes. The initial TWAS methods calculated weights, and imputed expression, one tissue at a time. Recently, multivariate regression methods have been developed to use information from multiple tissues concurrently. These multivariate methods improve accuracy of gene expression imputation by exploiting sharing of expression Quantitative Trait Loci (eQTLs) among tissues. However, these methods remain suboptimal as they analyze only one gene at a time, failing to exploit information about eQTL sharing that multiple genes can provide. Here, we describe new Bayesian multivariate regression methods that are able to use data at many genes to learn patterns of eQTL sharing across tissues, borrowing strength across genes. By using a flexible prior, these methods are able to cope with different patterns of effect sharing and specificity across tissues. We evaluated the accuracy of gene expression predictions using simulations across genomic heritability values, number of tissues and patterns of sharing. Our results suggest that these new methods can predict at least as well as existing multi-tissue methods in scenarios where effects are completely shared across tissues, and outperform existing multi-tissue methods in scenarios where there is no sharing of effects across tissues.

A statistical model for genomic predictions of high-dimensional traits

Dr Daniel Runcie¹

¹University of California Davis, USA

Session 20, November 11, 2020, 12:00 PM - 1:25 PM

Biography:

Research in my group focuses on statistical and functional approaches to study genetic variation and gene-environment interactions in plants. Our goals are to improve predictions of crop performance, learn about forces that shape the evolutionary histories of natural populations, and identify critical systems that limit plant responses to climate change.

Measuring and modeling multiple traits at once can accelerate the rate of genetic gain in breeding programs, whether their goal is to improve a single target trait or simultaneously improving many traits. Multi-trait data is widely available, from high-throughput phenotyping technologies, repeated measures, or multi-environment trials. However, the vast majority of statistical models used in plant and animal breeding today handle only a single trait (or at most a few traits) at a time. We have developed a new statistical model for genomic prediction that efficiently and robustly scales to thousands of traits, allowing simultaneous predictions of high-dimensional phenotypes using data from complex experimental designs. Our approach overcomes the computational bottlenecks and over-parameterization challenges of traditional multi-trait linear mixed models by combining recent innovations in computational algorithms and statistical theory. These include: i) efficient and tunable Bayesian priors that prioritize only the strongest, most informative signals in Big Data, ii) a latent factor structure for trait covariances, and iii) efficient approximation and implementation schemes for reducing computational costs in mixed models. We will demonstrate the utility of our approach in the context of multi-trait multi-environment genomic predictions using data from maize.

Pooled genotyping strategies for the rapid construction of genomic reference populations

Dr Pamela A. Alexandre¹, Dr Laercio R. Porto-Neto¹, Dr Emre Karaman², Dr Sigrid A. Lehnert¹, Dr Antonio Reverter¹

¹Commonwealth Scientific and Industrial Research Organization, Agriculture & Food, Brisbane, Australia, ²Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark

Session 21, November 11, 2020, 7:00 PM - 8:30 PM

Biography:

Dr Pâmela Alexandre holds a BS in Animal Science from the Faculty of Animal Science and Food Engineering of the University of São Paulo (FZEA/USP, 2007-2011) and a Masters in Animal Science from the same institution in the area of molecular biology applied to livestock (2012-2015). During the MSc, she completed a five-month research internship at the University of Copenhagen/Denmark with the Group of Animal Breeding, Quantitative Genetics & Systems Biology under the guidance of Prof. Dr Haja Kadarmideen (BEPE/FAPESP). She holds a PhD from the Animal Bioscience Program of FZEA/USP (2015-2019). During her PhD, she completed a six-month research internship at the Commonwealth Scientific and Industrial Research Organization (CSIRO-Australia) under the guidance of Dr Antonio Reverter (BEPE/FAPESP). She is currently a postdoctoral fellow at CSIRO Agriculture and Food, with the Animal Genomics Team. She is interested in applying high throughput technologies to uncover the molecular aspects behind complex phenotypes, to generate innovative approaches to improve animal production and sustainability.

To explore a cost-effective opportunity for genomic evaluations of multi-sire commercial herds, we compared the accuracy (ACC) of bulls' genomic estimated breeding values (GEBV) using different pooled genotype strategies and simulated genomic and phenotypic data for one low (t1) and one moderate (t2) heritability trait. Sires' GEBV were calculated using a univariate mixed model, with a hybrid genomic relationship matrix relating 200 sires to their 2,200 progeny, the latest grouped in pools of 2, 5, 10, 20, 25, 50 or 100 individuals. Pooling criteria were: at random, grouped sorting by t1, grouped sorting by t2, or grouped sorting by a combination of t1 and t2. Although the best accuracy was achieved for a given trait when pools were grouped based on that same trait (ACC-t1:0.50-0.56, ACC-t2:0.66-0.77), pooling by one trait impacted negatively on the accuracy for the second trait (ACC-t1:0.25-0.46, ACC-t2:0.29-0.71). Pooling progeny based on a combined measure may be an alternative to use the same pools to calculate sires' GEBVs for both traits (ACC-t1:0.45-0.57, ACC-t2:0.62-0.76). Pools of 10 progeny represented a good compromise between loss of accuracy (~10-15%) and cost savings (~90%) with genotype assays. Pools assigned at random presented the poorest results (ACC-t1:0.07-0.45, ACC-t2:0.14-0.70). In conclusion, pooling by phenotype is the best approach to implementing genomic evaluation at commercial level, particularly when pools of 10 individuals are evaluated. While combining phenotypes seems a promising strategy to allow more flexibility to the estimates made using pools, more studies are necessary in this regard.

Increasing the accuracy of genomic prediction for crossbred livestock: examples from dairy cattle

Dr Iona Macleod¹

¹*Agriculture Victoria, Australia*

Session 21, November 11, 2020, 7:00 PM - 8:30 PM

Biography:

Iona MacLeod is a senior research scientist with Agriculture Victoria Research based in Melbourne, Australia, at the Centre for AgriBioscience. After a varied career, Iona completed a PhD in 2013 in quantitative and population genomics and currently works in the field of quantitative livestock genomics.

There is increasing interest in generating accurate genomic predictions for crossbreds because various livestock production systems use crossbreds as breeding stock and for production. We investigated the accuracy and bias of genomic prediction for crossbred dairy cattle compared to purebreds using several approaches. The traits tested were milk, fat and protein yields. Genomic prediction was implemented with a Bayesian method and a custom set of 46,000 sequence genotypes selected from a multi-omics analysis of whole genome sequence. With these tools we explored the use of different multi-breed training populations: Holstein, Jersey and their crossbreds, where the latter included more complex crosses than just F1. The multi-breed reference populations included: a Holstein dominated set of over 36,000 pure and crossbred animals, and a set of over 15,000 pure and crossbred animals with equalised Holstein and Jersey breed proportions. In addition, we compared the accuracy and bias of genomic prediction using the custom set of sequence genotypes versus standard commercial SNP chip genotypes. The results indicated that it was preferable to equalise breed proportions in the multi-breed training approach rather than have a larger multi-breed set that was dominated by one of the breeds represented in the crossbreds. It was also advantageous to include crossbreds in the multi-breed training population. The selected set of custom sequence variants generally outperformed the standard commercial SNP array, highlighting the potential value of creating custom SNP arrays where variants are selected based on a comprehensive multi-omics data analysis.

Genomic analysis reveals new genes and causal mutations for the environmental variance of litter size in rabbits

Ms Cristina Casto Rebollo¹

¹*Institute for Animal Science & Technology, Spain*

Session 21, November 11, 2020, 7:00 PM - 8:30 PM

Biography:

My name is Cristina Casto-Rebollo and I am 26 years old. I studied Biotechnology at the University of Pablo the Olavide (Seville) before undertaking a MSc in Animal Breeding and Reproduction Biotechnology at the Universitat Politècnica de València, where I started a PhD in Animal Breeding and Genetics in 2018, under the supervision of Noelia Ibáñez-Escriche and Maria-Jose Argente. My PhD research focusses on trying to unravel the genetic background of the environmental variance of litter size (VE) in divergently selected rabbit lines for high and low VE, and understanding its effects on the gut microbiome.

The environmental variance (VE) is partly under genetic control and recently has been proposed as a measure of an animal's resilience. Unravelling the genetic background of the VE of complex traits could help to improve the animal's resilience. The objective of this study was to identify the candidate genes and the causal mutations associated with the differences on VE of litter size in rabbits. We combined results of a genome-wide association study (GWAS) and a whole genome sequencing (WGS) analysis using data from a divergent selection experiment for VE of litter size in rabbits. These divergent lines showed differences in resilience. Four associated genomic regions with the VE of litter size were identified in the chromosome 3, 7, 10 and 14 with two GWAS approaches (Single-marker regression and Bayesian multiple-marker regression). A total of 38 genes were in the associated genomic regions. We identified missense, UTR and splicing variants in 17 of the 38 genes. Functions of these 17 genes were related to the immune system, the development of sensory structures, the stress response and the gene expression, among others. The genes HDAC9, HUNK, ITGB8, MIS18A, ENSOCUG0000021276, URB1 and PAXBP were highlighted due to the presence of homozygote variants in one of the rabbit lines and not in the other. This is the first study combining GWAS and WGS analysis to unravel the genetic background of the VE. The results suggested a control of the VE of litter size through the immune system, the stress response and the nervous system.

Genetic parameter for variability of milk production in cattle as indicator of environmental sensitivity

Mr. Enrico Mancin¹, Dr. Cristina Sartori¹, Dr. Lucio Mota¹, Prof. Roberto Mantovani¹

¹*University Of Padova, Legnaro (PD), Italy*

Session 21, November 11, 2020, 7:00 PM - 8:30 PM

In many livestock sectors it is desirable that animals are robust to environmental changes. In dairy cattle environmental sensitivity has been studied mainly as variations in residual variance for production. In other species, studies have considered the variability of other productive character as environmental variance., e.g. variability of litter size in rabbits and swine, that is directly connected with individual fitness and disease resistance. This study aimed to investigate the possibility to apply the same approach in dairy cattle by estimating the genetic component of milk production variability and its genetic correlation with longevity. Aosta Chestnut cattle was considered as a case study. The final dataset contained 29620 lactation yields of 9046 animals (3.27 lactation/cow). Only cows with the whole career in one herd were maintained. Milk yield was pre-corrected for the effects of the number of lactation (NL) and the year-month of lactation. Variability was calculated as coefficient of variation (CV) and standard deviation (SD) for milk yield; longevity was expressed as maximum NL. Bitrait analyses between milk CV and milk SD and longevity were done using a Gibbs sampling algorithm. The heritability of milk CV and milk SD were 0.07 and 0.06, respectively and both traits had positive genetic correlations with longevity (0.6). Milk yield had a correlation of 0.2 with milk SD and, interestingly, of -0.5 with milk CV. Concluding, genetic components of environmental sensitivity are low but they show medium-high genetic correlations with a fitness trait as longevity.

Environmental influences on the effects of new mutations

Dr Katrina McGuigan¹

¹UQ, Brisbane, Australia

Session 21, November 11, 2020, 7:00 PM - 8:30 PM

Biography:

Katrina McGuigan is an evolutionary quantitative geneticist interested in understanding the evolution of genetic variation for quantitative traits, and how the high-dimensional covariation of genetic variation among traits affects phenotypic evolution. She is a senior lecturer at the University of Queensland, Australia. Her group is interested in finding new experimental and statistical approaches to gain insights into complex phenotypes. Current research projects focus on the effects of new mutations on quantitative genetic variation, and how that variation is affected by background genotype and fitness, and by the external macro-environmental variation.

Quantitative genetic variation is ultimately the result of new mutations, and the distribution of phenotypic or fitness effects of mutations have implications not only for the maintenance of genetic variance, but for a wide range of biological phenomena. The magnitude of mutational variance introduced each generation is quite variable, ranging over a couple of orders of magnitude. Some of this variation can be explained by factors such as variation in mutation rate, or in inferred number of contributing loci, but variability in estimates remains. Using *Drosophila serrata* as a model, we investigated the potential contribution of environmental variability to estimates of mutational variance, using a panel of inbred lines that accumulated mutations over 20 generations. We considered both micro- and macro-environmental effects on mutational variances. Repeated measures of size and wing shape traits across six sequential generations revealed substantial variation in the magnitude of mutational variance, which could not be explained by the effects of drift or selection on new mutations. Thus, where estimates of mutational variance come from single generation point estimates, there might be substantial variation due to micro-environmental variation from generation to generation. We also how mutational variance was affected by mild (i.e., not reducing viability) changes in developmental temperature and nutrition. We observed mutational variance in development time was strongly affected by nutrition environment, but not by temperature. These studies further highlight that environmental effects on mutational variances can be independent of effects on the trait mean (i.e., plasticity), and independent among traits.

Estimating mutation rate and characterising de novo mutations in pigs

Dr Christina Rochus¹

¹Wageningen University, The Netherlands

Session 21, November 11, 2020, 7:00 PM - 8:30 PM

Biography:

Dr. Christina Rochus is a postdoctoral researcher at Wageningen University and Research in the Animal Breeding and Genomics Group. Dr. Rochus' research interests include genetic diversity and adaptation in livestock. Currently she is focused on estimating mutation rate, characterising de novo mutations and estimating their effects on complex traits in domestic pig populations.

Mutation rate and the effects of de novo mutations (DNM) on quantitative traits are not well known, especially in livestock species. In humans, direct estimates of mutation rate have changed our understanding of evolutionary timing, and DNM have been associated with several developmental disorders. Livestock breeding programs provide very good designs for estimating the impact of DNM on complex traits because genotypes and phenotypes are available for a large number of offspring carrying the DNM. The aims of our project are to detect DNM in two commercial pig breeding lines, estimate mutation rate and predict the effect of DNM on quantitative traits. We sequenced 48 trios, for 150 bp paired end reads at 30X coverage. De novo SNP mutations were detected using the GATK trio aware method. All probands had between 30 and 425 offspring genotyped with a 50K SNP chip, and were phenotyped for 32 functional- and production-related traits to estimate mutation effects. A preliminary estimate of mutation rate in pigs is 1.8×10^{-8} per site per gamete and mutation rate is higher near the ends of the chromosome: both findings are consistent with estimates from humans and cattle. Most DNM are located in introns (67%) and upstream (22%) of genes. In the coming months, DNM will be phased and we will estimate their effects on quantitative traits. These results will generate fundamental knowledge on the contribution of DNM to complex traits in mammals, exploiting pig breeding programs as a powerful source of data.

Inherited genetic variation, acquired mutations and clonal selection

Dr Steve McCarroll¹

¹Harvard University, USA

Session 22, November 12, 2020, 7:00 AM - 8:25 AM

Biography:

*Steve McCarroll is Professor of Biomedical Science and Genetics at Harvard Medical School. His research program reflects the diversity of his training. He was a Ph.D. student in Cori Bargmann's lab (genetics and neuroscience in *C. elegans*) at U.C. San Francisco, then a postdoc in David Altshuler's lab (human genetics and genomics) at MGH and the Broad Institute. The focus of Steve's research is how human genomes vary and how this variation shapes human biology. His lab has led the development of single cell droplet gene expression sequencing and is using this technology to investigate allele specific gene expression in individual cell types.*

Inherited genetic variation and acquired mutations are generally studied separately in science as distinct and independent influences on human phenotypes – the former following strong mathematical patterns at a population level, the latter appearing random and capricious. I will discuss our recent and emerging findings that these forms of variation are intimately connected to one another and synergize in powerful ways. We are learning this by studying clonal hematopoiesis, the common age-related acquisition of large populations of blood cells (generally 1-50% of a person's blood cells) that share a somatic mutation in common and thus appear to have derived from a single somatic cell. By identifying clones in blood-derived DNA data sampled from hundreds of thousands of people (e.g. UK Biobank), and studying the acquired mutations in relationship to inherited variation, we are finding two powerful connections between inherited variation and acquired mutations. First, we have found seven genes so far in which inherited variation (at least 50 different segregating variants) makes it highly likely that an individual will develop clones with specific somatic mutations during their lifetime – so likely, in fact, that the acquisition of clones with specific somatic mutations segregates in families like a Mendelian trait. Second, we have found that acquired mutations and clonal selection routinely act upon polygenic risk, causing cells to “level up” their polygenic risk for proliferation-related traits. These emerging results are causing us to appreciate the biological connectedness of these forms of variation and to question our earlier belief that acquired mutations do not strongly contribute to heritable traits.

Multi-trait evolution in the presence of social interactions

Dr Jarrod Hadfield¹

¹University Of Edinburgh, Edinburgh, United Kingdom

Session 22, November 12, 2020, 7:00 AM - 8:25 AM

Biography:

Jarrod is an evolutionary biologist based in the Institute for Evolutionary Biology at the University of Edinburgh. He is interested in understanding how natural and kin selection operates in wild systems, and how the genetic basis of traits determines how they evolve in response to selection. His lab combines theoretical work and statistical development with empirical work primarily on wild populations of the bird. Current QG projects in the lab are 1) the importance of indirect genetic effects (parent-offspring, siblings) for the evolution of growth traits and the relationship between indirect genetic models and kin selection models 2) the role of phenotypic plasticity versus genetic differentiation in response to spatially and temporally fluctuating environments 3) the relative importance of mutation versus selection in determining levels of quantitative genetic variation and 4) theory and statistical methods for dealing with genetic and environmental skew. In addition to these biological questions, he is interested in general statistical methods for hierarchical models, particularly MCMC methods for non-Gaussian data.

Most evolutionary biologists use kin selection models and Hamilton's rule to understand the evolutionary process in the presence of social interactions. In contrast, most quantitative geneticists use indirect genetic effect models. Using causal analysis, I show how these two modelling frameworks are connected and derive a multi-trait version of Hamilton's rule that has close connections to the Lande equation. When multiple traits are under selection Hamilton's original rule is shown to only be applicable when all traits are at an evolutionary equilibrium. The rule can be salvaged by considering selection at the genetic level rather than the phenotypic level but by doing so the intuitive appeal of Hamilton's rule is lost. Understanding the evolution of a single trait that is correlated with a suite of other traits under selection is a daunting empirical task. Quantitative geneticists using indirect genetic effect models have sidestepped this issue by measuring the aggregate impact of many traits of a social partner on a focal individual's trait rather than the impact of each trait individually. In general, this approach gives the wrong answer, and the conditions under which it gives the right answer are quite restrictive. Nevertheless, this restrictive set of conditions might be fulfilled in many systems.

Simultaneous quantification of mRNA and protein in single cells reveals trans-acting genetic variation in gene expression

Dr Frank Albert¹

¹University of Minnesota, USA

Session 22, November 12, 2020, 7:00 AM - 8:25 AM

Biography:

*Frank Albert is an Assistant Professor in the Department of Genetics, Cell Biology, & Development at the University of Minnesota. His laboratory combines experimental and computational approaches to study the effects of genetic variation on gene expression, using the yeast *Saccharomyces cerevisiae* as a powerful and tractable model system. A particular focus is on the identification of causal DNA variants that alter gene expression in cis and trans, and on genetic influences on protein levels compared to mRNA abundance.*

DNA variants that influence the expression of distant genes in trans are a key source of heritable phenotypic variation. Studies in humans, mice, and yeast have suggested the existence of trans effects that specifically affect mRNA or protein levels, but mapped mRNA and proteins at different times, with different experimental designs, and low statistical power. We developed a system for simultaneous quantification of mRNA and proteins in single, live cells of the yeast *Saccharomyces cerevisiae*. Genes of interest are tagged with green fluorescent protein (GFP) followed by a CRISPR gRNA sequence flanked by ribozymes. After transcription, the ribozymes release the gRNA, which directs dCas9 fused to a transcriptional activator to drive an mCherry gene. Thus, GFP and mCherry report on protein and mRNA levels, respectively, in the same cells. Using this system, we mapped the expression of ten genes. In millions of recombinant cells from a cross between two yeast strains, we used fluorescence-activated cell sorting to collect thousands of cells with high or low mRNA or protein levels. Pooled whole-genome sequencing provided high statistical power to map loci that affect mRNA or protein. The ten genes were affected by 97 trans-acting loci (8% FDR). Only 30% of loci affected mRNA and protein for the same gene with the same direction of effect. Half the loci influenced protein but not mRNA. We fine-mapped one protein-specific locus to a nonsense mutation in the YAK1 kinase gene. These complex and distinct influences on mRNA and proteins have important implications for quantitative genetics.

Future application of genomics in the poultry industry

Dr Rachel Hawken¹

¹Cobb-Vantress, USA

Session 23, November 12, 2020, 12:00 PM - 1:25 PM

Biography:

Rachel Hawken currently leads a team at Cobb-Vantress developing genetic tools to maintain a competitive product for the broiler breeding industry. To achieve this goal her team explores genomic and statistical genetic approaches and engages innovative technologies and collaborative research. Rachel received her Ph.D. at The University of Melbourne, School of Veterinary Science. She moved to Minnesota to complete Postdoctoral Fellow and Research Fellow positions at the School of Veterinary Science at the University of Minnesota. Rachel then returned to Australia to join the Commonwealth Science and Industrial Research Organization (CSIRO) as a Senior Research scientist, before being recruited to Cobb in Arkansas. Rachel is now the Senior Director for Genetics (Genomics and Quantitative Genetics) at Cobb. The focus of Rachel's early academic career was the development and testing of genomic tools to enable the selection of superior breeding stock for the swine, sheep, dairy and beef industries. Since joining Cobb, Rachel has implemented such genomic tools for the genetic improvement of the Cobb broiler. In addition, she has had the privilege of designing and implementing logistic and sampling systems needed to engage genomic technologies. Rachel is also responsible for the statistical analyses of large broiler performance data sets produced by Cobb to achieve genetic improvement.

Poultry production needs to double in the next 25 years to meet the growing demand for safe and affordable animal protein globally. To meet the demands of newly regulated and emerging markets and build a sustainable future for the world's broiler industry we need to make genetic improvements for production, growth efficiency, environmental stresses and health and welfare. The integration of new technologies, such as genomics provides valuable data and new approaches toward addressing these challenges. This presentation will cover the impact and the future applications of genomics on our industry.

How can biology and breeding contribute to improving food systems and climate change?

Dr Edward Buckler¹

¹USDA, USA

Session 23, November 12, 2020, 12:00 PM - 1:25 PM

Biography:

Ed Buckler is a USDA-ARS crop geneticist based at Cornell University campus.

Ed's research focusses on developing genomic, statistical, and bioinformatic methods with applications in maize and other crops. Key questions are:

- How does genetic variation give rise to phenotypic variation?*
- How can we use genetics to make agriculture more efficient and share those efficiencies globally?*
- How can we reduce the impact of agriculture on the environment?*

Ed's tools and approaches have been applied to over 3000 other species – everything from human genetics, nearly every crop, and many species of ecological interest. Ed's latest papers have been on machine learning of RNA expression and the relationship between eQTL and fitness.

The demands of food production, fuels, nutrition, and climate change are going to require that thousands of species undergo genomic selection over the next two decades. The approaches for making genomic selection models are too inefficient for scaling to thousands of species. Here, we propose how to use evolution and machine learning of the functional elements to begin developing robust models that work across species. Evolutionary comparisons over 20 million years are efficiently identifying key distal regulatory elements. Maize and Arabidopsis diverged 140 million years ago, yet our initial machine learning models predicting gene expression have significant transferability. Finally, models trained in bacteria and archaea are showing promise for understanding plant protein adaptations. The most significant obstacle to building transferable models may be the social aspect of getting disparate communities of scientists to make mutually interoperable and transferable models.

The Open Chromatin Regulation of Complex Traits

Mr. Nathaniel W. Ellis^{1,2}, Dr. Annette M. Fahrenkrog³, Dr. Kelly M. Balmant², Dr. Daniel Conde², Mr. Christopher Dervinis², Mr. Jerald D. Noble^{1,2}, Dr. Marcio F. R. Resende^{1,4}, Dr. Jason A. Smith², Dr. William B. Barbazuk^{1,5}, Dr. Matias Kirst^{1,2}

¹*Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, United States*, ²*School of Forest Resources and Conservation, University of Florida, Gainesville, United States*, ³*Servivios CGV SpA, Osorno, Chile*, ⁴*Horticultural Sciences Department, University of Florida, Gainesville, United States*, ⁵*Biology Department, University of Florida, Gainesville, United States*

Session 23, November 12, 2020, 12:00 PM - 1:25 PM

Populus deltoides (North American Eastern cottonwood) is a short-rotation woody crop with strong potential for bioenergy production because of its fast growth and wide adaptation to the Midwest and Southern United States. *P. deltoides* is cultivated worldwide for lumber and biomass, but development of new cultivars is hindered by lengthy breeding cycles and difficulties in phenotyping traits such as disease resistance and yield. The long-term goal of this research is to uncover and apply genomic information to guide and accelerate improvement of poplar cultivars. In order to uncover genes regulating critical traits we are characterizing a genetically unrelated population of 425 *P. deltoides* individuals. Genome-wide association studies (GWAS) have been applied and include growth properties measured under field conditions and disease resistance to *Sphaerulina musiva* (Septoria Leaf Spot) in this population. In addition, we are using the Assay for Transposase-Accessible Chromatin combined with next-generation DNA sequencing (ATAC-seq) to uncover open chromatin regions across the genome to study variants contained within them. We have performed replicated ATAC-seq experiments upon six different tissue types allowing for a comprehensive evaluation of chromatin accessibility. Our initial evaluations observe the accessible regions of the genome to represent ~ 7% of the overall genome. Our hypothesis is polymorphisms within these accessible regions of the *P. deltoides* genome contribute to a large proportion of the phenotypic variance.

Detecting the functional consequences of human genetic variation using high-throughput differentiation of human induced pluripotent stem cells

Dr Daniel Gaffney¹

¹Wellcome Sanger Institute, UK

Session 24, November 12, 2020, 7:00 PM - 8:25 PM

Biography:

Daniel is a Group Leader at the Wellcome Sanger Institute. The long-term goal of Daniel's group is to understand the molecular and cellular consequences of genetic changes in gene regulatory regions. His research combines statistical genetics with high-throughput experimental techniques in human cells to address these questions. Much of the group's recent research has been focussed on using human induced pluripotent stem cells (hiPSCs) and cells derived from hiPSCs as model systems to map and characterise human noncoding genetic changes.

Despite substantial progress in mapping expression quantitative trait loci (eQTLs), there are significant gaps in our resources to interrogate the function of human noncoding genetic variation. In particular, although many cell types, particularly immune cells, change substantially upon exposure to environmental stimulus, capturing genetic changes that alter these responses is challenging. A second limitation is that many important cell states are transitional, and briefly during embryonic development. I will present two related pieces of work from our group that attempt to address these problems using human induced pluripotent stem cells (hiPSCs). In the first project, I will show how we use hiPSC-derived macrophages from hundreds of individuals to map eQTLs across 28 stimulated cell states. Our results highlight that, even in a single cell type, detectable eQTLs are present at most expressed genes when cells are stimulated in different ways, with many disease associations showing a condition-specific effect on expression. In the second part of my talk, I will demonstrate how pools of hiPSCs, profiled using single cell RNA-seq, can be used to detect eQTLs during neuronal differentiation. Here, we show that scores of loci for cognitive function and related traits can be confidently colocalised with a specific gene during hiPSC-neuronal differentiation, but are absent from GTEx. Finally, I will describe factors that govern efficiency of iPSC differentiation to a neuronal cell fate, show that this can be reliably predicted from expression in iPSCs, and discuss the implications for future work using hiPSCs to model human regulatory variation.

The revised title for this talk is:

Evidence of horizontal indirect genetic effects in humans

Professor Albert Tenesa^{1,2,3}, Dr Oriol Canela-Xandri^{1,2}, Dr Konrad Rawlik^{1,3}

¹University Of Edinburgh, Edinburgh, UK, ²MRC-Human Genetics Unit, Edinburgh, UK, ³Roslin Institute, Edinburgh, UK

Session 24, November 12, 2020, 7:00 PM - 8:25 PM

Biography:

Albert Tenesa studied agricultural engineering at the Polytechnic University of Valencia before undertaking an MSc in Quantitative Genetics and Genome Analysis at the University of Edinburgh, where he also completed his PhD in quantitative genetics, under the supervision of Peter Visscher and Sara Knott. After a postdoc with Professor Malcolm Dunlop working on colorectal cancer genetics, he became a group leader at the Roslin Institute and the Medical Research Council Human Genetics Unit at the University of Edinburgh. Albert's research aims to understand how genetic variation contributes to phenotypic variation of complex traits in humans and to develop the tools that will help in transitioning from study cohorts to whole populations. You can follow his research team on twitter: @GroupTenesa

Humans show an extraordinary range of phenotypic variation. From our height, to our hair colour and to our disease risk each of us is unique. We know that differences among individuals are driven by genetic and environmental factors. Discovery of these factors has been fuelled in later years by technology, co-operation among scientists and big datasets. Access to UK Biobank, which has deep phenotyping data and genotype information on 500,000 human participants, has allowed us to explore phenomena like assortative mating, indirect genetic effects, gene by environment interactions or elusive epistatic effects. Furthermore, it has allowed us to perform Genome-Wide Association Studies (GWAS) at an unprecedented scale in terms of cohort size and number of traits, thereby providing a first global insight into pleiotropy. The massively large datasets used mean that GWAS remain computationally intensive, which hinders progress in research, especially for groups without access to the necessary computational requirements. Whilst providing open access to summary statistics has democratised access to GWAS results, this solution remains unsatisfactory for a number of reasons. For instance, matching scale and expert knowledge of thousands of traits is difficult and reproducibility, even when using the same dataset, is not always guaranteed. To address this, we propose a new GWAS paradigm where users set up the model to be applied in their GWAS and summary statistics are generated and streamed instantly on demand. We hope this 'open access' approach to GWAS will speed up scientific discoveries and will catalyse improvements in population health.

Breeding Crops Resilient to Future Climate Change using Environment Covariant Enriched Genomic Prediction

Dr Hans D. Daetwyler¹

¹Agriculture Victoria, Australia

Session 24, November 12, 2020, 7:00 PM - 8:25 PM

Biography:

I hold a joint appointment as Associate Professor and Research Leader Computational Biology at La Trobe University and Agriculture Victoria, respectively. My group's focus is efficient large-scale statistical analysis using variety of methods and data types in agricultural plant and animal species. Our main aim is to use predictive statistics that accelerate genetic improvement for important traits. Often this is accomplished via genomic prediction, including the use of whole-genome sequence and functional genomics as well as metabolomics. Using computer simulations we investigate the impact of these methods on breeding population genetic gain and diversity in the long term.

Models fitting genotype-by-environment interactions (GxE) making use of multi-environment trials are commonly used to evaluate breeding lines. We developed mixed model methodology for the incorporation of environmental covariates (EC, e.g. rainfall, humidity, temperature) directly into genomic prediction and genome-wide association. This allows for the prediction of lines' performance in projected future climatic conditions and the investigation of loci-by-EC interactions. While the method is applicable to any crop and EC combination, we have validated it in a large wheat population being bred for heat tolerance consisting of >2000 genotyped lines derived from diverse parents sourced globally and within Australia. Lines were deeply phenotyped across multiple years and locations for in-field agronomic (e.g. yield) and phenomic (e.g. thermal imaging), seed quality (e.g. protein), and up to 20 seed micronutrient traits (e.g. zinc, iron, phytate). Two sowing dates (normal and late) were used to test heat susceptibility during three main growth phases: vegetative (pre-anthesis period), anthesis (daily EC seven days pre and post), and grain fill period. This resulted in a total of 447 EC for each plot/line/trial, which were reduced via principal component analysis to a much smaller set explaining the most environmental variation. GxEC modelling increased genomic prediction accuracy for yield by up to 19% and revealed putative loci-by-EC interactions. In-field validation with and without mobile heat chambers is underway. The germplasm resource and methods developed will increase our understanding of GxEC and allow for the genomic prediction of performance for projected future climate scenarios to pre-breed more climate resilient crops.

The impact of sex on gene expression and its genetic regulation across human tissues

Dr Barbara E. Stranger¹, Dr Meritxell Oliva², Dr. Manuel Muñoz-Aguirre^{3,4}, Dr. Sarah Kim-Hellmuth^{5,6}, GTEx Consortium

¹Northwestern University, Feinberg School Of Medicine, Chicago, United States, ²University of Chicago, Chicago, United States, ³Centre for Genomic Regulation, Barcelona, Spain, ⁴Universitat Politècnica de Catalunya, Barcelona, Spain, ⁵New York Genome Center, New York, United States, ⁶Max Planck Institute of Psychiatry, Munich, Germany

Session 25, November 13, 2020, 7:00 AM - 8:25 AM

Biography:

Dr Barbara Stranger, PhD is an Assistant Professor in the Section of Genetic Medicine at the University of Chicago Department of Medicine. In addition, she is a Faculty Fellow in the Institute for Genomics and Systems Biology and The Center for Translational Data Science. Her lab analyzes multi-dimensional human genomics data, particularly transcriptome data and genetic variation data, in the context of human health and disease. She develops effective analytic approaches for large-scale analysis of functional genomics data and applies systems biology methodologies to integrate data of different types to inform biology of complex traits. She is a member of the Analysis Working Group (AWG) of NIH's Genotype-Tissue Expression (GTEx) project and leads the Sex Differences Working Group within the GTEx AWG. Current projects in her laboratory focus on understanding mechanisms and consequences of sex differences in 1) The human transcriptome, 2) The human proteome, 3) The genetic architecture of neuropsychiatric traits, and 4) The genetics and genomics of cancer and response to therapy. She has a longstanding interest in evolution and also applies the tools and approaches of population and evolutionary genetics to her research areas.

Many complex human phenotypes exhibit sex-differentiated characteristics, however a comprehensive characterization of sex differences at the molecular level is lacking, and the underlying mechanisms of these differences remain largely unknown. Here, we explore the role of sex on a variety of transcriptome-related phenotypes. We have generated an extensive catalog comprising sex differences in gene expression and its genetic regulation across 44 human tissues surveyed by the Genotype-Tissue Expression (GTEx) Consortium (v8 release). This work demonstrates that sex strongly influences gene expression levels and cellular composition of tissue samples across the human body. The effect of sex on gene expression is widespread, with a total of 35% of all genes exhibiting sex-biased expression in at least one tissue. This suggests that many, if not most, biological processes are impacted by sex effects on the transcriptome. We expand the identification of cis-eQTLs with sex-differentiated effects and characterize their cellular origin. For cis-eQTLs derived by the GTEx Consortium, we performed a genotype-by-sex interaction eQTL analysis and identified 369 sex--biased eQTLs (sb--eQTLs). By integrating sb-eQTLs with genome-wide association study data, we identify dozens of gene-trait associations that are driven by genetic regulation in a single sex, including novel associations not detected with sex-agnostic approaches. Collectively, our integrative analyses provide the most comprehensive characterization of sex differences in the human transcriptome to date, with implications for complex traits.

A new interpretation for a 99 year-old equation: the Castle-Wright estimator is also a powerful test of natural selection

Professor Hunter Fraser¹

¹Stanford University, USA

Session 25, November 13, 2020, 7:00 AM - 8:25 AM

Biography:

We study the evolution of complex traits by developing new experimental and computational methods. Our work brings together quantitative genetics, genomics, epigenetics, and evolutionary biology to achieve a deeper understanding of how genetic variation shapes the phenotypic diversity of life. Our main focus is on the evolution of gene expression, since this is the primary fuel for natural selection. Our long-term goal is to understand the genetic basis of complex traits well enough to introduce them into new species via genome editing.

Distinguishing which traits have evolved under natural selection, as opposed to neutral evolution, is a major goal of evolutionary biology. Several tests have been proposed to accomplish this, but these either rely on false assumptions or suffer from low power. Here, I introduce a new approach to detecting lineage-specific selection that makes minimal assumptions and only requires phenotypic data from ~10 individuals. The test compares the phenotypic difference between two populations to what would be expected by chance under neutral evolution, which can be estimated from the phenotypic distribution of an F2 cross between those populations. Comparing its performance to the QTL sign test—an existing test of selection that requires both genotype and phenotype data—the new test achieves comparable power with 50- to 100-fold fewer individuals (and no genotype data). Applying the test to empirical data spanning over a century shows strong directional selection in many crops, as well as on naturally selected traits such as head shape in Hawaiian *Drosophila* and skin color in humans. Applied to gene expression data, the test reveals that the strength of stabilizing selection acting on mRNA levels in a species is strongly associated with that species' effective population size. In sum, this test is applicable to phenotypic data from almost any genetic cross, allowing selection to be detected more easily and powerfully than previously possible. Most surprisingly, the test is identical to the well-known Castle-Wright estimator, which has been widely used since 1921 to estimate the number of loci underlying quantitative traits.

Come join the multiple testing party!

Professor Matthew Stephens¹

¹*University of Chicago, United States*

Session 25, November 13, 2020, 7:00 AM - 8:25 AM

Biography:

Matthew Stephens is Professor of Statistics and Professor of Human Genetics at the University of Chicago, and a Gordon and Betty Moore Investigator in Data-Driven Discovery. He received a BA in Mathematics (1992) and Diploma in Mathematical Statistics (1994) from the University of Cambridge UK, and a D.Phil in Statistics (1997) from the University of Oxford UK. Dr Stephens is a statistician and data scientist, who has made seminal contributions to the practice and applications of statistics in genetics. His research interests include analysis of population structure, analysis of genetic association studies (including both complex traits, and molecular phenotypes, such as gene expression and chromatin accessibility), large scale regression and multiple testing. His lab distributes several widely-used software packages for statistical analysis, including PHASE and fastPHASE (for haplotype inference), GEMMA (for association testing) and BIMBAM (for genotype imputation and association testing). Dr Stephens's Honors include the Guy Medal in Bronze by the Royal Statistical Society in 2006, was honoured as a Medallion Lecturer by the Institute for Mathematical Statistics in 2014, and inclusion in the Thomson-Reuters list of Highly Cited Researchers 2014.

Multiple testing is often described as a "burden". But in genetics researchers nevertheless seem ever intent on collecting more and more data, and consequently conducting more and more tests. Examples include typing more and more genetic variants in more and more study groups being phenotypes for more and more different traits. Why do researchers subject themselves to this unnecessary burden? I believe that, in fact, most researchers implicitly view large number of tests, as more of an opportunity than a burden. Further, I believe that instead of labouring under this burden we should embrace this opportunity, and be looking for better statistical analysis methods -- particularly methods to share information across tests of different traits -- to better distinguish signal from noise. I invite you to a multiple testing party.

POSTER PRESENTATIONS

233

Autozygosity detection: a comparison of outcomes of different software and marker densities

Mr Sowah Addo

Biography:

I am originally from Ghana where I studied Agricultural Sciences (BSc.) and worked as a Teaching Assistant in 2007/2008. I further trained as an Erasmus Mundus Animal Breeding and Genetics (EM-ABG) scholar with a split stay at Wageningen University in The Netherland and Kiel University in Germany. In March 2016, I joined the Animal Breeding Group of Kiel University as a Research Assistant. There, I worked on a European Innovation Partnership Project that sought to develop population management strategies for local breeds. In May 2019, I moved to Kassel University and I have been pursuing the same research interest.

Runs of homozygosity (ROH) analysis is nowadays the preferred method of quantifying individual autozygosity. ROH detection is largely implemented in PLINK using an observational genotype-counting algorithms relying on user-defined parameter settings. Consequently, outcomes of different studies are often non-comparable. RZOOROH, a recent software, implements an efficient and accurate model-based approach to identify homozygous-by-descent (HBD) segments and it is devoid of parameter preconditioning. The aim of the current study was to ascertain how the outcomes of PLINK-based ROH detection compare to autozygosity detection using RZOOROH. Genetic data consisted of 40,753 single nucleotide polymorphism (SNPs) markers on 46 sheep, 34,066 SNPs on 51 pigs and whole-genome sequence (WGS) data on 6 horses. Furthermore, 26,932 SNPs featured on the EquineSNP50 Genotyping BeadChip were extracted from the WGS data to create an array-like data for the horse breed. Applying optimised parameter settings across species, estimates of average PLINK-based inbreeding coefficient were 8.13% (sheep), 26.54% (pig) and 21.0% (horse). These estimates are slightly lower but not significantly different (except for sheep) from those estimated using RZOOROH. Generally, RZOOROH detected more HBD segments including those not detectable by PLINK given its parameter thresholds. The HBD segments were grouped to infer the age of inbreeding and the oldest was found in the horse. We observed large differences in autozygosity estimates between WGS (21.0%) and array data (11.1%) mostly due to the accrual of inbreeding from very remote ancestors when using WGS data. RZOOROH provides valuable information for the optimisation of rule-based ROH detection analysis.

Whole-genome sequencing reveals population structure and demographic history of Nigerian indigenous pigs

Dr Adeniyi Charles Adeola

Biography:

My research investigates the biodiversity and evolutionary history of wild and domestic animals in Africa to positively impact conservation efforts for the sustainable utilization of genetic resources by using genomic data (typically mitochondrial DNA sequences) for genetic diversity assessment, population demography, phylogenetic analysis (NJ, ML, BI trees and MJ network) and estimation of divergence times (Bayesian analysis). My research has provided comprehensive biodiversity data that has served as a baseline for initiating effective conservation policies. Recent works include integration of genomic and transcriptome data using comparative and population genomic approaches to investigate genomic basis of animal adaptive evolution in Africa.

The population structure and evolutionary history of African indigenous pig is essential and fundamental for its genetic diversity. Herein, we analyze whole-genome sequences of 51 Nigerian Indigenous pigs (NIP) from three different geographic regions plus 210 wild and domestic pigs from Europe, Near East and Far East, Near East. NIPs clustered into three distinct subpopulations with varying levels of European and East Asian pig ancestry. Phylogenetic, PCA and ADMIXTURE analyses revealed that two subpopulations clustered closer to European Wild boar and domestic pigs, and the remaining one to East Asian wild and domestic pigs. Significant gene flow was detected from European wild boar into the NIPs, indicating that majority of NIP are of European ancestry. The two NIP subpopulations with substantial European ancestry showed a high level of genetic diversity and rapid LD decay, while the subpopulation with East Asian ancestry showed a lower genetic diversity and slower LD decay. It implies that the NIP with East Asian ancestry might have a smaller effective population size, consistent with our observation of its contribution to the small fraction in the sampled NIPs. Demographic analysis revealed the divergence of NIP subpopulations with East Asian and European ancestries around 200K years ago, indicative of the possibility of another *Sus* species. Our study provides insights into NIP population structure and demographic history, and would aid in future conservation efforts.

Predicting wheat yield from genotypes and environmental data using four machine learning approaches

Dr Hawlader Abdullah Al-Mamun

Biography:

I'm a Research Scientist at Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia with a formal education in computer science, bioinformatics and quantitative genetics. My current research focuses on the development of machine learning methods to integrate high-dimensional multi-omic data to i) improve prediction accuracy of commercially important phenotype in both plants and animals and ii) understand the various factors determining biological outcomes.

Specific research interests

- *Artificial intelligence and machine learning for optimization of biological problems*
- *Computational methods and statistical analysis of high-throughput genomic data*
- *Use of functional knowledge in genomic selection*
- *Functional integration of GWAS and gene expression data*

The ability to predict yield of untested genotypes has been a goal of plant breeding programs for many decades. In recent years the volume of data with potential to improve prediction accuracies has increased substantially. Machine learning has been shown to be a valuable tool for many prediction problems involving Big Data. Algorithms that have been utilised include deep neural networks (DNN), which have gained much attention due to their ability to solve complex prediction problems in many research areas, as well as decision tree-based ML methods, which have been a popular choice in biological domain because of their ability to identify important variables and the relationships between variables. Using genotypes from multiparent advanced generation intercross (MAGIC) wheat populations and dense environmental data, we compared the predictive accuracy of DNN and three state-of-the-art decision tree-based methods for yield prediction. The data included genotypes from CSIRO's four-way and eight-way MAGIC populations and environmental data such as temperature, precipitation and day length. These populations were grown and phenotyped in four consecutive years at three different locations. Using 10-fold cross validation (CV), the performance of DNN, random forest (RF), gradient boosting machine (GBM) and model tree (MT) were compared. The results showed that the decision tree-based methods outperformed DNN. Of the tree-based methods, GBM obtained the lowest mean RMSE and highest average correlation in most analyses. The achieved 10-fold CV accuracies (e.g. 0.61 for GBM) confirm the value of ML approaches for yield prediction using dense genomic and weather data.

Finding treasure (epistatic interactions) in a dark random forest

Dr Hawlader Abdullah Al-Mamun

Biography:

I'm a Research Scientist at Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia with a formal education in computer science, bioinformatics and quantitative genetics. My current research focuses on the development of machine learning methods to integrate high-dimensional multi-omic data to i) improve prediction accuracy of commercially important phenotype in both plants and animals and ii) understand the various factors determining biological outcomes.

Specific research interests

- *Artificial intelligence and machine learning for optimization of biological problems*
- *Computational methods and statistical analysis of high-throughput genomic data*
- *Use of functional knowledge in genomic selection*
- *Functional integration of GWAS and gene expression data*

Many phenotypes and disease traits in human, animals and plants are complex in nature and involve many genes and their interactions. Random Forest (RF) is a popular machine learning tool that is regularly reported as identifying epistatic interactions. To date, the RF based approaches for identifying interactions are based on variable importance measurements which cannot distinguish whether a discovered interaction is a true interaction or simply two variables with strong marginal effects. Additionally, if the interacting variables have small marginal effects, they will not appear near the top of the variable importance list. This means that detecting interactions based on variable importance can be problematic. To enable the identification of epistatic interactions when the interacting variables have an interaction effect with small or no marginal effects, a two-step approach was designed and implemented. First, pairs of variables occurring as parent-child pairs in the forest are tested against the null hypothesis. Pairs identified as potentially interacting are then tested in a second step to see if they have statistical evidence of a true interaction. The approach was evaluated on multiple simulation datasets and two real datasets. Simulation results demonstrated the method was able to identify true interactions. The real data analyses found a small number of interactions that were subsequently shown to improve the accuracy of prediction when included as interaction variables in input data. Although the method has only been used to identify marker interactions, the approach is equally applicable to detect Genotype by Environment (GxE) interactions.

Temporal Genomic Prediction: Using Historical Data to Predict Future Genotype Performance.

Dr Vivi Arief

Biography:

Vivi Arief is currently a Lecturer in Biometry for the School of Agriculture and Food Sciences, The University of Queensland. She graduated with B.Sc. in Agriculture from Bogor Agricultural University (Indonesia) and the MAgrSt and PhD in Plant Breeding and Quantitative Genetics from the University of Queensland. Her research focuses on the area of interface between statistics and plant breeding. Her current research involves developing QU-GENE simulation platform for genomic selection.

The application of genomic prediction methodology to historical data allows an evaluation of the pattern of performance of the breeding germplasm over the range of observed environments. By utilising appropriate relationship matrices, temporal genomic prediction enables a complete prediction of all tested genotypes to the past and present environments. Thus, it extends the period for which the breeding material is evaluated and enables the evaluation of long-term performance. The characteristics of these past environments, which are known and documented, can be used to estimate the likelihood of a similar type of environment occurring in the future. Therefore, the performance of genotypes in these past environments are predictors for their future performance in similar environments. This application of temporal genomic prediction is demonstrated using historical data from a wheat breeding program.

Genomic mechanisms underlying behavioral traits across livestock species and their potential application in genomic prediction of breeding values

Ms Amanda B. Alvarenga

Biography:

Amanda Botelho Alvarenga is a PhD student in Quantitative Genetics and Genomics in the Department of Animal Sciences at Purdue University, USA. Amanda has an honors BSc degree in Animal Science from the Federal University of Lavras, Brazil. She completed her MSc in Genetics and Animal Breeding at the “Luiz de Queiroz” College of Agriculture, University of São Paulo, Brazil. She is interested in statistics, quantitative genetics, and genomics applied to genetic improvement of livestock species. Her current PhD project focuses on integrating multiple data sources to maximize the performance of genomic predictions for temperament traits in American Angus cattle.

Behavioral traits directly influence livestock farming, including the safety of farmers, animal performance, longevity, and meat quality. Due to these economic and welfare implications, we performed a systematic review of 68 studies to better understand the underlying genomic mechanisms for various behavioral indicator traits in cattle (37 traits), pigs (55 traits), and sheep (22 traits). A total of 2,226 genomic regions and 715 candidate genes were retrieved for the three species. The behavioral traits with the highest number of associated genomic regions were temperament (cattle), backtest traits (pigs), and vocalization (sheep). The main overlapping genomic regions across species were located at BTA29 (GRM5 gene), SSC8 (NR3C2 gene), and OAR7 (LRRRC49) in the cattle, pigs, and sheep genome, respectively. There is strong evidence of conserved genes across mammal species. In this study, six genes were commonly identified in cattle and pigs (NR3C2, PITPNM3, RERG, SPNS3, U6, ZFAT). Additionally, 313 out of 715 genes were also identified in humans for behavioral, mental, and neuronal disorders. The genes identified were enriched in biologically important pathways such as dopaminergic mechanism, insulin secretion, neurotrophin, and hippo signaling. Additionally, suggestive biological pathways related to GnRH and estrogen signaling, steroid, olfactory, and in utero embryonic development were identified. These findings contribute to a better understanding of the genomic mechanisms underlying behavior. Incorporating the findings into a genomic prediction of breeding values, no differences in the accuracy were observed between ssGBLUP and weighted-ssGBLUP based on the genes identified. However, biology-driven outperformed the likelihood-dependent weighting techniques (e.g. no-prior information).

Characterization of healthy aging in blood cells at single-cell resolution

Ms Elyssa Bader

Biography:

Elyssa Bader is a PhD student in Molecular Genetics department at the University of Toronto under the supervision of Dr. Philip Awadalla. Her research focuses on understanding the variation of aging in blood cells. With integration of multi-omic datasets from large population cohorts such as the Canadian Partnership for Tomorrow's Health (CanPATH) and the UK Biobank, she is investigating how germline and somatic variation manifest in peripheral blood cells and affect their function. Elyssa is also a founding member of HER CODE CAMP, an organization that provides free coding workshops and networking opportunities for under-represented groups in computer science.

Although the aging process generally has negative effects on blood cells, some elderly individuals have exceptional blood health. Here, we use large population cohorts to isolate individuals at the extremities of the blood health spectrum and identify functional variation that is associated with healthy blood aging. We evaluate blood health using a risk score calculated from complete blood count (CBC) data. In the UK Biobank (n = ~500 000) and the Canadian Partnership for Tomorrow Project (n = ~320 000) cohorts, individuals with low CBC risk scores are less likely to be diagnosed with cardiovascular and cardiometabolic diseases, and some cancers. Further, individuals with high CBC risk scores were subsequently diagnosed with a cancer 1 month – 3 years later. To understand the functional differences between blood cells from individuals with low versus high CBC risk scores, we performed single cell RNA sequencing on over 400 000 cells and ATAC sequencing on PBMCs. We observe distinct changes in cell type proportions and transcriptional profiles dependent on CBC risk score and age. Further, to determine if some cellular populations can better discriminate healthy blood aging, we compare transcriptional profiles within cell populations between groups of individuals matched for age or CBC risk score. Finally, we discriminate germline from somatic variation to determine the timing of somatic events and associate these variants with functional changes across cell types. Our results demonstrate the utility of natural variation in healthy agers to uncover mechanisms that prevent or slow down the negative effects of aging.

Single-Step Genomic Selection in Turkeys for Health, Welfare, Efficiency and Production Traits

Professor Christine Baes^{1,2}, Flavio Schenkel¹, Emhimad Abdalla¹, HakimehEmamgholi Begli¹, Alexandra Harlander-matauschek³, Shai Barbut⁴, Sarah Adams¹, Emily Leishman¹, Ryley Vanderhout¹, Michelle Yahiro⁴, Nienke VanStaaveren^{1,3}, Ben J Wood^{5,6}

¹Centre for Genetic Improvement of Livestock, University of Guelph, ²Institute of Genetics, Vetsuisse Faculty, University of Bern, ³Campbell Centre for the Study of Animal Welfare, University of Guelph, ⁴Department of Food Science, University of Guelph, ⁵Hybrid Turkeys, C-650 Riverbend Drive, Suite C, ⁶School of Veterinary Science, University of Queensland, Gatton Campus

Biography:

Dr. Christine Baes is an Associate Professor and Canada Research Chair in the Department of Animal Biosciences at the University of Guelph in Canada. She also holds a researcher position at the University of Bern in Switzerland. Dr. Baes was born and raised on a dairy farm in Southwestern Ontario and learned to appreciate Canadian agriculture at a young age. After completing her Bachelors degree at Guelph, she went on to finish a Master of Science degree in Animal Welfare at the Universität Hohenheim and a PhD in Quantitative Genetics at the the Leibniz Institute for Farm Animal Biology and the Christian Albrechts Universität zu Kiel in Germany. She became involved in various large-scale livestock breeding projects (swine, horses, dairy cattle, goats), and, together with her team, strives to bridge the gap between cutting-edge research and practical application of new knowledge. Dr. Baes has extensive knowledge in the areas of quantitative genetics and statistical genomics as it relates to the genetic and genomic evaluation of livestock. She works with large data (SNP chip data as well as NGS data), and her programming experience includes FORTRAN, R, UNIX, and other languages. Dr. Baes has taught multiple courses on genetics, genomics, and statistics in Germany, Switzerland, and Canada. In her spare time, she runs a small farm outside of Maryhill, in Ontario Canada.

The implementation of single-step genomic BLUP (ssGBLUP) selection programs can increase the accuracy of breeding values, therefore improving the ability to estimate the genetic merit of livestock. In particular, traits with low heritability are amenable to genomic selection. Improvement of traits such as livability, disease resistance, fertility, and other health and welfare traits in turkeys could considerably advance breeding programs.

Effective application of ssGBLUP requires genomic information, population-based pedigree data, and high-quality phenotypes. The aim of this study was to apply ssGBLUP to commercial turkey lines (15,000 turkeys genotyped at 65K). This reference population was used to assess the increase in accuracy of selection using genomic information, which ranged from 0 to +0.3, depending on the trait. A further goal of the study was to monitor inbreeding within the different lines. A large number of high-quality phenotypes related to fertility, growth, production, and carcass composition were collected; additional health and behaviour phenotypes related to livability are currently being investigated and developed for use in performance testing. Furthermore, meat quality (e.g. white striation, water holding capacity, pH, sheer force and colour, etc.) and total carcass composition phenotypes were analysed.

With improved methodology, more detailed phenotypic information, and comprehensive data collection and integration, we present more accurate selection of parent stock for application in applied poultry breeding programs.

Unraveling molecular mechanisms of complex traits in pine with genome wide association and gene coexpression networks

Stephanie Karenina Bajay¹, Alexandre Hild Aono¹, Anete Pereira de Souza¹

¹University of Campinas, Brazil

Biography:

Biologist graduated from the Federal University of São Carlos, Stephanie completed her master's degree in 2017 in Plant Genetics and Breeding from the State University of Campinas (UNICAMP). From 2018 to 2019 she completed part of her PhD at the Forest Genomics Lab, University of Florida - Gainesville / FL / USA, in the area of forest breeding and genomic selection in tree species. She is currently a PhD student in Genetics and Molecular Biology-UNICAMP, with the project entitled "Genome Prediction and Genomic Association in Rubber Tree: Main Renewable Source for Rubber Production". She has experience in the following subjects: plant breeding, bioinformatics, molecular markers, plant genetics and genomics and population genetics.

Loblolly pine (LP), a long-lived tree species, is the most productive and economically important wood species in the southern United States. Genetic improvement programs for pine trees in this region have focused on survival, early and rapid growth, resistance to diseases and pests and stem shape. Although marker-assisted selection for these traits has been successful, most of them are quantitative and presumably influenced by the action of an unknown network of genes, interacting through complex molecular mechanisms. The extremely large size (1C=21.6G) and high complexity of the Pinus genomes have presented challenges for its characterization, task for the sequencing and computational analysis. In this study, we present the first comprehensive integrated analysis in LP of genome-wide association study (GWAS) and gene co-expression networks to provide an improved characterization of the gene space and to identify patterns of selection among orthologous gene families. We used populations with full-sib progenies from 45 crosses (~50 ind/fam) (hereafter family) tested in seven sites of Cooperative Forest Genetics Research Program (CFGRP) in the 2nd cycle of Florida LP selection. A total of 1,999 individuals were phenotyped and genotyped using capture probes targeting putative genes based on an elite germoplasm transcriptome from loblolly pine through a service provided by Rapid Genomics (Gainesville, FL). Narrow-sense individual-tree heritability varied between sites from 0.05 to 0.34 for rust; 0.05 to 0.1 for stem water potential (SWP) and 0.17 to 0.32 for volume. A total of 31,589 SNPs, called using FreeBayes v.1.0.1, was carried out to detect population structure and linkage disequilibrium. GWAS was performed through a multilocus mixed model using FarmCPU method implemented in R software. For the construction of weighted gene coexpression networks, three transcriptomes were assembled based on data from different pine species (Pinus taeda, Pinus elliotti and Pinus radiata). Based on GWAS analysis, we selected the putative genes associated with the target traits and searched for the cascade of related molecular mechanisms within coexpression networks. We found significant marker-trait associations with the three traits measured and also had correspondence in the assembled transcriptomes. The results advance our understanding of the genetics influencing wood traits and identifies candidate genes for future functional studies and to increase our understanding of quantitative genetics and the genomics of complex phenotypic variations in LP.

Acknowledgement the Fapesp (Fapesp 2019/13452-3, 2019/03232-6), Capes, USDA and CFGRP

Aquaculture Genome-Wide Association Studies: What I learned about admixture from human geneticists

Prof Elizabeth Boulding¹

¹*Department of Integrative Biology, College of Biological Sciences, University of Guelph, Canada*

Biography:

I became interested in population and quantitative genetics while I was conducting my B.Sc. and M.Sc. research on selection by shell-breaking crabs on their bivalve prey. I focused on the theoretical and empirical quantitative genetics of predator-resistant traits of shelled gastropods during my Ph.D. research at the University of Washington in Seattle. In 1993 I joined the faculty in the College of Biological Sciences at the University of Guelph in Ontario, Canada. I maintain long-term predatory crab-addition field experiments where my students and I investigate the demographic and evolutionary responses of the surrounding snail populations. My students and I also investigate the genomic basis of quantitative traits in the North American subspecies of Atlantic salmon in collaboration with Canadian and European university researchers and aquaculture company scientists. We are also now working on the genetic basis of propensity to remove sea lice from salmon clients by two species of cleaner fishes.

Human geneticists have developed sophisticated statistical methods for genome-wide association analysis (GWA) that adjust for population structure and for kinship independently. These methods allow very large and statistically powerful GWA of unmixed and admixed populations from different continents. However, aquaculture geneticists have been slow to incorporate these strategies even though it is not unusual for them to conduct GWA on aquaculture strains containing admixed individuals from highly diverged populations or even species. I applied kinship-free methods of correcting for population structure to a historical year-class of North American Atlantic salmon (*Salmo salar*) that contained a number of admixed individuals with European ancestry. Even after correction for population structure, I found that the GWA results showed less genomic inflation when individuals with more than with >10% European ancestry were removed from the analysis. The resulting GWA found no associations for age at sexual maturity on chromosome 25 - which harbours a major locus for this trait in European populations - but did detect a SNP with experiment-wide significance on its homeolog, chromosome 21. I hypothesize that the differences in karyotypes and genetic architecture that developed during the 600,000 years since the North American and European Atlantic salmon subspecies diverged makes it impossible to correct a GWAS containing admixed individuals with kinship-free methods. An alternative I have used in salmon, where the family sizes are typically large, is to use within-family only methods of GWA where correction for population structure unnecessary. I conclude that large differences in the genetic architecture of traits between highly-diverged and admixed populations from different continents, may require new statistical methods of correction for population structure that have yet to be developed.

Classifying true and false reciprocal translocation breakpoints from whole genome sequence data using deep learning

Dr Aniek Bouwman

Biography:

Dr. Aniek Bouwman is a researcher at Wageningen University & Research. She holds a PhD in Animal Breeding and Genomics. The central theme of her research thus far is using genomic information to explain heritable variation in livestock. She has experience with (meta) genome-wide association studies, imputation of genotypes, and genomic selection for a range of different traits, using different genotyping densities from SNP-chip to whole genome sequence data. Her current challenges are to make optimal use of whole-genome sequence data in animal breeding.

Balanced reciprocal translocations are rare but detrimental events that lead to subfertility. They are usually detected using chromosome staining, i.e. karyotyping. However, they can also be detected in whole-genome sequence data. The detection of such translocations is not straightforward, and bioinformatic tools tend to detect a lot of false positive results, therefore manual inspection of the results remains an essential laborious step. The objective of this study was to investigate deep learning model performance to classify true positive and false positive translocations from images of aligned sequence reads at the position of breakpoints detected by bioinformatic tools.

The aligned sequence data from 39 animals was analyzed for the presence of inter-chromosomal translocations using Delly, resulting in 2992 breakpoints after QC filtering. Integrative Genomics Viewer was used to automatically create images of these breakpoints for visual classification. After visual inspection, 2448 breakpoints were classified as false positive, while 544 showed the correct characteristics of a reciprocal breakpoint (a combination of normal reads and split and discordant reads). A convolutional neural network was trained on 80% of the data of 34 animals and validated on the remaining 20%. A separate test set of 5 animals was used for model testing.

The model had an accuracy of 0.92, a sensitivity of 0.78, and a specificity of 0.96. The specificity is good, but the sensitivity is somewhat low. Wrongly classified breakpoints seemed to require a wider window to properly visualize the breakpoint, hence a wider window size for the image might improve results further.

Comparison of univariate and multivariate models for marker selection and genomic prediction

Ms Charlotte Brault

Biography:

I'm Charlotte Brault, I work in Montpellier at INRAE as a PhD student since October 2018, on genomic prediction in grapevine, in collaboration with grapevine breeders. I have an AgroParisTech master degree in agronomy and plant breeding I did two master internships: the first one (2.5 months) at KWS company in Germany on abiotic stress tolerance in maize, the second one (6 months) at INRAE, in the team I'm currently working with, on genomic prediction in grapevine. I'm also involved in plant breeding teaching at the Institut Agro in Montpellier.

Statistical methods for finding associations between genetic markers and phenotypes have two main outcomes of interest: marker selection (QTL detection) for deciphering genetic architecture of complex traits and genomic prediction for increasing breeding speed. Among them, multivariate (multi-trait) methods may increase predictive ability or marker selection power.

We compared several uni- and multivariate methods for both marker selection and genomic prediction on simulated data with contrasted genetic architectures and correlations between traits. The univariate methods we tested are simple and multiple interval mapping, ridge regression, LASSO and elastic net; the multivariate ones are multi-task group-lasso and multivariate structured elastic net (Chiquet, 2017). Our results indicate that methods using interval mapping always gave the worst results, both to predict genotypes and to select markers. Others perform quite similarly for prediction and marker selection, despite moderate-to-strong genetic correlations between traits, with a 32 % increase in mean predictive ability and a 12 % increase in average power at 5% error rate, compared to interval mapping methods. However, we showed that the largest difference is for multi-task group-lasso in marker selection when analyzed traits had common loci. Thus multi-task group-lasso is interesting to predict and also to select markers.

Based on these results, we analyzed real grapevine data, a dozen traits measured on a phenotyping platform. On average, multi-task group-lasso yielded 7.5 times more QTLs than interval mapping, and 95% of QTLs found by interval mapping were also found by multi-task group-lasso.

An efficient BayesR that has comparable speed to GBLUP and enables large multi-trait analyses

Dr Ed Breen

Biography:

Edmond J. Breen

Currently Senior Research Scientist

Computational Biology Agriculture Research Division Agriculture Victoria. Using big data and high performance computing for data analysis of genomic selection and prediction. Performing image analysis for automated plant phenotyping. Interested in computational techniques for uni and multivariate analysis mixed effects models, image segmentation and mathematical morphology. Has had positions as Lead of Bioinformatics at Australian Proteome Analysis Facility Sydney. Bioinformatics Manager at Eli-Lilly, Head of Informatics Proteome Systems Sydney, and Senior Researcher Image Analysis CSIRO's Division of Statistics and Mathematics.

Bayesian methods of genomic prediction are more accurate but slower than GBLUP. Here we present a new method for implementing BayesR which is 18 times faster than the fastest previously reported EM version of BayesR. This is achieved by processing SNPs in blocks and residual updating. We apply this blocked BayesR to two large data sets of dairy cattle: a single trait analysis of milk, fat and protein yields using 97,000 animals with high density genotypes; and a multi-trait analysis of 20 principal components extracted from milk MIR recorded on 5000 cows. Both analyses find SNPs in common with high association probabilities with these traits.

Model development and estimation of genetic parameters for docility in American Angus cattle

Dr Luiz Brito

Biography:

Amanda Botelho Alvarenga is a PhD student in Quantitative Genetics and Genomics in the Department of Animal Sciences at Purdue University, USA. Amanda has an honors BSc degree in Animal Science from the Federal University of Lavras, Brazil. She completed her MSc in Genetics and Animal Breeding at the "Luiz de Queiroz" College of Agriculture, University of São Paulo, Brazil. She is interested in statistics, quantitative genetics, and genomics applied to genetic improvement of livestock species. Her current PhD project focuses on integrating multiple data sources to maximize the performance of genomic predictions for temperament traits in American Angus cattle.

Behavioral traits has been identified as a key breeding goal in the worldwide cattle industry because it has detrimental effects on animal welfare and performance, handlers' safety, meat quality, and longevity in the herd. Temperament can be improved through genetic selection. However, the success of the breeding program depends on a proper definition of statistical models and accurate estimates of variance components, consequently, accurate estimates of breeding values. In total, 266,029 American Angus cattle with temperament score (1-6 scale) and a 10-generation pedigree depth were analyzed. The non-genetic variables were selected based on a forward model selection method using the R `lm` function. The following effects significantly ($P < 0.05$) affected temperament: calf age deviation from 365 days, if individuals were born from embryos transference, age of dam, and contemporary group (concatenation of date, gender and herd at birth, weaning, and temperament stages; creep-feeding system; if individuals had ultrasound records). Four genetic models (1: reduced model with direct genetic effect; 2: direct and maternal genetic effects; 3: direct genetic and maternal permanent environmental effects; 4: complete model) and variance components estimation were performed using the THRGIBBS1F90 software (500k-1M chain length; 50-70% burn-in; 10 thin). The convergence was evaluated based on the Geweke test (boa R package). The heritability estimates on the liability scale were from 0.38 to 0.44 (0.01) indicating that direct genetic selection for temperament will be very effective. The best-fitted model was the second evaluated (direct genetic and maternal permanent environmental effects) indicating the maternal influences on animal behavioral outcomes.

Identifying Gene Contribution to Neuropsychiatric Disease: PrediXcan X PheRS

Ms Jessica Brown

Biography:

Jessica E.H. Brown has a BS in Biology and Statistics from Elon University. She is currently a fifth year Human Genetics PhD student and a third year Biostatistics masters student at Vanderbilt University in Nancy J. Cox's lab. Her work focuses on neuropsychiatric disorders and the continuum between Mendelian and complex diseases. While the majority of her work uses computational techniques, she also dabbles in some wet lab work, using zebrafish as a model organism.

To identify genes contributing to neuropsychiatric disorders, we combined PrediXcan, for predicting gene expression based on cis-regulatory variation, and phenotype risk scores (PheRS), for quantifying clinical features of Mendelian disease from its associated phenome. PheRS is calculated by adding up the number of phenotypes within a set (for a given Mendelian condition) an individual has, with weights for each phenotype reflecting, for example, prevalence in a population or significance of association of the phenotypic feature with the disease. While PheRS were created for Mendelian diseases, their concept can be more widely used to capture any designated phenotypic constellation. To create a neuropsychiatric phenome, all phenotypes in Vanderbilt's biobank BioVU with the category of neurologic or psychiatric were used. A neuropsychiatric PheRS was calculated for each individual in the 10k PredixVU database (the application of PrediXcan to BioVU). To evaluate the association between PheRS and predicted gene expression, regressions were run using restricted cubic splines with three knots for the predicted gene expression covariate to allow flexibility in the model and additional covariates of gender and the first three principal components. Genes were limited to human orthologs of zebrafish matrisome genes, to test our hypothesis that matrisome genes contribute to risk of neuropsychiatric disease in addition to other medical phenome. Using this method, multiple genes show a high association between neuropsychiatric PheRS and predicted gene expression, including ANGPT2, FBN1, and TNF. Further refinement of targets will include ascertaining associations with multiple early onset phenotypes expected to be observable in zebrafish.

Genotype by fertility season interaction for farrowing rate at first insemination

Ms Annika Bunz

Biography:

Annika Bunz graduated in 2014 with a masters in Agricultural science from Christian Albrecht University of Kiel and since then has been working in Genetics for five years at Rivalea Australia, one of the largest pork producers in Australia; PhD candidate at AGBU University of New England investigating the "Genetic analysis of seasonal infertility in sows and boars" and is the recipient of the postgraduate research scholarship from Australian Pork Limited.

Farrowing rate in pigs is an important reproduction trait impacted by seasonal infertility. Other studies have used seasons based on calendar months or fitted temperature and photoperiod information at mating dates separately to investigate genotype by season interactions. However, the combined effects of temperature and photoperiod experienced by sows around mating have not been used to define trait-specific fertility seasons. Partitioning Around Medoids clusters were used to group mating dates into four fertility seasons according to their similarity in photoperiod and temperature patterns in the time period 35 days prior and post each mating date. A series of bivariate animal models were used to estimate genetic parameters. Heritabilities for farrowing rate were highest and lowest in the most and least stressful fertility season, respectively. Further, genotype by fertility season interactions were found based on estimates of genetic correlations for farrowing rate in different fertility seasons. This methodology to define fertility seasons can be applied to other traits. The implications of these genetic parameters for the design of pig breeding programs that incorporate genotype by fertility season interactions should be evaluated.

A meta-GWAS reanalysis of twenty years of quantitative traits in maize

Miss Merritt Burch

Biography:

Merritt Burch is a third-year Ph.D. student in the Buckler Lab at Cornell University in the section of Plant Breeding and Genetics. Merritt's research investigates the extent to which pleiotropy controls phenotypes within the maize genome as well as the role of transposable elements in the regulation of gene expression.

Here we remap, via GWAS, phenotypes from across maize populations to find the features that characterize complex traits. Past mapping experiments have been difficult to compare to each other because they have focused on one population, utilize different genome versions, and have only analyzed a few traits. In this study, we collected thousands of previously published morphological, metabolic, and expression phenotypes across maize landraces (n = 5000), public inbred lines (n = 2815), and half-sib lines (n = 5000) in multiple environments. We then mapped this dataset in a meta-GWAS using consistent model parameters. Due to the high genetic heterogeneity within and between these populations, we use a combination of global and local principal components to control for linked loci arising from population structure and kinship. We hypothesize that the properties of these mapped variants can predict loci that have a functional impact, uncover regulatory networks of gene activity, reveal genotype by environment effects, and uncover loci showing statistical pleiotropy. We will compile these results in an online, queryable database for maize geneticists and breeders to mine for associations of interest for gene prioritization and gene editing.

Using statistical and crop growth models to characterize genotype by environment interaction over time

Dr Daniela Bustos-Korts

Biography:

Daniela Bustos-Korts works as a researcher at Wageningen University (NL). She graduated with a BSc in Agriculture and an MSc in Crop Physiology (Universidad Austral de Chile). In 2017, she obtained her PhD from Wageningen University. Her PhD thesis was about the use of statistical and crop growth models for phenotype prediction across multiple environments. Her research interests are in the development of strategies that combine statistical and crop growth models to help breeders designing effective phenotyping and prediction schemes. These strategies involve the use of crop growth models like APSIM and linear-mixed models for multi-trait and multi-environment genomic prediction.

Climate, soil and management typically change across environments, inducing genotype-specific responses that might lead to heterogeneous genotypic ranking for the target trait, e.g. yield. These genotype-specific responses are an expression of genotype-by-environment interaction (G×E). G×E for a complex target trait is an emerging property of agricultural systems that results from the interplay between a hierarchy of secondary traits involving the capture and allocation of environmental resources during the growing season. Traits underlying yield differ in their contribution to adaptation across environmental conditions and have different levels of G×E. The hierarchy of traits underlying yield ranges from basic traits that correspond to response mechanisms/sensitivities and are less prone to G×E, to intermediate traits that integrate a larger number of processes over time, showing a larger amount of G×E. Here, we present a framework to study the performance of genotype to phenotype (G2P) modelling approaches for systems exhibiting G×E. We generate response surfaces, or adaptation landscapes, for yield and yield related traits, emphasizing the organisation of the traits in a hierarchy and their development and interactions over time. We use the crop growth model APSIM-wheat with genotype-dependent parameters as a tool to simulate non-linear trait responses over time with complex trait dependencies and apply it to wheat crops in Australia. We illustrate how such simulated data can be used to evaluate statistical genotype-to-phenotype models for multiple traits and environments and to characterize relationships between traits over time and across environments, as a way to identify traits that confer specific adaptation

A genome-wide methylation study of body fat traits in the Norfolk Island Isolate

Mrs Thao Van Cao

Biography:

After working as a research assistant in a biostatistics group at Oxford University Clinical Research Unit in 2017 in Vietnam, Van was awarded the Australian Government Research Training Program scholarship to commence an MPhil at Queensland University of Technology and became a member of the Genomics Research Centre (GRC) in 2019. Van's project is to identify genes and loci that are associated with body fat traits and is continuing to expand her works on understanding epigenetics aging and the heritability of loci in relation to body fat traits.

Background: While several studies have been investigated obese cohorts and found over 500 genes associated with obesity in both genome-wide and epigenome-wide association studies, it is still unclear how epigenetic profiles are associated with body fat traits among adult subjects. To address this, we analysed DNA methylation profiles of 48 healthy participants (24 males and 24 females) from the isolated Norfolk Island.

Method: Six body fat traits were collected carefully including body fat percentage, body mass index, hip circumference, waist circumference, waist-hip-ratio and body weight. Peripheral blood samples were collected for DNA methylation array measurements. We used Principle Component Analysis (PCA) to capture covariance among phenotypes. Then, stepwise linear regression was applied to identify differentially methylated positions (DMPs) at genome-wide significance ($P \leq 2.4 \times 10^{-7}$). Gene set enrichment analysis (GSEA) was performed in genes at suggestive DMPs ($P \leq 1.0 \times 10^{-4}$).

Result: Two components covering 89% of the phenotypic variance were identified by PCA. Five significant DMPs, which mapped to GOT2-APOOP5-CDH8, LYSMD3, HIBADH, ADGRD1 (GPR133)-LINC01257 and EBF4 genes, were found in the total cohort. In addition, Cadherin (28 genes, $P_{adj} = 6.76 \times 10^{-7}$) and Wnt signaling pathways (38 genes, $P_{adj} = 7.78 \times 10^{-6}$) were enriched from 848 genes containing suggestive DMPs through GSEA.

CONCLUSION: This study identified epigenetically influenced genes and pathways that potentially involving in natural body fat mechanism in healthy adults and provide targets for consideration in future studies.

Accuracy of Imputation to Whole-Genome Sequence in Nelore Cattle

Dr Roberto Carvalho

Biography:

Dr Roberto Carvalho is a Researcher in Animal Breeding and Genetics at the Department of Animal Science of São Paulo State University (Unesp), Jaboticabal Campus, SP, Brazil. Roberto has a good connection with the beef industry as he had worked for 10 years (2004-2013) at GenSys as a consultant for important commercial beef cattle breeding programs in Brazil. For the last 7 years, working as a researcher, he has authored and co-authored more than 80 peer-reviewed papers, focused mainly on quantitative genetics and genomic studies applied to livestock and aquaculture breeding (<https://orcid.org/0000-0002-4506-0555>).

Advances in next-generation sequencing techniques associated with a drastic decrease in sequencing costs is favoring the development of strategies to explore the complete DNA sequence in genetic evaluations, theoretically including causal mutations and allowing better genomic predictions. However, cost is still a limitation to sequence a large number of animals. An alternative and cost-effective strategy would be to sequence key ancestors of the population, and to impute the genotypes, of animals genotyped with SNP arrays, to the whole-genome sequence. Our research group has a database with about 10,000 Nelore animals, from commercial breeding programs, genotyped with Illumina SNP arrays with different densities. Using the sequence data of approximately 150 influential Nelore bulls, we will carry out the imputation of these genotypes to whole genome sequence, which will increase the amount of genomic information per animal. A previous investigation of the imputation accuracy is essential to determine the feasibility of this process. This issue is being investigated in the present study, in order to assess the accuracy of imputation to whole-genome sequence in Nelore cattle, and to develop a pipeline to impute sequence data of our existing genotype database. Acknowledgments: São Paulo Research Foundation (FAPESP, grants 09/16118-5, 16/19514-2, 17/10630-2, 18/10109-3 and 19/12434-1) and Coordination for the Improvement of Higher Education Personnel - Brasil (CAPES - #001).

Allelic bias and genetic mapping of differentiation potential revealed by single-cell transcriptomics in recombinant hybrid mouse embryonic stem cells

Dr Frank Chan¹, Mr Moritz A. Peters¹, Mr Volker Soltys¹, Ms Insa Hirschberg¹, Dr Marek Kucka¹

¹*Friedrich Miescher Laboratory of the Max Planck Society*

Self-renewal and pluripotency are twin hallmarks of embryonic stem cells. Every cell in an early embryo must balance between differentiation and proliferation, often involving irreversible fate decisions. However, there are very few studies that directly investigate how these gene pathways evolve between species, because crosses between species usually fail. Here, we combine in vitro recombination (IVR) and single-cell transcriptomics in F1 hybrid embryonic stem (ES) cells to determine the genes and loci influencing the decision between renewal vs. differentiation. Through RNAi suppression of Blm helicase, we effectively produced recombinant ES cells between the laboratory C57BL/6N mouse and *Mus spretus*. We then induced differentiation and obtained single-cell transcriptomes on ~20,000 IVR cells. Using reference scRNAseq datasets from naïve and differentiated non-recombinant cells, we show that in vitro recombination boosted genotypic and expression diversity. We found that the genetic reshuffling under IVR produced cells that can sustain robust Nanog stemness marker expression despite differentiating treatments. Moreover, we observed that allele-specific expression increased with differentiation, possibly from recombinant cells. We also found that SPRET alleles were enriched in highly differentiated cells. Our work shows that by combining in vitro recombination and single-cell assays, it is now possible to directly map genes controlling otherwise inaccessible cellular phenotypes across species.

The nature of variation from spontaneous mutations in mammals – a large-scale mutation accumulation experiment in four strains of mice

Dr Jobran Chebib

Biography:

I am an ERC Postdoctoral Research Fellow at the University of Edinburgh in Peter Keightley's lab. My curiosity has led me to engaging in fundamental research that tackles questions about the genetic mechanisms that facilitate and hinder evolution. More specifically, I study genetic architectures using both theoretical models and empirical data, which are both necessary to understand the complexities of biological systems, including pleiotropy, linkage and mutation.

Understanding how new variation arises in quantitative traits each generation through spontaneous mutation is an important goal in evolutionary genetics. Yet little is known about the extent of new variation in mammals even though there is evidence that it may be as much as 1% per generation. In order to tackle this issue we have started the first large-scale mutation accumulation (MA) experiment in mammals with 4 inbred mouse strains. Each strain has been maintained through full-sib mating to create a large number of MA lines and phenotypic data has been collected from each individual including measurements on fertility, viability, body weight, and tail length. Whole genome sequencing has also been performed on progenitor individuals as well as eighth-generation descendants. Using this data we will present average mutation rates at the nucleotide level, including variance in mutation rates within and among strains, and average changes in phenotypic variation arising in populations from new mutations per generation. We will be able to compare the observed molecular variation in the inbred lines to theoretical expectations of pairwise nucleotide diversity per site at mutation drift-balance, and to compare observed phenotypic changes to previous estimations. The results of this study will help to illuminate the processes and effects of spontaneous mutations in populations and contribute to answering fundamental questions in evolutionary genetics.

Improving analyses of GWAS summary statistics by detecting data heterogeneity and errors

Dr Wenhan Chen

Biography:

Wenhan here. I am an Early Career Researcher at the University of Queensland. My interests lie in algorithm design, software development and the application of the two in biotech and diseases.

Summary data publicized from large-scale genome-wide association studies have facilitated the development of various summary-data-based methods. However, analyses using these methods can suffer from biases caused by low data quality and data heterogeneity when multiple data sets are used. Here a quality control method is proposed, which leverages linkage disequilibrium among SNPs to detect and eliminate errors in GWAS data and heterogeneity between data sets to improve summary-data-based analyses. We showed by simulation that our method substantially reduced the false positive rate in detecting secondary GWAS signals in summary-data-based conditional and joint association (COJO) analysis, enabling the application of COJO to rare variants. We further showcased that our method as a QC step can improve heritability estimation and putative causal gene discovery. The method has been implemented in a freely available software tool DENTIST.

Exact Distribution of Linkage Disequilibrium in the Presence of Mutation, Selection or Minor Allele Frequency Filtering

Professor Hao Cheng

Biography:

Dr. Hao Cheng is an assistant professor of Quantitative Genetics in the Department of Animal Science at University of California, Davis. He got his Ph.D. in Genetics and Statistics at Iowa State University. His research interests are broadly involved in the development of statistical and computational methods for the genetic improvement of populations in agriculture through more accurate and efficient genetic analysis. He has focused on the use of phenomics, genomics, pedigree, and other sources of big data in various species to better predict a wide variety of traits.

Linkage disequilibrium (LD), often expressed in terms of the squared correlation (r^2) between allelic values at two loci, is an important concept in many branches of genetics and genomics. Genetic drift and recombination have opposite effects on LD, and thus r^2 will keep changing until the effects of these two forces are counterbalanced. Several approximations have been used to determine the expected value of r^2 at equilibrium in the presence or absence of mutation. Here we propose a probability-based approach to compute the exact distribution of allele frequencies at two loci in a finite population at any generation t conditional on the distribution at generation $t-1$. As r^2 is a function of this distribution of allele frequencies, this approach can be used to examine the distribution of r^2 over generations as it approaches equilibrium. The exact distribution of LD from our method is used to describe, quantify and compare LD at different equilibria, including equilibrium in the absence or presence of mutation, selection, and filtering by minor allele frequency. We also propose a deterministic formula for expected LD in the presence of mutation at equilibrium based on the exact distribution of LD.

Genotyping strategies for rainbow trout breeding programs affected by genotype by environment interactions

Dr Think Tuan Chu

Biography:

I am Think Tuan Chu, a postdoctoral researcher at Center for Quantitative genetics and Genomics, Aarhus University. My major interest is breeding planning for chicken, pigs, and fish.

Selective genotyping of phenotypically elite animals may lead to bias and lowered accuracy of GEBV. Selective genotyping selection candidates based on phenotypes measured in the breeding (B) environment is not necessarily a good strategy when a breeding program is to improve animals' performance in the commercial (C) environment. This simulation study compared random versus selective genotyping of selection candidates in a breeding program for rainbow trout when genotyping efforts were limited, and genotype-by-environment interactions existed. The selective genotyping strategy for B fish was done by selecting an individual with the best phenotype in each random sample of 20 fish. A number (varied in different scenarios) C fish were randomly genotyped and phenotyped for the reference population. Pedigree of non-genotyped individuals were not registered. Heritability of the trait was 0.3. The genetic correlation (r_g) between the performances measured in B and C was 0.2, 0.5 or 0.8. It was found that the selective genotyping strategy led to significantly higher genetic gain than random genotyping selection candidates. The difference in genetic gain between the two strategies increased with increasing r_g . Interestingly, with r_g of 0.2 and 0.5, selective genotyping did not lead to bias of GEBV of performance in C, while with r_g of 0.8 the variance of GEBV was deflated slightly. In conclusion, selective genotyping of selection candidates is still the recommended strategy in breeding programs for rainbow trout when a limited number of individuals can be genotyped, even if the selection of genotyping is based on a correlated trait.

A joint use of pooling and imputation for SNP genotyping

Miss Camille Clouard

Biography:

Camille Clouard is a PhD student at Uppsala University, Sweden. Early as a high school student, she got the opportunity to conduct an internship in Plants Genetics at the INRAE (National Research Institute for Agriculture, Food and Environment, France), which motivated her to enter and later graduate from AgroParisTech (Paris Institute of Technology for Life, Food and Environmental Sciences, France) where she discovered her enthusiasm for computational science applied to biology-specific challenges.

The latest decade's development biotechnologies have substantially cut the costs of DNA sequencing. Nevertheless, when performing for example genotyping on large data sets i.e. with thousands of individuals in a cost-sensitive context, whole-genome or targeted sequencing can remain expensive. Using techniques from the field of group testing theory has been performed on single-marker PCR genotyping in the 1990s. By implementing pooling strategies, several studies have shown it is an effective and accurate means for reducing DNA processing costs. However, few works specifically consider pooling samples on SNP-arrays for genotyping purposes.

Pooling samples results though in a loss of information about individual genotypes. The uncertainty introduced by pooling appears as missing individual genotypes. Computational methods for genotype imputation have been widely used for dealing with missing genotypes. They mainly implement iterative procedures in a Bayesian statistical framework for determining the most likely genotypes at unassayed markers. Traditionally, imputation infers additional markers states in a study population genotyped at low density by using a reference population with similar genetic structure which was genotyped at high density. We present tentatively promising results for combining a pooling scheme for SNP genotyping with computational genotype imputation both in humans and important foodcrop species. We also suggest how modelling the possible outcomes for uncertain pooled genotypes can inform the imputation algorithms further in order to fully and accurately reconstruct the genotypes of all individuals within each pool, while using less than half the number of assays needed for non-pooled genotyping of each individual.

Investigating the relationship between imputation accuracies and relatedness

Natalie K Connors, Mohammad H. Ferdosi

Biography:

Natalie has been involved in the application of genomics data for Australia's beef industry for the past 5 years, including poll genetics research and data quality control for BREEDPLAN genetic evaluations.

Single Step Best Linear Unbiased Prediction (ssBLUP) is used in the Australian beef industry's genetic evaluation, BREEDPLAN, for the prediction of Estimated Breeding Values (EBVs), which uses genomic information in a Genomic Relationship Matrix (GRM). Imputation of missing Single Nucleotide Polymorphisms (SNPs) and imputation of low density to high density genotypes is essential to combine various SNP densities for building the GRM. EBV accuracies are dependent on an individual's relationship to the rest of the population. Similarly, a target population's imputation accuracy is dependent on relatedness to the reference population. This study introduces a 'relatedness score', calculated in a similar way to EBV accuracies, to indicate an animal's relatedness to the reference population. The objective of this study was to identify how well the relatedness score can predict imputation accuracies. For this purpose QMSim was used to simulate 10 generations of genotypes (20 chromosomes and 2000 SNPs – 200 cM) with 40 males and 800 females in the historical population. Generations 4 to 10 were used to evaluate the relationship between imputation accuracies and relatedness score. The results demonstrated a non-linear correlation between imputation accuracies and relatedness score when individuals exist across multiple generations and with densities greater than 1000 SNPs were used. These results indicate a relatedness score may explain low EBV accuracies and EBV instability within BREEDPLAN data, due to low imputation accuracies.

Understanding the genetics of fertility and temperament in Northern beef cattle using genomic technologies

Mr James Copley

Biography:

PhD candidate at QAAFI, Centre for Animal Science. Exploring the genetics of beef cattle fertility in northern Australia.

Fertility and temperament of beef cattle are important drivers of productivity in the northern beef industry. Fertility of breeding females is measured by their reproductive performance, the ability to reproduce annually being optimal. Temperament is similarly important, poor temperament is a danger to handlers and incurs higher management costs. The purpose of this study was to use genomic prediction to estimate heritability of fertility and temperament traits and to probe the genetic correlation between them. The study analysed data from *Bos indicus* influenced heifers in the Beef CRC and Smart futures projects. Phenotypic data consisted of fertility trait records (PPAI, AGECL and Tscore) across three separate datasets, Brahman (n=936), Tropical composite (n=1097) and Smart Futures (n=3696), and temperament was measured as Flighttime (n=4645). Genotypes were imputed up to 728k SNP density, and GBLUP results were obtained using GCTA. Heritabilities were 0.56(0.08), 0.37(0.08), 0.44(0.11), 0.24(0.08), 0.19(0.03) and 0.33(0.03) for AGECL (Brahman), AGECL (Tropical composite), PPAI (Brahman), PPAI (Tropical composite), Tscore (Smart Futures) and Flighttime respectively. Genetic correlations were -0.01(0.10), -0.06(0.11), -0.03(0.07) and 0.32(0.09) between AGECL and Flighttime, PPAI and Flighttime, Tscore and Flighttime and PPAI and AGECL respectively. Genetic correlation was moderate between AGECL and PPAI but low to negligible between Flighttime and all fertility traits. The results suggest that while the individual traits are sufficiently heritable to respond to direct selection, temperament and fertility appear to be independently heritable and selection for improved temperament could not be used to indirectly select for improved fertility.

An Explicit Model to Estimate Narrow Sense Heritability in ‘Chimeric’ Organisms

Dr Andrés J. Cortés

Biography:

Dr. Andrés J. Cortés holds an Associate Research position as Geneticist at the Colombian Corporation for Agricultural Research (AGROSAVIA). He graduated as Plant Geneticist (PhD) from Uppsala University (Sweden), and as Biologist (BSc Hons, MSc) from Universidad de los Andes (Colombia). His research experience dates back to the International Center for Tropical Agriculture (CIAT), the University of Fribourg (Switzerland), the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL SLF), and the Swedish University of Agricultural Sciences (SLU). Dr. Cortés has investigated the genetic adaptive potential in plants and trees of agro-ecological interest using genomic, evolutionary, and ecological tools.

How different genetic backgrounds interact to shape a unique phenotype has been one of the most pervasive questions in quantitative genetics and molecular evolution. Horizontal gene transfer and allopolyploidy are often regarded as the typical processes that lead to the interaction of various genomes within a single organism. However, a commonly disregarded yet ancient process that also produces genetic chimeras is grafting, which joins the root system (rootstock) of one genotype to the shoot (scion) of another. Despite grafting is a common practice across a phylogenetically diverse array of plant species, an explicit estimation of rootstock effects (i.e. rootstock-mediated heritability – h^2) is lacking. Therefore, here we expanded a marker-based ‘genetic prediction’ model, which relies on a linear predictor to estimate the additive genetic contribution to phenotypic trait variation, to a grafted clonal species with seedling rootstocks of mixed ancestry. Model’s significance was assessed using its predictive ability (r) and two randomization strategies, of the phenotypic vector and the markers’ matrix. A first validation of this model was carried out using a reference dataset of 20 traits measured during three years in 240 grafted fruit trees from eight plantations. False-negative rate was low because only five traits, known to be influence by the rootstock, exhibited significant rootstock-mediated heritability estimates from 0.32 to 0.46 and model fits (r) from 0.60 to 0.74. Further validation datasets will be desirable. Explicit estimation of trait heritability in grafted systems will ultimately speed-up tree breeding programs while help understating the phenotypic consequences of genetic ‘chimeras’.

Improvement of genomic prediction accuracy for residual feed intake by prioritizing genetic markers identified by genome-wide association and gene expression

Dr Sara de las Heras Saldana¹, Dr Nasir Moghaddar¹, Dr Samuel A. Clark¹, Professor Julius H. J. van der Werf¹

¹*University of New England*

Genomic selection strategies applied to complex traits like residual feed intake (RFI) could improve the selection for feed efficiency in livestock. In the last decade, more information about the underlying genetic architecture for traits like RFI have been obtained through genomic wide association studies (GWAS) and transcriptomic studies. The aim of this study was to test whether combining information from GWAS and gene expression significantly associated (GSA) results could improve the accuracy of genomic prediction. We evaluated the gain in accuracy of prediction of RFI in 2190 Angus steers using medium-density and high-density SNP panels (770k) and by adding pre-selected SNPs (top SNPs) most significant in GWAS and close GSA from a gene expression. Two cross-validation designs were compared, one where the same dataset was used for training and GWAS discovery (4CV) and one where the discovery was separated from the training set (4x4CV). There was no improvement in prediction when using 770k compared with the medium density SNP panel. The 4x4CV design increase in accuracy by 1.2 and 2.7 percent point when top-SNPs ($-\log_{10}(P)=3.5$) were used, compared to using only 50k or 770k, respectively. The 4CV design showed lower accuracy when using top SNPs and the predictions were much more biased. The use of top SNPs in combination with selected SNPs located inside GSA reduced the bias in prediction compared with using only top SNPs in 4x4CV and slightly increased the accuracy of prediction. Genomic prediction accuracy can be improved when using selected SNPs from GWAS and GSA.

Progenies selection for resistance to *Fusarium verticillioides* in maize crop

Dr João C. De Souza¹, Dra. Paula Vieira, Dra Rafaela Guimaraes, Dr Renato Vasconcelos, Dr Brena K.S. Silva, Dr. Flávio Henrique Medeiros

¹Ufla, ²Ufla, ³Ufla, ⁴UFLA, ⁵UFLA, ⁶UFLA

Fusarium ear rot is a disease in maize crop that affects yield, reduce nutritional and phytosanitary quality of the grains. The aimed was investigate the relationship between rot grains and *F. verticillioides*. The experiments were conducted at UFLA, Brazil. Segregating populations (F2:3, RC11, RC12) and F1 were obtained from the cross between L75 (resistant) x L43 (susceptible) lines. In the 2017/2018 crop season, all generations were inoculated and were subjected to the blotter test. For the evaluation of severity and incidence, a diagrammatic scale was used (0 at 5). With the weight of 100 grains data, was performed the correlation between the incidence and weight of 100 grains and the principal component analysis (PCA) was obtained by the biplot chart. The incidence, severity, healthy grain weight and rot grains weight data were compared by PCA. Pearson's correlation between incidence and weight of rot grains was not significant, so it is not possible to conclude that the correlation between these characters is not zero. PCA shows that principal component PC1 and PC2 account for 32.33 and 24.55% of the variation. The first component shows strong associations between healthy weight and total weight. The first component also has associations for severity and incidence. The PCA graph indicates that there is no direct correlation between total weight and incidence, this is also supported by the correlation results, indicating that the disease incidence is not directly related to production. Therefore, no inferences can be made about resistance to *F. verticillioides* from rot grains results.

Genetic improvement program for goats in India: Experiences and impact

Dr M S Dige

Biography:

Dr M S Dige is working as a scientist in ICAR-Central Institute for Research on Goats, Makhdoom. He has specialized in Animal Genetics and breeding, is currently working in a field of Animal breeding and Genetics. He has contributed significantly in the conservation and improvement of Jamunapari and Barbari goat, two most important goat breeds of north India. He is actively involved in All India co-ordinated Research Project (AICRP) which is a long term program to bring upon genetic improvement and conservation of goat genetic resources of India.

All India Co-ordinated Research Project (AICRP) on Goat Improvement is a long term programme aimed to bring upon genetic improvement and conservation of goat genetic resources of the country in their natural habitat. Presently, AICRP on goat improvement is working with 15 registered goat breeds and five local genotypes in ten different eco-zones ranging from humid to semi-arid and cold arid to hot arid. The major thrust of the project is to build up long term capacities of goat keepers through introduction & dissemination of improved genes(bucks), technology transfer, creation of knowledge base, application of health management practices for enhancing production potentials on sustainable basis. This is operational at 461 villages and also working in more than 35 tribal villages and contributing for a better livelihood in the tribal region. The performance recording was carried out in 62784 animals. Jamunapari, Barbari and Sirohi nucleus herds are working as best model for in- situ conservation in the natural home tract of the breed. The AICRP have significantly produced and distributed 1022 improved animals in a year to farmers for breed improvement as well as up-gradation of local germplasm. The programme has significantly contributed by increasing fecundity, population growth, milk production, body growth and reducing morbidity and mortality. Technological interventions under the project have benefited more than 2277 goat rearing families in different production system over 15 states of the country. It has provided average employment ranging from 80 to 140 man days and has improved income of farmer's significantly.

Diagnostics for genotypes from low-depth sequencing

Dr Ken Dodds

Biography:

Ken Dodds is a member of the Bioinformatics and Statistics Team in AgResearch, a New Zealand crown research institute. Ken's research is focused on the utilisation of genomic information for genetic improvement in the primary industries. Current projects are genomic selection in sheep, deer and goats and analysis methods for genotyping by sequencing data for use in livestock, forage, aquaculture, conservation and ecology.

Advances in genomic science have enabled the characterisation of genotypes using sequencing technology. In many applications it is sufficient to genotype only a proportion of the genome, meaning that restriction enzyme-reduced representation sequencing methods can be used. Statistical methods have been developed that allow the use of uncertain genotypes, specifically genotypes supported by few sequence reads. These developments enable relatively low-cost studies of any species including those where there is currently little genomic knowledge. As there are many steps (both lab-based and computational) involved in generating the genotypic information, it is important to implement diagnostic checks of workflows and assumptions used in analyses. A set of mostly graphical diagnostics is presented. These help diagnose 1) contamination from different or the same species, 2) different individuals being labelled as being the same, 3) sequence reads from duplicated regions erroneously assigned to the same position, 4) non-binomial sampling of alleles for sequencing, and 5) non-optimised lab protocols. These diagnostics are available in scripts in the AgResearch GitHub repository (<https://github.com/AgResearch>).

Predicting the purebred-crossbred genetic correlation from genetic variances within, and covariance between parental lines

Mr Pascal Duenk

Biography:

Pascal Duenk has recently obtained his PhD degree at Wageningen University and Research, and is now a researcher and lecturer at the Animal Breeding and Genomics group. His PhD research was focused around the use of data on crossbred animals for genomic prediction. In addition, Pascal worked on the impact of non-additive effects on genetic correlations between populations, and between purebreds and crossbreds. Currently, his research interests include quantitative genetics, machine learning, and genomics in animal breeding.

Breeders benefit from heterosis and breed complementarity by mating individuals from genetically distinct purebred lines to produce crosses. While the aim of such breeding programs is to improve crossbred (i.e. hybrid) performance, selection usually takes place in the purebred lines based on purebred performance. The response to selection in crossbred performance therefore depends on the genetic correlation between purebred and crossbred performance (r_{pc}). The r_{pc} can be lower than one due to non-additive effects in combination with differences in allele frequencies between parental lines. Hence, this suggests that r_{pc} can be expressed as a function of parameters in the parental lines. In this study, we derive expressions for r_{pc} based on genetic variances within, and genetic covariance between parental lines. These expressions were derived for a genetic model with only additive and dominance effects (A+D), and only additive and epistatic additive by additive effects (A+AA). We validated our expressions using simulations of purebred lines that were either positively or negatively selected, and their crosses. Finally, using these simulations, we investigated the value of r_{pc} for more complex genetic models (e.g. with dominance and complementary or multiplicative epistasis), for which expressions could not be derived. Results show that our expressions provide exact estimates of r_{pc} for models A+D and A+AA, and accurate upper and lower bounds of r_{pc} for more complex models, using information of parental lines only. In conclusion, our work demonstrates the impact of non-additive effects on r_{pc} , and aids in determining bounds of r_{pc} for different genetic models.

Leveraging genotype-by-environment interactions across discrete climate regions to select more sustainable beef cattle

Ms. Harly J Durbin¹, Mr. Troy N Rowan¹, Dr. Jared E Decker¹, Dr. Stephen P Miller²

¹University Of Missouri Animal Sciences, ²Angus Genetics, Inc.

Biography:

Harly Durbin is a Ph.D. student in the Division of Animal Sciences at the University of Missouri - Columbia. Her research interests include using quantitative and population genetics methods for livestock improvement, particularly environmental adaptation and sustainability

Compared to other livestock systems, beef cattle are produced in a wide range of climates with minimal management interventions. While environmental heterogeneity is a strength of the beef cattle industry, it also increases the potential for genotype-by-environment interactions. This may present issues for sustainable beef production, especially for breeders in stressful environments purchasing genetics from outside their region.

Random regression models using a single environmental variable may not capture all stressors unique to a region (i.e., local pathogens and forages) and their interactions. Therefore, we generated environmental region-specific genetic predictions. Using k-means clustering on 30-year-normals for temperature, rainfall, and elevation, we assigned growth trait records from cattle registered in the American Angus Association to one of 7 discrete climate regions based on herd zip code.

Data was collected by American Angus breeders from 1973-2018 and included 7,080,318 weaning weight and 3,468,487 post-weaning gain measurements. We then used univariate & multivariate restricted maximum likelihood models to calculate breeding values and genetic correlations between regions for weaning weight and post-weaning gain. We find that while genetic trends are similar between regions, there is significant re-ranking of breeding values. Genetic correlations across regions ranged from 0.67 to 0.82 for weaning weight direct, 0.64 to 0.86 for maternal milk, and 0.63 to 0.91 for post-weaning gain. In the future, these approaches could be incorporated into genetic evaluations to help beef cattle producers identify animals better suited to their environment.

Comparison of genomic prediction methods for longitudinal female fertility in Brahman cattle

Dr Bailey Engle

Biography:

Dr. Engle's research focus has centred on the understanding and improvement of complex production traits in beef cattle, using innovative genomic technologies. Combining a practical, first hand knowledge of beef cattle production with robust experience in applied quantitative genetics, Dr. Engle has crafted a niche for herself in the area of genomic selection in tropically adapted beef cattle, with particular emphasis on female fertility and cow productivity traits. Her current work is developing genomic breeding values for novel cattle phenotypes, such as stayability, longevity, and heifer maturity, with the specific target of eventual commercial adoption by cattle breeders.

Yearly cow fertility is a primary determinant of profitability in beef production. This lowly heritable trait is inordinately impacted by variable environmental conditions, making improvement difficult. Given this, there is a heightened interest from industry to develop selection tools that may assist the selection of breeding animals for improved female fertility. Therefore, the objective of this project was to explore methodologies for the genomic prediction of longitudinal beef cow fertility. In this study, Brahman cows ($n = 4395$) that had at least 2 sequential calving records beginning as a heifer (2 or 3 years) and up to 18 years of age were used for prediction of female fertility. All animals were pedigreed ($n = 35670$) and some ($n = 3759$) genotyped at 36k SNP density and imputed up to approximately 728k SNP. Variance components and EBVs were estimated using ssBLUP in the blupf90 software suite. Breeding values for stayability, or a threshold measure of productivity to a set age, were univariately estimated starting at 4 years of age to represent heifer fertility, and at older ages to represent sustained cow longevity. Alternatively, lifetime number of calves was used to represent a singular measure of a cow's reproductive longevity. Heifer fertility and sustained lifetime fertility are separate, but highly interrelated and correlated traits. Here we present a comparison of prediction schemes that take into account early- and late-in-life fertility both independently and in combination.

No evidence of genetic relationship between theta/beta ratio and Attention Deficit Hyperactivity Disorder

Mr Geoffrey A English

Biography:

Geoffrey is a 3rd year PhD student at QIMR Berghofer Medical Research Institute, Herston, Qld, Australia. He is studying the genetic and environmental heterogeneity influencing the heritability of ADHD subtypes.

Background: ADHD is a heritable childhood disorder with three subtypes (inattention, hyperactivity/impulsivity and combined) with a complex and subjective diagnostic procedure. In 2013, the U.S. Food and Drug Administration approved the use of an EEG device (NEBA) which estimates the ratio of the theta/beta power spectrums (TBR) at the Cz electrode, to aid ADHD diagnoses. However, the test is expensive (\$475) and studies examining its efficacy are highly conflicting. Furthermore, no formal test for a genetic link between TRB and ADHD has been performed. The present study estimated the phenotypic and genetic relationships between ADHD across all subtypes and the TBR using both eyes open (EO) and closed (EC) EEG recordings.

Methods: In an Australian adolescent twin and family cohort using GREML we estimated the relationship between TBR and ADHD subtypes (using three measures: SWAN mothers (Neo=515, Nec=636) self-report (Neo=581, Nec=722) and ASRS (Neo=370, Nec=454)); and between TBR and a Polygenic risk score (PRS; Neo=838, Nec=1036) calculated from the recent PGC case/control ADHD meta-analysis (Demontis 2018). All analyses controlled for age and sex.

Results: Despite being highly heritable (EO - $h^2 = .862$, SE = .016; EC - $h^2 = .891$, SE = .021) across TBR measures, analyses found no evidence of any relationship between TBR and ADHD subtypes or between TBR and ADHD PRS scores.

Conclusions: Given the cost of these tests and the failure here to detect any phenotypic or genetic relationship with ADHD, more research is recommended to justify using TBR as a diagnostic instrument for ADHD.

Clinical and Genetic Heterogeneity in Type 2 Diabetes

Ms Annika Faucon¹, Dr Xue Zhong¹, Dr Lea K. Davis¹, Dr Nancy J. Cox¹

¹*Vanderbilt University Medical Center*

Diabetes is a complex, heterogeneous disease with many clinical comorbidities. The extent to which genetic heterogeneity drives clinical heterogeneity in Type 2 Diabetes (T2D) is not known. In this set of experiments, I compare clinical risk conferred by genetic risk from three clusters of T2D risk variants. I constructed the three genetic risk scores (GRSs) for individuals in Vanderbilt's integrated EHR biobank (BioVU). I ran logistic regressions to calculate the change in probability of having each diagnostic code group (termed phecodes) for each one standard deviation increase in genetic risk for each of the three GRSs, after accounting for important covariates.

Liver disease and acute renal failure phecodes were significantly associated only with the BMI-Dyslipidemia risk pathway (OR=1.14, P=8e-8) and (OR 1.11 P=1e-7), respectively, not with Insulin Secretion (OR=0.99, P=0.566) and (OR=1.00, P=0.874) nor Insulin Action pathways (OR=1.07, P=0.002) and (OR=1.05, P=0.015), respectively. Use of BMI as a covariate removed signals of statistical significance (Bonferroni p-value= in associations between the BMI-Dyslipidemia GRS and blood glucose (OR=1.043, P = 6.7e-6) and blood urea nitrogen (OR=1.040, P= 1.6e-5), but adjustment for BMI did not bring significant p-values closer to 1 for the insulin secretion nor the insulin action GRS - lab associations. This indicates that BMI may be a more useful measurement to predict downstream risk in participants with T2D risk sourced from the BMI-dyslipidemia pathway.

Increasing the goodness-of-fit of genomic prediction model with addition of maternal genomic relationship matrix.

Dr Mohammad Ferdosi¹, Dr Natalie Connors¹ and Dr Majid Khansefid²

¹ Animal Genetics and Breeding Unit, University of New England, Armidale, Australia

² AgriBio Centre for AgriBioscience, Agriculture Victoria, Bundoora, Australia

Biography:

I am a research fellow at Animal Genetics and Breeding Unit. I received my PhD degree in animal breeding from the University of New England in 2016. During my PhD I developed HSPHase that has been submitted to The Comprehensive R Archive Network. I have been working on imputation, haplotype inference and their applications in animal breeding. More recently, I am working on implementation of genomic pipeline for single step BLUP. The genomic pipeline has been implemented as part of BREEDPLAN which is Australia national beef genetic evaluation.

Genomic prediction models use a genomic relationship matrix (GRM) to quantify relationships. However, the GRM relationships do not distinguish between the maternal and paternal origins, thereby ignoring parent-of-origin effects such as imprinting. Genomic imprinting is a phenomenon whereby gene expressions within progeny are varied based on the parental origin of haplotypes or alleles, generally due to epigenetic effects such as DNA methylation. Genomic imprinting has been reported for many economically important traits in livestock such as weight. In this study, we explored the effect of fitting a maternal and/or paternal genomic relationship matrix (GRM), in addition to a regular GRM, on the goodness-of-fit of a genomic prediction model for 600 day weight by measuring the log-likelihood of restricted maximum likelihood (REML). The results showed the log-likelihood of the model was improved significantly when using the combination of regular GRM and maternal GRM simultaneously suggesting a better model. This result could be due to maternal imprinting, however further research is required to differentiate the maternal effects from parent-of-origin-dependent effects.

Breeding for flavor: linking metabolomics and genomics for blueberry improvement

Dr Luis Felipe Ferrão¹, Dr Timothy Johson¹, Dr Juliana Benevenuto¹, Rodrigo Amadeu¹, Dr Patrick Edger², Dr Thomas Colquhoun¹, Dr Patricio Munoz¹

¹University Of Florida, ²Michigan State University

Fruits are important sources of micronutrients in the human diet. After decades of intense breeding, breeders have mainly focused their activities on the improvement of yield, shelf-life, and appearance; which indeed meet the demand of growers and industry. In sharp contrast, consumers have been dissatisfied with the drop off in flavor quality for years. A prime reason for this inconsistency is the complexity of the target trait. Flavor is considered a multifactorial and subjective trait, involving a combination of multiple human senses. More specifically, it is the interaction between our olfactory system and the volatile organic compounds (VOCs) released by the food that provides each unique aroma and, therefore, a distinct flavor experience. Although many fruit crops had already characterized their VOC composition, the VOC roles on consumer preference nor their genetic architecture is well understood. Herein, we reported the use of a targeted genotyping and gas chromatography/mass spectrometry volatile extraction approaches of 1,438 individuals from a blueberry breeding population. By integrating metabolomics and genomics approaches, our contribution were three-fold: (i) showed that some volatiles are controlled by few major genomic regions, some of which harboring biosynthetic enzyme-coding genes; (ii) demonstrated that marker-assisted selection for flavor is a feasible and efficient tool to be implemented in practical breeding programs; (iii) showed that volatiles modulate consumer preference, which should provide direction for breeding. In conclusion, our study provides a new step forward for fruit flavor research, and illustrate the scientific basis needed for applying molecular breeding to improve fruit flavor perception.

Modelling the genetic structure of a priori brain networks

Anna Fürtjes

Biography:

Anna Fürtjes is a second year PhD student at the Social, Genetic and Developmental Psychiatry centre at King's College, London. Her research uses statistical genetic and neuroimaging techniques to address questions surrounding individual differences in cognitive abilities and age-related cognitive decline. She uses open research practices including pre-registrations and publicly available data analysis code.

Biologists assume based on Cheverud's conjecture that phenotypic correlations between morphological traits mirror genetic correlations. It has not yet been investigated whether this applies to the human brain. Phenotypic studies described the brains' organisation into networks of interconnected regions, and that properties of structural connectivity within these networks are associated with age and cognitive abilities. This study focuses on those structural properties to test whether phenotypic indices of brain connectivity recapitulate genetic connectivity which is what Cheverud's conjecture would suggest. There remains scepticism towards the validity of brain networks and genetics may uncover stronger biological foundations than can be assumed from phenotypic studies. The latest genome-wide association study (GWAS) summary statistics for brain volumes were used to represent network connectivity with covariance-based indices extracted using Principal Component Analysis. We present a novel methodology summarising by-variant effects across multiple GWAS phenotypes representing network connectivity in univariate summary statistics. Genetic characteristics of whole-brain and network-specific connectivity were described and compared with phenotypic ones. Dimensions of genetic connectivity accounted for less than half the interindividual variance across brain volumes (17%) compared to phenotypic connectivity (41%). Correlations between phenotypic whole-brain connectivity and age-volume associations were not recapitulated by genetic indices. Canonical networks were genetically uncorrelated with performance on cognitive tests. Such discrepancies between properties of phenotypic and genetic brain connectivity would not be predicted based on Cheverud's conjecture. This study highlights the importance of non-genetic influences on brain organisation and raises questions about the existence of a biological foundation of canonical brain networks.

Using robots to characterize genotype by environment interactions through developmental time

Dr Joseph Gage

Biography:

Joseph Gage is a NSF Postdoctoral Fellow at Cornell University, mentored by Ed Buckler and Richard Vierstra. Prior to his postdoc, he completed a MS in Biometry and a PhD in Plant Breeding and Plant Genetics in Natalia de Leon's group at the University of Wisconsin - Madison. His research integrates novel phenotyping methods, quantitative genetics, and molecular -omics to understand plant productivity.

Technological advances are enabling collection of immense phenotypic datasets from multiple timepoints that can be leveraged to study how GxE changes throughout development within a single location. Using semi-autonomous sub-canopy rovers, we collected in-field Light Detection and Ranging (LiDar) scans of thousands of maize hybrids on a weekly basis, which allow three dimensional reconstructions of entire field trials at various discrete time points during the growing season. Because LiDar point clouds are noisy, complex, and high-dimensional, they are not well-suited to analysis by traditional means. We used machine learning to generate Latent Space Phenotypes (LSPs), which are heritable traits that summarize plant architecture in an unsupervised manner free from human biases. LSPs were used to describe plant development throughout the growing season. Individual time points within a single location can be considered different environments, enabling study of within-location GxE. We hypothesize that genomic regions associated with GxE within a location (Aurora, NY), will also be associated with GxE between locations assayed as part of the Genomes to Fields initiative (G2F). The scalability of sub-canopy rovers holds promise for evaluation of numerous environmental locations at densely spaced timepoints in future studies.

Exploring the impact of selection at hair colour loci across different human populations

Mr Francesc Ganau Penella

Biography:

Francesc Ganau (Lleida, Spain, 1995) graduated with Honors in Biomedical Sciences in 2017 from the University of Lleida. He did a MSc in Bioinformatics at the Universitat Autònoma in Barcelona (2017/18) and a MSc in Quantitative Genetics and Genome Analysis at the University of Edinburgh in United Kingdom (2018/19). In 2019/20 he did a MSc by Research at the University of Edinburgh on the recent evolution of human hair colour. He currently works on his PhD at the Centre for Genomic Regulation of Barcelona. He suffers from legal blindness (20/400 vision) since he was five years old.

When humans left Africa and dispersed around the world, they met novel environments to which their phenotypes were not adapted. The resulting local selective pressures on each of these geographically distinct populations varied in direction and strength, and led to a plethora of divergent phenotypic adaptations. This research studies the genomic fingerprint left by these processes on hair colour, a trait with substantial inter and intra-population phenotypic variation, that we use as a proof-of-concept for differential adaptation.

Genomic regions strongly associated with hair colour are defined based on a published GWAS on the red, brown and blonde hair colours of the white British population. Association between hair colour and selection is tested using permutation-based methodologies. Significant regions from the hair-colour GWAS are tested for selection across all 26 populations present in the 1000 Genomes database (<https://www.internationalgenome.org/>), comparing how positive selection shaped each of them.

Results show that blonde and brown hair underwent strong, recent selection in most European populations, although there is no evidence of this selection being driven by hard sweeps. Conversely, African populations show loci in all 3 hair colours to be generally conserved, matching our hypothesis that hair evolved to adapt to new environments out of Africa. Finally, some populations in Eastern Asia show strong positive selection patterns for loci involved in brown and red hair. This may suggest that the black hair present in areas like Japan may not stem from the same alleles than black hair in Africa.

Detecting unrecorded environmental challenges and evaluating genetic determinism of resilience in lambs

C.A Garcia-Baccino^{1,2}, C Marie-Etancelin¹, F Tortereau¹, D Marcon³, J.L. Weisbecker¹, A Legarra¹

¹GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, ²Facultad de Agronomía, Universidad de Buenos Aires, ³Unité Expérimentale INRAE, Domaine de La Sapinière, INRAE

Biography:

I am from Argentina. I studied at Universidad de Buenos Aires. I graduated in Agricultural Engineering (2012) and obtained my Master (2017) and PhD (2019) in Quantitative Genetics and Animal Breeding. Since 2019, I am working in INRAE Toulouse in the GenPhySE unit in a two-year contract as a postdoc. I am actually working in the Genetic and Genomic Modelisation group with Dr. Andres Legarra and Dr. Zulma Vitezica. I am involved in SMARTER project and my work is mainly methodological, related to quantitative genetics and genetic evaluation of livestock.

Resilient animals are capable of remaining productive under different environmental conditions. Detection of environmental challenges affecting the entire population can provide a unique opportunity to select animals more resilient to those events. Here we present a method that consists in, first, inferring the existence of highly variable days (indicator of environmental challenge) via mixture models applied to frequent phenotypic records and second, using the inferred probabilities of occurrence of an environmental challenge in a reaction norm model to evaluate genetic determinism of resilience to these events. We illustrated the method using an ovine dataset with daily feed intake (DFI) records. The estimated probabilities of occurrence of an unrecorded environmental challenge proved to be informative and useful to include as a covariate in a reaction norm animal model. We estimated breeding values for environmental sensitivity of the genetic potential for DFI. The level and slope showed to be negatively correlated (-0.46 ± 0.21) showing that a hypothetical selection for increased or decreased DFI would result in increased susceptibility to stress. The reranking of individuals observed along the environmental gradient and the low genetic correlations among extreme environmental conditions confirmed the existence of GxE interaction. The method is promising and seems viable to identify unrecorded environmental challenges events, useful when selecting resilient animals and only productive data are available. It is general enough to be applied to a wide variety of phenotypic records from different species and useful when dealing with large datasets. In the frame of H2020 project SMARTER n°772787.

In-silico, in-vitro and in-vivo approaches to investigate a GWAS-identified risk locus implicate GPX3 as lead candidate gene in ALS

Restuadi Restuadi, Frederik J. Steyn, Edor Kabashi, Shyuan T. Ngo, Fei-Fei Cheng, Marta F. Nabais, Mike Thompson, Ting Qi, Anjali K. Henders, Leanne Wallace, Chris R Bye, Bradley J. Turner, Laura Ziser, Susan Mather, Pamela A. McCombe, Merrilee Needham, David Schultz, Matthew C. Kiernan, Wouter van Rheenen, Leonard van den Berg, Jan H Veldink, Roel Ophoff, Alexander Gusev, Noah Zaitlen, Allan F. McRae, Robert D. Henderson, Naomi R. Wray, Jean Giacomotto, **Fleur C. Garton**¹

¹*Institute for Molecular Biosciences, University Of Queensland*

Fleur Garton is an early career postdoctoral research fellow working with the Wray group on understanding the genetic contributions to motor neurone disease.

Amyotrophic lateral sclerosis (ALS) is a complex late-onset, neurodegenerative disease with a poor prognosis. Large ALS studies demonstrate that both common and rare variants contribute to the disease liability and genome-wide association studies in ALS and controls have identified ten risk loci to date. We were the first to identify an ALS risk locus on chromosome five using a multi-ethnic GWAS approach in 2017 1 which was subsequently detected in a larger, partially overlapping, European only ALS GWAS 2. Current association analysis data alone cannot determine which gene in the locus (two initially implicated GPX3 and TNIP1) is the most plausible contributing to ALS risk. Here, we report on a broad set of follow-up studies conducted to provide independent evidence that could support the relevance of one, both or neither in the context of ALS risk.

We used in-silico (COJO, FUMA, LDSC, PoPS, SMR, TWAS), in-vitro (human motor neurons) and in-vivo (expression in ALS case/control plasma, zebrafish) approaches to narrow down the likely candidate mechanism that alters susceptibility to ALS.

Both TNIP1 and GPX3 are implicated across a broad range of in-silico analyses, as a reflection that rs10463311 is an eQTL for both genes. The in-vivo expression analyses in ALS cases (preliminary n= 48 and replication n=198) suggests GPX3 protein expression decreases in plasma with worsening disease (ALSFRS-R, adjusted R2 = 0.042, p = 0.0055, replication) with TNIP1 not detected. In-vitro GPX3 and TNIP1 knock-down in human motor neurons did not identify a phenotype however, functional validation in-vivo indicates a pathogenic role of GPX3 loss-of-function causing motor deficits in zebrafish embryos, which were reduced (rescued) with MO-insensitive gpx3-mRNA.

This research demonstrates an investigation of an ALS risk locus to identify a candidate gene for follow-up studies. We use complementary lines of evidence demonstrate support for GPX3 contributing to ALS risk which has implications for understanding mechanisms of disease and targeted therapeutic approaches. We highlight that aspects of this pipeline (particularly in-silico, which is high-throughput and cost-effective) could be utilised to counter the inherent difficulty of identifying lead candidates in ALS and other neurodegenerative diseases.

Incorporation of GWAS data on genomic prediction accuracies assessed in the small, unadmixed, unstructured population of Icelandic dairy cattle

Mr Egill Gautason

Biography:

I am currently a PhD student at the Center for Quantitative Genetics and Genomics at Aarhus University, under the supervision of Senior Researcher Goutam Sahana. My background is agricultural science. I have a BS degree in agricultural science from the Agricultural University of Iceland, and an MSc degree in agrobiolology from Aarhus University. My main research interests are genetic diversity of livestock and application of genomic breeding methods in small livestock populations.

Genomic estimation of breeding values has greatly increased genetic gain for major dairy cattle breeds, but has not been widely implemented in smaller populations. In small cattle populations, the implementation of genomic selection is restricted by low numbers of progeny tested reference bulls, which limits the prediction accuracy of genomically estimated breeding values (GEBVs). Methods to increase accuracy of GEBVs for small cattle populations are needed for them to remain competitive.

We use genotype data from 8,100 Icelandic cattle, including cows and progeny tested bulls, to study the effect of enriching genomic prediction with GWAS results on prediction accuracies. Icelandic cattle is a dairy breed that consists of approximately 26,000 breeding females, and has an active (non-genomic) breeding program. We have genomically characterized the population, demonstrating that Icelandic cattle belongs to a clade of northern Nordic indigenous breeds and has been isolated for a long time. Admixture analysis revealed only very small traces of foreign import into the population, and the population is unstructured. Linkage disequilibrium is higher in Icelandic Cattle than in Red Nordic and Holstein-Friesian cattle. We have compared prediction accuracies of pedigree-based BLUP with single step GBLUP and found that genomic prediction accuracies for yield traits are 30-60% higher than pedigree-based BLUP. Future research is to enrich genomic predictions with functional annotation, or GWAS results and assess accuracy gains.

Polymorphism of Leptin Gene and Its Association with some Growth Traits in Lori Bakhtiari and Lori Bakhtiari-Afshari Crossbreed Sheep

Dr Ali Ghazi Khani Shad¹, Mr Mohammad Kazem Sharifi Shoorabi²

¹Azad University of Saveh, ²Azad University of Saveh, Iran

Biography:

I have PhD in Animal Breeding and genetics and I am working as university professor from 2003 to 2020 in Azad University of Saveh, Iran. I have a B.Sc. in Animal Production Engineering (1996-2000) (GPA Score: 17.14 from 20) and also, I am a graduate of an M.Sc. in Animal Science (Animal Breeding and Genetics) from the University of Tehran (2000-2002) (GPA Score: 17.08 from 20). I did my thesis in the field of poultry breeding. My Ph.D. course started in 2002 at the Science and Research University of Tehran in the discipline of Animal Breeding and Genetics. In my Ph.D. thesis, I compared different selection and mating strategies by using real data and stochastic simulation to maximize genetic improvement while minimizing the rate of inbreeding. The main area of my research in the past was in Quantitative Genetics and Molecular Genetics. With new progress in recent years in molecular genetics, bioinformatics, and computer strategies, it is necessary to apply these tools to accelerate genetic improvement and improve selection accuracies in animal breeding programs. This is the most critical reason that I will be working on new technologies in my future research work.

The purpose of this study was evaluation of leptin gene polymorphism by PCR-SSCP and its relationship with some growth traits in Lori Bakhtiari and crossbred of Lori Bakhtiari- Afshari sheep. Blood samples were collected from sheep (male and female) of Lori-Bakhtiari in Shahr-e-Kord Sholi station and sheep (male and female) of Lori Bakhtiari-Afshari crossbreed from villages of Shahr-e-Kord. DNA was extracted, using extraction kit of Sinnagen co. Evaluation of quality and quantity of DNA was performed using agarose gel electrophoresis and spectrophotometry. Polymerase chain reaction (PCR) was conducted to amplify bp fragment of exon of leptin gene. Then single strand conformation polymorphism (SSCP) of PCR products was performed and genotypic patterns were obtained using acrylamid gel and silver staining. For leptin gene in Lori bakhtiari sheep, band patterns including L to L and for crossbreeds sheep, band patterns including L to L were obtained. There was a high polymorphism in exon of the leptin gene in both breeds. Also results of means comparison showed that the leptin gene was significantly associated with weight at and months.

Genome-wide study of the human lipidome and links to cardiovascular disease risk

Dr Corey Giles¹, Dr Gemma Cadby², Dr Kevin Huynh¹, Ms Natalie A Mellett¹, Mr Gavriel Olshansky¹, Dr Alexander Smith¹, Ms Anh Nguyen¹, A/Prof Michael Inouye⁴, Prof Eric Moses^{3,5}, Prof Peter J Meikle¹
¹Metabolomics Laboratory, Baker Heart And Diabetes Institute, ²School of Population and Global Health, Faculty of Health and Medical Sciences, The University of Western Australia, ³The Curtin UWA Centre for Genetic Origins of Health and Disease, Faculty of Health Sciences, Curtin University and School of Biomedical Sciences, The University of Western Australia, ⁴Systems Genomics, Baker Heart and Diabetes Institute, ⁵Menzies Institute for Medical Research, University of Tasmania

Biography:

Corey Giles is an early career research scientist at the Baker Heart and Diabetes Institute (Melbourne, Australia). His passion and research interests lie in understanding the mechanisms through which lipid metabolism affects chronic diseases – including cardiovascular disease, type 2 diabetes and Alzheimer’s disease. His career goals are to identify the metabolic pathways that play a causal role in disease aetiology. Utilising high-throughput lipidomics, Corey aims to profile tens of thousands of human plasma samples from a broad range of population- and clinical-cohorts. Integrating lipidomic data with existing genomic resources opens new avenues for defining lipid species and pathways causally associated with disease aetiology and progression. Supported by a dedicated and diverse research team – led by Professor Peter Meikle – Corey is developing and applying novel statistical techniques to a fast-growing collection of human lipidomes.

Dysregulation of lipid metabolism is as an important – and modifiable – risk factor for the initiation and progression of cardiovascular disease (CVD). Although lipid metabolism is established as crucial to numerous biological processes, the genetic factors that influence inter-individual variation are still not well understood. To address this issue, we apply an integrative approach to link genetic variants with altered lipid metabolism and CVD.

Lipidomic profiling was performed using liquid chromatography coupled electrospray-ionization tandem-mass spectrometry on 4,492 individuals from the Busselton Family Health Study. We performed genome-wide association analysis of 596 lipid species and 33 lipid classes using linear-mixed models, correcting for age, sex, their interactions, 10 genomic principal components and an empirical genetic relatedness matrix. To account for lipoprotein mediated associations, the analysis was repeated with HDL-C, triglycerides and total cholesterol as covariates. Additionally, collider bias was avoided by conditioning using multi-trait-based conditional and joint analysis.

Over 70,000 genome-wide significant ($p < 5e-08$) associations were identified, with 451 lipid species having at least one significant association. Approximately 70% of the observed associations are independent of lipoprotein measures. Pleiotropic associations of genetic variants were determined by integration of results from expression- quantitative trait locus (QTL) and protein-QTL studies. We find significant genetic correlations between lipid species and CVD in the UK Biobank and CARDIoGRAMplusC4D consortium. Leveraging two-sample Mendelian Randomization, we further show that several lipid species have putative causal roles in CVD.

By linking lipid species to CVD, through genetic associations, we highlight potential therapeutic targets for monitoring, prevention and treatment.

Haplotype associated RNA expression (HARE) improves prediction of complex traits in maize

Dr Anju Giri

Biography:

Anju Giri is currently a postdoctoral researcher working with Professor Ed Buckler at Cornell University in Ithaca, New York. She got her Ph.D. in wheat breeding from Kansas State University in 2019, where her dissertation was on improving wheat for heat and drought stress tolerance. Her research area includes genetics, genomics, breeding, and physiology.

Gene expression is an important phenomenon that involves the regulation in cis, trans, and their interactions. Here we imputed haplotype expression (RNA expression of genes by haplotype) and evaluated its transferability across tissues and its utility in genomic prediction. The average correlation between haplotype Associated RNA Expression (HARE) and measured expression ranged from 0.43 to 0.49 across different tissues. We show that HARE was more transferable across tissues than the measured transcript expression. Genomic prediction models were evaluated across two diverse maize populations: the Goodman Association Panel (diverse mapping population of ~280 individuals) and the US Nested Association Mapping panel (NAM; 25 families of 200 individuals per family for ~5000 total individuals) to predict 26 complex traits. HARE often resulted in higher prediction accuracy than randomized values of haplotype expression, suggesting that they carry functional information in addition to information about haplotype structure. HARE increased prediction accuracy up to 14% over random expression values, however, the accuracy increase was dependent on the trait and the training population. The largest increase was observed when the model was trained in NAM and tested in the Goodman Panel. By leveraging the diverse NAM founders' high-quality assemblies through the Practical Haplotype Graph (PHG) we showed its utility in imputing expression in related and unrelated samples in the population.

Determining the clinical epidemiology, genetic architecture, and sex-differences in psychogenic nonepileptic seizures

Ms Slavina Goleva

Biography:

My name is Slavi Goleva, I earned my BS in Biology from the University of Kentucky in 2012. I am currently working towards completing my PhD at Vanderbilt University in Dr. Lea K Davis's lab where my research focuses on the clinical and genetic epidemiology of psychiatric disorders.

Psychogenic nonepileptic seizures (PNES) are often clinically confused with epilepsy, but have different causation/treatment. This condition is debilitating, understudied and there is no consistently utilized ICD code for PNES. Previous studies have shown that patients with PNES are more frequently female, with more comorbid psychiatric diagnoses and obesity.

The goal of this study was to characterize the clinical and genetic characteristics of PNES.

We identified 4,267 PNES patients (prevalence 0.17%) in the Vanderbilt University EHR (VU-EHR). PNES patients were identified by a phenotyping algorithm using inclusion (seizure ICD codes, PNES- and EEG-specific keywords) and exclusion (epilepsy codes) criteria applied to EHR data (80% PPV).

We determined which diagnoses in the VU-EHR were most comorbid with PNES using a PNES case/control Phenome Wide Association Study. We identified expected associations with psychiatric disorders, and novel associations with cerebrovascular disease (OR=1.15, $p=2.8 \times 10^{-13}$).

73% of algorithm-identified PNES patients were female. To determine PNES-comorbidity sex differences, we calculated odds ratios between PNES and 13 common psychiatric disorders in females vs. males. Several psychiatric disorders were more likely to be comorbid with PNES in females than in males (e.g. PTSD, OR=2.72).

We hypothesized that genetic risk for commonly comorbid disorders would be higher among PNES cases than controls. We calculated polygenic risk scores (PRS) for 11 comorbid disorders in algorithm-positive PNES cases (N = 296) and controls (N = 60,049). The PRS for T2D (OR=1.15, $p=0.02$), suicide attempts (OR=1.15, $p=0.02$), focal epilepsy (OR=1.19, $p=0.01$), and generalized epilepsy (OR=0.87, $p=0.01$) were correlated with PNES case/control status.

Exploring tolerance to parasitic disease in Nellore cattle

Miss Gabriela C. Gouveia

Biography:

I am a veterinary, graduated from the Federal University of Minas Gerais - UFMG (2014), and I hold a Master's degree in Animal Science at UFMG, focusing on Animal Breeding and Genetics (2017). I am currently a visiting research student at The University of Queensland - Australia, and a Ph. D. candidate in Animal Science at UFMG with a focus on Animal Breeding and Genetics at UFMG. I have a background in quantitative genetics, animal improvement programs' design, and genome-wide association studies for phenotypes of productive and reproductive interest in the animal production industry.

The tolerance mechanisms through which hosts do not reduce parasite infection but alleviate the negative fitness consequences of pathogen burden instead might be an efficient selection criterion to cope with beef cattle parasitic diseases. Thus, we aimed to estimate genetic variability of tolerance to different parasites and study the genotype by parasitic load interaction. For this, we used body weight at different ages and ticks, gastrointestinal nematodes eggs and *Eimeria* spp. oocysts counts from 1712 Nellore young bulls raised in Brazil and 1083 out of these were also genotyped (30k SNP panel). Tolerance was analyzed as the reaction norm slopes for parasite burden where the body weight was regressed on both age and each of three parasites burden trajectory using linear splines with 7 and 2 knots, respectively. The 2 knots described the low and high infestation where median of ticks, nematodes and *Eimeria*'s counts per batch=0 for knot 1 and 33; 12 and 16.5 for knot 2, respectively. Our results showed that additive variance estimates for tolerance increase with pathogen burden trajectory, ranging from 63.67 ± 16.68 to 296.89 ± 75.97 for ticks, from 58.13 ± 16.68 to 206.80 ± 46.50 for nematodes and from 57.19 ± 16.29 to 648.25 ± 141.76 for *Eimeria* spp., revealing significant additive variance for parasitic load tolerance. Our results suggested presence of genotype by parasitic load interaction since we observed low genetic correlation of 0.20 for ticks; 0.16 for nematodes and -0.03 for *Eimeria* spp. between low and high infestation. Thus, variation in tolerance can induce genotype re-ranking, contributing to environment-dependent genetic responses to selection.

Selection for worm resistance in sheep resulted in an increase in genetic variance over time

Dr Johan Greeff

Biography:

Dr Johan Greeff

Senior Geneticist – Sheep industry, Department of Primary Industries and Regional Development

Johan was educated in South Africa and came to Australia in 1993 to take up a geneticist position in the Department of Agriculture Western Australia in Katanning. Since 1993 he has been researching the genetics and the underlying biological causes of worm resistance in the Rylington Merino flock.

Selection theory predicts that selection changes gene frequencies and that it can lead to an increase in homozygosity, which reduces the genetic variance. This paper will report on the changes in additive genetic variance in the long term Rylington Merino sheep selection line that have been closed and selected for an increase in worm resistance using faecal egg counts (FEC) over 21 years. FEC at weaning and FEC at hogget age were investigated. The dataset was subdivided in four time phases (5, 5, 5, and 6 year periods) and each phase was analysed separated with mixed model methodology, and using the same pedigree structure. The data were standardised to have the same variance for each year and sex. As the standardised data were skewed, the data were also log and square root transformed. An animal model was used and the results showed that there was an increase in the genetic additive variance over time, while the residual variation stayed relatively the same. This is contrary to expectation and resulted in a general increase in the heritability estimates that nearly doubled for both traits from the first to the fourth phase. The results will be discussed.

Genetic Analysis of Pathogen-Specific Intramammary Infection in Dairy Cows

Mrs Saranya Gunasegaram Narayana

Biography:

I am Saranya Narayana, PhD student at the University of Calgary, Canada. Currently, I am in my third year of PhD working on a collaborative project between University of Calgary and University of Guelph. My PhD project focuses on genetic and genomic analysis of mastitis in dairy cattle. I have obtained my master's degree in Animal Breeding and Genetics at University of Guelph, Canada on 2014. Moreover, I have also bachelor's degree in Animal Science and Fisheries from University of Peradeniya, Sri Lanka. My aspiration is to become a geneticist in the future.

The overall objective of this study was to investigate genetic variation of overall and pathogen-specific intramammary infection (IMI) in non-clinical multiparous cows. Data and milk samples were collected over a 2-y interval as part of the Canadian Bovine Mastitis Research Network. The final dataset contained records of 46,900 quarter milk samples from 3,382 clinically healthy multiparous Holstein-Friesian cows from 84 dairy herds. Seven traits were considered for the genetic analysis: overall IMI, non-aureus staphylococci IMI, contagious pathogen IMI, environmental pathogen IMI, major pathogen IMI, minor pathogen IMI, and natural logarithm of somatic cell count (LnSCC). Data were analyzed at the quarter level using a threshold probit model. Prevalence of IMI traits at quarter-level ranged from 6.8 to 45.5%. Heritability estimates (quarter-level) of overall IMI and pathogen-specific IMI traits ranged from 0.017 to 0.073. Genetic correlation ranged from 0.18 to 0.97 among pathogen-specific IMI traits and with overall IMI. IMI traits had low to moderate genetic correlation with LnSCC. Estimated moderate to high (0.31 to 0.87) Spearman rank correlation between estimated breeding values (EBV) for overall IMI and pathogen-specific IMI traits indicated possible re-ranking of sires. Moreover, percentage of daughters with IMI caused by different pathogen groups ranged from 13 to 80% and 38 to 94% for the best (10% decile) and worst sires (90% decile) ranked by their EBV, respectively. In summary, despite low heritability, there is utilizable genetic variation for pathogen-specific IMI traits. Hence, pathogen-specific IMI resistance can be improved through long-term genetic selection.

Partitioning of variance between multiple relationship matrices in BLUP analyses

Dr Phillip M. Gurman¹, Dr Li Li¹, Dr Andrew A. Swan¹, Dr Nasir Moghaddar², Professor Julius HJ van der Werf²
¹Animal Genetics And Breeding Unit, ²School of Environmental and Rural Science

Biography:

Phillip Gurman is a Research Fellow with the Animal Genetics and Breeding Unit, University of New England. Phillip's role is focused on the development of methods for the use of genomic data in routine genetic evaluation systems. This role is primarily focused on sheep industry but is also extended to other species.

GWAS analyses have resulted in SNP sets that are more predictive for specific traits. Combining these SNPs in a genomic relationship matrix (GRM) with non-selected SNPs may dilute their predictive ability. Instead, predictive SNPs could be treated separately with their own variance. This study examines the partitioning of variance between multiple genetic effects defined by multiple relationship matrices. Univariate REML analyses were performed using GCTA for intramuscular fat (imf), carcass eye muscle depth (cemd), and carcass fat (ccfat) measured on approximately 9.5k genotyped sheep from multiple breeds. Genetic relationship matrices fitted included numerator relationship matrix (NRM) and two GRMs, one based on a standard SNP array (GRMC, 48.5k) and one based on SNPs selected from whole-genome sequence (GRMP, 2.7k). GRMs were constructed with either breed-specific allele frequencies, or population allele frequencies. Breed structure was accommodated by fitting random genetic groups. For GRMs constructed with population allele frequencies, the proportion of genetic variance attributed to GRMC was between 0.14 for ccfat and 0.38 for imf, while for GRMP it was between 0.36 for imf and 0.73 for cemd. The remaining genetic variance was explained by the NRM (range 0.02- 0.38). Similar proportions were observed for the multi-breed GRM. Proportions of genetic variances estimated for the NRM and GRMs can be used in single-step models to increase prediction accuracy, but questions remain regarding the impact of co-linearity between effects. For example, using a breed-adjusted GRM resulted in an increase in genetic group variance relative to the other genetic effects.

Resource Allocation Optimization in Cereals

Dr Lucia Gutierrez

Biography:

Dr. Gutierrez is an Associate Professor and the Cereals Breeder and Quantitative Geneticist at the University of Wisconsin-Madison. Her research focuses on understanding the genetic architecture of complex traits and their response to the environment. She integrates state of the art genotyping technologies with large phenotyping experiments to study complex traits. She studies the mechanisms employed by plants for local adaptation including the study of biotic and abiotic interactions. Her research program has also an applied component, which combines strong theoretical development, genomic tools, and high throughput phenotyping to release cereals cultivars to serve the U.S. agricultural systems.

Sustaining the increasing human population will require major progress in cultivar productivity. Billions of dollars are yearly spent on cultivar evaluation to discern genetic merit from the environment and noise components. Controlling the micro-environment (i.e. plant-to-plant variations due to field heterogeneity) and the macro-environment (i.e. genotype by environment interaction due to locations and years, GxE) is fundamental for breeding success. We will compare methodological approaches to optimize resource allocation for plant breeding using genomic and large multi-environment phenotypic information. Optimization of experimental design strategies based on both micro-environmental variation (local control of field heterogeneity with experimental designs) and macro-environmental variation (GxE due to years and locations) are discussed, comparing prediction accuracy as well as response to selection. Finally, resources were optimized for multi-trait studies. Our results show that resources can be optimized within and across locations when field heterogeneity and GxE is modeled with genomic information. Additionally, easy or inexpensive traits can be phenotyped at a higher depth than expensive or difficult to phenotype traits to optimize phenotyping resources. These strategies increase the response to selection while using the same phenotyping resources by a combination of using larger population sizes and increasing phenotyping efficiency.

Genomic prediction and genotype x environment characterization in unbalanced historical cereals variety trials

Dr Lucia Gutierrez

Biography:

Dr. Gutierrez is an Associate Professor and the Cereals Breeder and Quantitative Geneticist at the University of Wisconsin-Madison. Her research focuses on understanding the genetic architecture of complex traits and their response to the environment. She integrates state of the art genotyping technologies with large phenotyping experiments to study complex traits. She studies the mechanisms employed by plants for local adaptation including the study of biotic and abiotic interactions. Her research program has also an applied component, which combines strong theoretical development, genomic tools, and high throughput phenotyping to release cereals cultivars to serve the U.S. agricultural systems.

Genomic prediction using a large amount of historical multi-year, multi-location datasets have rarely been utilized in genomic prediction. Utilization of historical multi-year, multi-location (environment) variety trials could help in developing robust genomic prediction models enhancing genomic selection in a breeding program. However, historical variety trial data are generally highly unbalanced as the same sets of varieties may not be tested across several years and locations. We examined multiple strategies to utilize unbalanced historical datasets from oat and wheat breeding programs in the Midwestern regions of the United States to characterize genotype x environment interactions, identify mega-environments with the similar ranking of genotypes across multiple environments, and perform genomic selection integrating GxE interactions in each environment, mega-environment, and all-environments combined. The main goal of this project was to build a variety selector tool for cereals such as wheat and oat for each geographical region in the Midwest by modeling the GxE interaction to group environments with similar ranking of varieties and then predicting the performance of all varieties for a specific geographic region using well established mixed models developed by our group. This tool will then be evaluated by a group of farmers that will empirically test the performing of the chosen varieties against their own best choices for the area. This variety selector tool will help growers and the scientific community in identifying best performing varieties of cereals such as wheat and oat in their testing location as well as assist in identifying mega environments in their region.

Improving genomic evaluation for male fertility -can information from female fertility increase reliability in dairy cattle?

Dr Mekonnen Haile-Mariam

Biography:

My name is Mekonnen Haile-Mariam and I am a senior research scientist at Agriculture Research Victoria, Melbourne, Australia. Most of my work over the last 20 years has focussed on the genetic improvement of Australian dairy cattle to increase their overall profitability. My research interest is improving functional traits (reproduction, health, welfare) in particularly using moderately heritable indicator traits to improve genetic evaluation for lowly heritable traits.

In dairy cattle selection for milk production has led to decline in female fertility. As a result, genetic improvement programs have also focused on female fertility but ignored male fertility assuming the artificial insemination industry was able to properly screen and standardize the quality of semen before it is widely used. With the recent trend to use genomically selected bulls before adequate screening, understanding bull fertility early has become an emerging area of research. This study explored the potential to use genomic information to select young bulls before they are extensively used for semen collection. Success or failure of insemination outcomes of about 3000 Holstein bulls with high density genotype data were used to estimate heritability and assess the accuracy of prediction for validation bulls. Female and bull fertility data on cows and bulls, respectively, were used to estimate variance explained by dominance variance in addition to additive variance.

The proportion of variance explained by SNP data was about 9% in male fertility which a fifth of that in female fertility, but the use of genomic data increased the heritability by over 10 times compared to pedigree data. By accounting for dominance relationship and joint analyses with female fertility it is possible to identify and use SNPs information to improve reliability of both male and female fertility, given the relationship between the two is high in beef cattle where unselected data are used. Genetic relationship between female and male fertility and SNPs that affect both traits were also discovered.

Proportionate Responses to Genetic Variation

Miss Bailey Harrington

Biography:

Bailey Harrington is a final-year, computational PhD student in the MRC Human Genetics Unit at the University of Edinburgh. Her PhD work is centred on the genetics of human skeletal traits, which are derived from medical images. As a result, another area of her work has been looking into how to use existing medical images to derive new information about, or obtain additional phenotypes for, the individuals in a cohort. This work involves high-performance computing, existing software, and custom software developed by Bailey and others in the institute. Bailey's background is in forensic science, with an emphasis on forensic anthropology.

Morphological traits have been studied extensively in humans, first under the subsequently discarded belief that they would be simple to explain genetically, and now in an attempt to understand what we know to be extremely complex traits. This can be seen with height, a trait which is influenced by several different developmental components of growth. These include factors affecting the length of long bones, factors affecting the length of the torso, and factors affecting overall growth, as well as environmental factors, like diet. We aim to understand genetic control over these components of growth and their subsequent effects on variation in body proportions. To this end we have run GWAS and downstream analyses on torso length, as well as long bone lengths obtained from dual-energy x-ray absorptiometry (DXA) whole-body scans ($N \sim 11,000$; heritability estimates range from $.334 \pm .0494$ to $.5802 \pm .0451$). We are interested in variation in these traits that exists independent of variation in overall height, so height was treated as a covariate in the GWAS models. These GWAS have uncovered novel genome-wide significant associations with loci that neither harbour known height loci, nor are known to be in linkage-disequilibrium with them. Cross-referencing genome-wide significant hits between our GWAS shows there are compensatory effects of these loci between the length of leg long bones and torso length, such that a locus increasing one, decreases the other. Sex-stratified analyses have shown there is also sexual dimorphism in proportional variation.

Genome-wide association study on meat tenderness using genotypes imputed to whole-genome sequence in a diverse New Zealand sheep population

Dr Andrew Hess

Biography:

Andrew Hess is currently a post-doc at AgResearch, primarily focused on utilizing whole-genome sequence data to identify causative mutations for production traits in New Zealand sheep. Andrew obtained his PhD from Iowa State University in Interdepartmental Genetics and Genomics with a minor in Statistics. He is interested in methods development focused on the utilization of sequence data for genomic analysis as well as developing methods for improved reference assemblies to get more value out of sequence data. He has recently accepted an Assistant Professor position at the University of Nevada, Reno, starting July 2021 and is interested in discussing collaborative opportunities.

Genotyping for DNA-based parentage and genomic selection is routine in the New Zealand sheep industry. To-date, ~23,000 New Zealand sheep have been genotyped with the 600K SNP arrays. Imputation to whole-genome sequence (WGS) has the potential to identify causative mutations associated with traits of economic importance to the New Zealand sheep industry. The sheep population in New Zealand is very diverse, with over 50 purebred or composite breeds observed in the Sheep Improvement Limited database, and frequent crossbreeding. The goal of this study was to utilize whole-genome sequence data to impute individuals with 600K SNP array genotypes to WGS using Beagle 5.1. Individuals with WGS data from Run2 of the International Sheep Genomics Consortium (ISGC), including 213 individuals from New Zealand, were used as a reference to impute individuals with 600K genotypes to WGS. Overall imputation accuracy was high (average individual accuracy 0.97) and was consistently high among all major breeds with high density genotype data, showing the ability to utilize the individuals with WGS sequence data from the ISGC to impute the diverse New Zealand sheep population. Genotypes imputed to WGS were used in a genome-wide association study on meat tenderness, which identified additional QTL not identified using 600K genotypes. Including these variants for genomic prediction resulted in an increase in prediction accuracy for meat tenderness. This information can be used to improve the design of the available genotyping arrays.

Sampling variance of additive-by-additive genetic variance using GWAS data from a sample of unrelated individuals.

Dr Valentin Hivert¹, Pr Jian Yang¹, Dr Loic Yengo Dimbou¹, Pr Peter M. Visscher¹

¹Institute For Molecular Bioscience, University of Queensland

Biography:

Valentin was awarded a PhD from the French National Institute of Higher Education in Agricultural Sciences and the University of Montpellier I (France) in 2018. His thesis was focused on methodological developments for genetic differentiation analysis in the Next Generation Sequencing era in both a neutral and adaptive context. Since 2019, he works as a Post-doctoral researcher at the University of Queensland in the Program in Complex Trait Genomics group under the supervision of Professor Peter Visscher. His current research focuses on the estimation of non-additive genetic variance in human complex traits.

Epistatic variance for complex traits is notoriously difficult to estimate in non-laboratory species, including humans, since the environment cannot be controlled so that non-additive genetic variance can be confounded with environmental covariance among relatives. In principle, epistatic variance attributable to common DNA variants can be estimated from a random sample of unrelated individuals with genome-wide SNP data. In this design, the proportion of phenotypic variance explained by epistatic effects, hereafter denoted σ^2_{epi} , can be quantified in a classical variance component framework using the Hadamard product of a Genomic Relationship Matrix (GRM) with itself. Here, we derive the sampling variances of two estimators of σ^2_{epi} : the Haseman-Elston OLS estimator ($\hat{\sigma}^2_{\text{HE}}$) and the restricted maximum likelihood estimator ($\hat{\sigma}^2_{\text{REML}}$). We show that the sampling variances of both estimators can be expressed solely as a function of the sample size (N) and the sampling variances of the diagonal and off-diagonal elements of the GRM, which can be expressed as a simple function of the effective number of independent markers M_e . We show that $R_v = \text{var}(\hat{\sigma}^2_{\text{HE}}) / \text{var}(\hat{\sigma}^2_{\text{REML}})$ asymptotically converges towards $3/2$, although R_v can be much larger for small sample sizes. Finally, we show through simulations that our theory agrees with the observed standard deviation (across replicates) as well as the empirical standard error for each replicate dataset. Our results imply that statistical power to detect σ^2_{epi} is substantially larger using $\hat{\sigma}^2_{\text{REML}}$ as compared to $\hat{\sigma}^2_{\text{HE}}$, in contrast to the estimation of additive variance where the sampling variance of the two estimation methods are approximately the same.

Temperature effects on genetic and phenotypic plasticity of sex determination in zebrafish (*Danio rerio*)

Dr Shahrbanou Hosseini

Biography:

Dr. Shahrbanou Hosseini is a postdoctoral researcher in the Animal Breeding and Genetics group in the Department of Animal Sciences at the University of Goettingen in Germany. She successfully finished her PhD last year with the grade summa cum laude. She studied genotype by temperature interaction effects on phenotypic plasticity in a model animal, the zebrafish. Her study has been given the young scientist award by H. Wilhelm Schaumann Stiftung in the European Federation of Animal Science in 2018. She works in a very challenging area in between genetics, evolution and developmental biology.

Teleost fish species exhibit sexual plasticity depending on genetic and environmental factors. The underlying mechanism of environmental effects on their sex determination is not still fully understood even in the widely used teleost research model animal, zebrafish. The main objective of this study was to investigate the genetic × environment interaction on phenotypic plasticity in a polygenic sex determination system of zebrafish. For this purpose, zebrafish eggs were exposed to two temperature treatments (control: 28°C, temperature-treated: 35°C during embryogenesis) and larvae were kept until sexual maturity. Adult fish were imaged in order to classify the sex using two machine learning methods: Deep Convolutional Neural Networks (DCNNs) based on whole body image and Support Vector Machine (SVM) based on caudal fin picture. We detected differences in colour intensity in the caudal fin of males and females using SVM. Further to investigate the genetic mechanism underlying sex-linked colour dimorphism, we collected 48 gonad and caudal fin tissue samples for transcriptome analysis. The predicted sex from the DCNNs demonstrated highly in agreement with the real sex in both experimental groups. However, SVM showed a less intense colour in caudal fin of a subset of temperature-treated males, suggesting these animals are likely masculinized. Transcriptome profiles demonstrated male- and female-specific gene expression patterns associated with sex determination in gonad. However, no differential expression of colour genes was identified in the caudal fin, suggesting that the differences in colouration between males and females might be due to post-transcriptional regulation of key enzymes involved in pigment synthesis.

Effects of Broiler Strains on Body weight, Morphometric traits and their relationship during the Starter and Finisher stage

Mr Demilade Israel Ibiwoye

Biography:

I am Ibiwoye Demilade Israel. I am a graduate with a Master of Science degree in Animal Production (majored in Animal Breeding and Genetics) from University of Ilorin, Nigeria. I have been able to demonstrate and distinguish myself in academic excellence by coming top 1% of my graduating year with distinction. I am a researcher in animal science and have strong interest in animal breeding and genetics particularly poultry. I have been an Agricultural Science Teacher for about 4 years and also an Animal Science lecturer for the Joint Admission Preliminary Examination Board (JUPEB) for University of Ilorin, Nigeria.

Eight hundred (800) broiler chicks consisting of two hundred (200) each of Ross (RS), Arbor Acre (AA), Hubbard (HB) and Marshall (MS) strains reared under the same nutritional and environmental condition for a period of eight weeks were used to estimate variations between strains and the relationship among the body weight (BW) and some morphometric body parameter such as Body Length (BL), Body Girth (BG), Shank Length (SKL), Shank Circumference (SKC), Thigh Length (THL), Keel Length (KL), Wing Length (WL), Drumstick Length (DSL), Body Height (BH), Beak Length (BKL), Comb Length (CBL), and Neck Length (NKL). The results showed that significant differences ($p < 0.05$) existed across the strains at the age of four and eight weeks. There was a significant ($p < 0.05$) correlations coefficients (-0.320 to 0.855) between BW and morphometric parameters at four (4) weeks of age. A significant level of correlation coefficient that ranged from -0.013 to 0.838 existed between BW and morphometric body parameters of the broiler chickens at eight weeks of age and the highest correlation value was obtained between BKL and SKC ($r = 0.838$), while BW was significantly ($p < 0.05$) correlated with SL ($r = 0.615$) and KL ($r = 0.513$) in the four strain during this period.

This study showed that variations and significant relationship exist among the broiler chickens strain raised under the same nutritional and environmental condition with respects to their body weight and morphometric body parameters. Also, some morphometric body parameters contribute positively to overall body weight in the four strains of broiler studied.

Initial Findings from the Australian Genetics of Stuttering Study

Dr Victoria E Jackson

Biography:

Dr Vicki Jackson is a Research Fellow at the Walter and Eliza Hall Institute of Medical Research, Melbourne. Vicki has a background in Medical Statistics and Genetic Epidemiology, with expertise in GWAS, and analysis of sequencing and other 'omics data. Vicki's current work focuses on exploring the genetic underpinnings of speech disorders.

Characterised by dysfluent speech, stuttering is a complex communication disorder that has a profound, long lasting effect on an individual's social and mental wellbeing. Up to 11% of children commence stuttering by 4 years of age, with approximately one third of those developing a persistent stutter. The cause of stuttering is unknown, but genetic factors are thought to play an important role.

We hypothesise that common genetic variation makes a substantial contribution to the risk of stuttering and are investigating this by conducting a genome-wide association study (GWAS), using a large population-based sample of people who stutter. So far there have been no published GWAS for stuttering, hence there is a clear gap in our genetic understanding of this disorder.

We are recruiting an international stuttering cohort, comprising both adults and children over age 7 years, via media and online campaigns, and have begun with Australian patients. Participants provide a range of phenotypic data by completing online questionnaires, together with online speech recordings. Genotyping of the Australian arm of our cohort has been undertaken using the Global Screening Array. We will provide an overview of our cohort to date, and preliminary results for the GWAS of our Australian samples.

We will also describe our ongoing efforts to build a network of cohorts, that have collected information on stuttering; we will bring these studies together, allowing us to undertake a large-scale GWAS meta-analysis.

Comparing the effect of SNP density on quantitative genetic analyses of a wild population

Miss Caelinn James

Biography:

I am Caelinn James, a second year PhD student at the University of Edinburgh with a interest in evolutionary genetics, and my PhD project is focused on investigating quantitative genetic traits in Soay sheep. Prior to starting my PhD, my background was in biomedical genetics – my undergraduate degree acted as a ‘gateway’ into the world of quantitative genetics and introduced me to my passion; evolutionary genetics. I chose to do a PhD focusing on a wild population in order to see quantitative and evolutionary genetics from a different point of view and learn new methods for studying quantitative genetics.

The Soay sheep (*Ovis aries*) of the St. Kilda archipelago are an unmanaged population from the primitive domestic sheep breed introduced to the island over three thousand years ago. Those sheep residing in the Village Bay area of Hirta (one of the islands in the archipelago) have been subjects of an individual based study since 1985, which has been recording phenotypic data of the sheep (e.g. birth weight or leg length), their pedigree, census location, as well as environmental factors such as vegetation density. In addition, the sheep are also genotyped; 7,700 Soay sheep have been genotyped on the Illumina OvineSNP50 Genotyping BeadChip, which contains 50K SNPs. Using these data, SNP-based trait heritabilities have been estimated and variants associated with observed phenotypic differences identified.

Recently, 188 Soay sheep have been genotyped using the Illumina Ovine Infinium HD SNP BeadChip containing 450K SNPs, allowing for imputation from the low-density SNP panel to the higher density panel for the rest of the sheep. We repeated the heritability estimation and mapping analyses performed with the lower-density genotypes with the high-density imputed genotype data, and compared the results to answer the following questions: (I) do heritability estimates differ between the two SNP densities?, (II) can we recapitulate the associations found with the low-density genotype panel with the high-density one?, and (III) do we find novel associations with the high-density SNP panel? Overall, we aim to elucidate the importance of SNP density when investigating quantitative genetic traits in this sheep population.

Untangling environmental determinants for explaining crop performance by combining Genomic selection methods and weather data

Dr Diego Jarquin¹, Researcher Reyna Perez Sandoval¹

¹University Of Nebraska-Lincoln, USA

Biography:

Dr. Jarquin is a Research Assistant Professor in the Department of Agronomy and Horticulture at the University of Nebraska-Lincoln (UNL). Dr. Jarquin is a statistician who merges statistical methodology, computer algorithm development, data science and collaborative work with plant sciences (plant breeding, biometrics, biostatistics, quantitative genetics, etc.).

Dr. Jarquin is actively collaborating on several projects (Genomes to Fields, SoyNAM, etc.) with public (University of Tokyo, ICRISAT, CIMMYT, IRRI, EMBRAPA, CENICANA, etc.) and private (Advanta Seeds) sectors. Recently, Dr. Jarquin has earned the 2020 National Association of Plant Breeders Early Career Scientist Award.

Genomic prediction models have shown to be a powerful tool for predicting crop performance when applied to scenarios where the environmental conditions of the testing set are same than those from the training set (i.e., CV2: incomplete field trials and CV1: predicting newly developed lines); however, the performance of these models is highly affected when the environmental stimuli from the testing set is not similar to the training set. In these cases, the predictive ability decreases drastically being in some cases zero or negative. We developed a method that examines the mean crop performance (yield and days to pollen) of maize hybrids populations (G2F: Genomes To Fields Initiative) as function of variable time windows. Using only the environmental mean performance of environments in the training set, we conducted a search to identify the most informative window of time of crop season for predicting these means for 14 environmental covariates. Then, we use the same windows of time of these covariates for the environments in testing set for performing predictions of tested and untested genotypes in unobserved environments via the reaction norm model. The obtained results showed that the predictive ability was improved between 150-180% for the case when we targeted the prediction of already tested genotypes in unobserved environments (CV0) and between 200-250% when predicting untested genotypes in unobserved environments (CV00). These results are very encouraging and thus we are confident that these developments can also be applied to other crops and traits and in different breeding stages.

Improving effective use of germplasm collections: The USDA Soybean Germplasm collection as a model

Dr. Diego Jarquin¹, Dr. George Graef¹, Dr. Kent Eskridge¹, Dr. Hao x Hao Xiaojuan¹, Dr. Brian Diers², Dr. Aaron Lorenz³, Dr. Asheesh Singh⁴, Dr. Randall Nelson⁵

¹University Of Nebraska-Lincoln, ²University Of Illinois, ³University Of Minnesota, ⁴Iowa State University, ⁵USDA-ARS and University of Illinois

Biography:

Dr. Jarquin is a Research Assistant Professor in the Department of Agronomy and Horticulture at the University of Nebraska-Lincoln (UNL). Dr. Jarquin is a statistician who merges statistical methodology, computer algorithm development, data science and collaborative work with plant sciences (plant breeding, biometrics, biostatistics, quantitative genetics, etc.).

Dr. Jarquin is actively collaborating on several projects (Genomes to Fields, SoyNAM, etc.) with public (University of Tokyo, ICRISAT, CIMMYT, IRRI, EMBRAPA, CENICANA, etc.) and private (Advanta Seeds) sectors. Recently, Dr. Jarquin has earned the 2020 National Association of Plant Breeders Early Career Scientist Award.

A major challenge in plant breeding is how to choose lines from a large germplasm collection where phenotypic data for important, complex traits like yield are absent or limited. The USDA Soybean Germplasm Collection contains ~20,000 accessions. Fewer than 20 accessions account for >90% of the genes in US commercial cultivars. This becomes a sampling question. The USDA soybean collection has genotype information for nearly every accession in the form of 50K SNP data. The objective of this research was to compare three sampling methods for their ability to identify the most informative set of individuals for a given sample size from a finite population using only molecular marker information. We compared random(RAN), cluster(CLU) and supersaturated design(SSD) to sample genetic diversity from the collection in Maturity Groups I-IV. Five-hundred accessions were sampled across three methods, with ~160 accessions in each sampling group. High-quality phenotypic information was collected from yield test plots across the north central USA in 14 environments for each MG test across two years. The SSD sampling group more effectively represented the total genetic variation in the collection than did the RAN or CLU samples. Phenotypic and genotypic variance for the SSD group was double that for RAN and CLU samples, and predictive ability in unobserved environments was improved between 11% and 40%. Efficient and effective sampling, maximized genotypic variance, efficient evaluation and addition of high-quality phenotypic information, coupled with genomic information have important implications for improved genetic gain and more effective use of germplasm collections worldwide.

Identifying microRNA Biomarkers in Prostate Cancer: A Transcriptome-Wide Association Studies (TWAS)

Mrs Dulari Kaushalya Jayarathna

Biography:

Ms Dulari Jayarathna obtained her Bachelor of Science (Statistics) degree from the University of Peradeniya, Sri Lanka. After her graduation, she worked as a teaching assistant and applied data analysis on health-related projects in collaboration with researchers from the faculty of medicine, faculty of dental science and faculty of allied health sciences in University of Peradeniya, Sri Lanka. These collaborations fascinated her to select the field of mathematical and computational biology in doctoral studies. Currently, she is pursuing her PhD with the School of Chemistry and Physics, Queensland University of Technology and QIMR Berghofer Medical Research Institute.

Prostate cancer (PrCa) is one of the leading causes of mortality and is the most commonly diagnosed cancer among men in Australia. The identification of novel biomarkers in PrCa could be valuable in the design of therapeutics and the identification of their molecular targets. Genome-wide association studies (GWAS) have identified about 170 PrCa risk variants and certain of them can be located within non-coding ribonucleic acid (RNA) regions such as microRNAs [1]. This study investigates the role of microRNAs in PrCa risk using a transcriptome-wide association studies (TWAS) approach, summary data-based mendelian randomisation (SMR). Six-microRNAs were identified by SMR and they were further analysed using HEIDI (heterogeneity in dependent instruments) method to evaluate single variants which effect both microRNAs and risk variants. The HEIDI test recognised hsa-miR-204-5p and hsa-miR-4661-5p as significant in PrCa risk. Moreover, these miRNAs were predicted as tumor-suppressive by SMR effect sizes. The experimental works have already established that hsa-miR-204-5p repress PrCa oncogenes and hsa-miR-4661-5p was recorded for the first time which requires wet-lab experiments for validation[2]. After validating results, these microRNAs can be utilised in therapeutic drug designs. The given analytical approach can be implemented for other cancers which would be helpful to build up comprehensive microRNA profile in cancers.

1. Benafif, S., Z. Kote-Jarai, and R.A. Eeles, A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). *Cancer Epidemiol Biomarkers Prev*, 2018. 27(8): p. 845-857.
2. Lin, Y.-C., et al., Tumor suppressor miRNA-204-5p promotes apoptosis by targeting BCL2 in prostate cancer cells. 2017. 40(5): p. 396-406.

DNA Methylation Signatures of Area Based Deprivation

Mr Anders Jespersen¹, Dr. Daniel McCartney¹, Dr. Heather Whalley¹, Dr. Riccardo Marioni¹, Dr. Mark James Adams¹, Dr. Kathryn Evans¹, Generation Scotland¹, Professor Andrew McIntosh¹

¹University Of Edinburgh

Low socioeconomic status (SES) is associated with decreased life expectancy and an increased risk of poor physical and mental health. This association is partly due to known environmental exposures such as smoking, increased alcohol consumption and higher body mass index (BMI), all of which have also been associated with changes in DNA methylation (DNAm). To date no large single study has addressed whether SES is associated with differences in DNAm.

We conducted an epigenome wide association study (EWAS) with data from the Generation Scotland cohort (n = 7373) using an area-based measure of deprivation (Scottish Index of Multiple Deprivation). Blood based DNA methylation was obtained using the Illumina HumanMethylationEPIC (850k) DNA methylation Array. We also aimed to identify whether any associations could be accounted for by common environmental factors.

SIMD had widespread (N = 19) associations with differences in DNAm that were attenuated when smoking, BMI and alcohol were included in the model. Significant findings included two CpGs previously associated with smoking (cg05575921 p = 3.88e-09, cg21566642 p = 6.56e-09) and one CpG proximal to the KIAA1671 gene previously associated with recurrent depression (cg15446156 p = 1.49e-08). After adjusting these models for the cg05575921 CpG (AHRR probe) no significant SIMD associations with DNAm remained.

This study demonstrates that low SES is associated with differences in DNAm near genes associated with respiratory disease, recurrent depression, bone mineral density and red blood cell count. These can be largely attributed to known environmental factors, and smoking in particular.

Estimating heritability in honeybees: comparison of three major methods based on empirical and simulated datasets

Dr H el ene Jourdan-Pineau

Biography:

I'm an evolutionary ecologist, interested in traits evolution of agriculturally important species, such as agricultural pest, disease vectors.

A central question in biology is whether observed variation in a particular trait is due to environmental factors or genetic factors. The level of the genetic contribution to phenotypic variation (namely the heritability) determines the response to selection and the possibility to improve characters of economic importance. In honeybee, the haplodiploid sex determination does not allow the straightforward use of classical quantitative genetics methods. Specific methods have been developed for about 40 years but some challenges remain. Our aim was to compare three major methods based on sib-analyses (a model using the average colony relatedness, a half-sibs/full-sibs model, and an animal model) using experimental and simulated datasets.

Our experimental dataset is composed of 5 colonies in Reunion Island and 5 colonies in Mauritius (N=853). All individuals were genotyped to reconstitute the pedigree, and the proboscis and cubital veins length were measured. We simulated phenotypic datasets with varying levels of heritability and common environmental effect and with genetic correlation between traits.

The simulation approach showed that the average colony relatedness performed badly whereas the half-sibs/full-sibs and the animal model gave reliable estimates of heritability and genetic correlations. Using the animal model, we found that wing-related traits had high heritabilities, allowing the use of those characters to discriminate between populations. On the contrary, the proboscis length is more heritable in Reunion Island than Mauritius due to substantial phenotypic plasticity. Finally, the A vein was genetically correlated with the B vein and the palpus length, indicating that those traits could not evolve independently.

Performance of Genomic Prediction using Different Multi-Allelic Genomic Relationship Matrices for Genotyping-by-Sequencing (GBS) Data

Mr Jie Kang^{1,2,3}, Dr Phillip Wilcox¹, Assoc. Prof Michael A. Black¹, Dr Stephen Byrne³, Dr Dan Milbourne³, Dr Marty Faville², Dr Jeanne M. E. Jacobs², Dr Ken G. Dodds²

¹University of Otago, ²AgResearch, ³Teagasc

Biography:

"My name is Jie Kang, I'm a PhD student in quantitative genetics at the University of Otago. My research focuses on identifying and utilizing short haplotypes (shortHaps) in genotyping-by-sequencing (GBS) data to improve genomic prediction. This project is jointly funded by AgResearch (New Zealand) and Teagasc (Ireland)."

Advances in sequencing technologies enable us to characterise variation in the genome of non-model but agriculturally important species. Approaches such as Genotyping-by-Sequencing (GBS) can produce abundant markers at relatively low cost. This has encouraged implementation of Genomic Selection (GS) to accelerate genetic gains in animal and plant breeding, but has also raised the challenge of how to make better use of the genomic information.

Unlike animal breeding, where high-quality reference genomes and well-developed modelling strategies already exist, a versatile analysis pipeline is needed for out-breeding plant species, such as perennial ryegrass (*Lolium perenne*). In addition, we want approaches that take the highly polymorphic nature of ryegrass into consideration when analysing (low-depth) GBS data. We are investigating whether genome-wide prediction can be improved by accounting for short haplotypes or 'ShortHaps', that is, multiple variants in small genomic segments such as those captured within a GBS read

In this study, we explore different ways to construct multi-allelic genomic relationship matrices using ShortHaps, and evaluate the performance of these methods within the genomic best linear unbiased prediction (GBLUP) framework. Latest results will be presented.

Algorithmic framework for assessing and improving heritability models

Mr Anubhav Kaphle

Biography:

Anubhav Kaphle is a second year PhD student with Prof. David Balding at Melbourne Integrative Genomics, The University of Melbourne. His PhD work focuses on developing improved heritability models and statistical methods for prediction of complex traits. His interest also lies in trans-ethnic genetic studies and transferability of genetic models in non-european populations.

There is an ongoing debate regarding the best approach to model how heritability varies across the genome. Different groups have proposed different heritability models, and there is a lack of robust framework to assess which models are most realistic. Recently, Speed et al. proposed a new likelihood approximation for genome-wide statistical models based on summary statistics, which provides a statistical framework for assessing heritability models. This framework also facilitates the estimation of model parameters as well as SNP heritability using Maximum Likelihood-based Estimation (MLE) techniques. In this work, we propose an efficient implementation of the statistical framework proposed by Speed et al., which employs a stochastic gradient descent-based algorithm, ADAM. The software can be used to estimate heritability parameter as well as heritability shared by different SNP categories for any assumed heritability models. We used the software to estimate these parameters for some UK biobank traits. We also estimated heritability contributed by eQTLs and heritability enrichment across tissues using GTEx data. Our method also allows for fine-scale model selection based on huge datasets to identify traits-specific heritability model.

Keywords: Likelihood approximation, stochastic gradient descent, SNP heritability, model selection, UK Biobank, summary statistics

Estimation of genetic and phenotypic trends for service based heifer and cow fertility traits for South African Holstein cattle

Ms Ramadimetje D Kgari^{1,3}, Prof Kennedy Dzama³, Dr Carel J Muller³, Prof Mahlako L Makgahlela^{1,2}
¹Agricultural Research Council, ²University of Free state, ³Stellenbosch University, South Africa

The negligence of fertility due to its low heritability while selecting for increased milk yields led to a reduction in reproductive efficiency in the past years. However, inclusion of fertility into dairy cattle breeding heightened worldwide from as early as the 1990s due to its economic importance. This study aims to estimate breeding values (EBV), genetic and phenotypic trends for heifer and cow fertility traits. Data from on-farm management systems included 64 464 records from 18 South African Holstein herds. The EBV were estimated using THRGIBBSF90 of BLUPf90 family of programs . The average breeding values for heifer traits did not show any particular trend. However, the phenotypic trends of heifer traits were showing a downward trend, with a decrease of 0.14 and 0.13 months/year for age at first service and age at first calving, respectively. Phenotypic trends for cow fertility traits appeared to be undesirable, with average increases of 0.16 and 0.83 days per year for calving to first service and days open, respectively and 0.02 more services per year for number of services per conception. In general, average breeding values for all traits explored did not show any particular trend. Similarly, phenotypic trends were fluctuating over the period studied. This could mean that the breeders are not assessing the herd reproductive performance. Genetic and phenotypic trends that appeared to follow trait improvement could have been coincidental. Thus, efforts have to be made towards the improvement of heifer and cow fertility traits in this study.

A practical approach for optimised partitioning of genomic relationship across chromosomes

Dr Majid Khansefid¹, Dr Mohammad Ferdosi²

¹*AgriBio Centre for AgriBioscience, Agriculture Victoria, Bundoora, Australia*

²*Animal Genetics and Breeding Unit, University of New England, Armidale, Australia*

Biography:

Majid studied his BSc and MSc in animal breeding and genetics at the University of Tehran, Iran. He did his PhD on "Genomic selection for residual feed intake" with the University of Melbourne and DairyBio in 2016. He joined Agriculture Victoria Research (AVR) as a computational biologist in 2017. He is interested in using omics data and computational theoretical principles to improve genomic predictions for complex traits in sheep and dairy cattle.

In genomic best linear unbiased prediction (GBLUP) models, genotyped markers are used to make a single genomic relationship matrix (GRM) and consequently each marker contributes similarly in explaining the genetic variance of traits. Some new methods incorporate markers effects in genomic prediction by applying different weights to markers in the GRM. These models often show small improvements in accuracy, but sometimes an increase in the bias of prediction. Alternatively, multiple GRMs made from markers located on each chromosome can be fitted in a GBLUP model. So, the chromosomes containing mutations with large effects on a trait can be used to explain more of the genetic variance. However, fitting many GRMs in a model is not always practical. In this study, for analysing final weight in Hereford cattle (2n=60), initially, we ran 30 models with 2 GRMs made from markers located on each chromosome (GRM_chr) and the markers from the remaining chromosomes (GRM_remaining-chrs). We found GRM_chr for chromosome 6 and 20 explained 20% and 23% of the total genetic variance, respectively, but the rest of GRM_chr failed to absorb any variance. Finally, the prediction model with 3 GRMs, GRM_chr for chromosome 6 and 20 and GRM_remaining-chrs, explained 22%, 26% and 52% of genetic variance, respectively, and compared to the model with a GRM made from all markers, log-likelihood was improved significantly ($p < 0.001$). Although, our results show potential in improving the goodness-of-fit of genomic prediction model, further analyses are required to validate the improvement in accuracy of genomic prediction.

Multi-tissue transcriptome-wide association study identifies genetic mechanisms underlying endometrial cancer

Miss Pik-Fang Kho¹, Endometrial Cancer Association Consortium, Dr Gabriel Cuellar Partida², Dr Amanda Spurdle¹, Dr Dylan Glubb¹, Dr Tracy O'Mara¹

¹Qimr Berghofer Medical Research Institute, ²Diamantina Institute

Despite continued efforts to identify genetic variants associated with endometrial cancer risk, mechanistic insights underlying its genetic predisposition are lacking. We performed transcriptome-wide association study (TWAS), integrating GWAS summary statistics of endometrial cancer (12,906 cases and 108,979 controls of European ancestry) and eQTL summary data across 48 tissues to identify candidate susceptibility genes underlying endometrial cancer risk. TWAS identified 26 candidate genes for endometrial cancer (FDR<0.05), of which 21 genes had not been previously reported as candidate genes. A majority of identified candidate genes (70%) were supported by colocalization analysis. Probabilistic fine-mapping analysis using FOCUS prioritised candidate genes BHLHE41, SRCIN1, SNX11, and RNF217 (posterior probability for causality>0.8). By comparing gene expression imputed from GWAS data with drug-induced gene expression profiles from the Connectivity Map database, we identified ten drug-repurposing candidates including an inhibitor of mTOR, which has shown promise as an endometrial cancer drug target. Exploring pleiotropic effects of TWAS-identified candidate genes in CTG-VIEW found eight candidate genes that were associated with endometrial cancer risk factors such as menopause, diabetes and body mass index. In summary, using TWAS we have prioritised candidate endometrial cancer risk genes for future experimental study, and highlighted compounds that could be repurposed for endometrial cancer treatment.

The genetics of peripartum depression

Dr Jacqueline Kiewa

Biography:

Jackie Kiewa completed her Bachelor of Applied Science (with Distinction) with RMIT (Melbourne) in 2017 and completed the Master of Bioinformatics with the University of Queensland in 2019, winning a prize for the highest grade point average for her program. Since then her PhD has focused on the genetics of peripartum depression. Previously, Jackie worked in education, and had a teaching role at Griffith University (Brisbane), working in the field of outdoor education and adventure therapy.

Background

The period around pregnancy and after childbirth is a time of increased risk of depression, termed perinatal depression (PND). PND is often conceptualised as a subset of major depressive disorder (MDD), and a prior history of MDD is the strongest predictor of PND. It is of interest to understand to what extent genetic risk factors are shared between PND and MDD. We conducted a GWAS of PND and estimated the genetic correlation with MDD.

Methods

Data (3,804 cases, 6,134 controls) came from the Australian Genetics of Depression Study (AGDS) and QSkin. LD score regression was used to compare significant genetic correlations for both PND and MDD. MAGMA was used to assess if gene association results were enriched in tissues. Polygenic Risk Scores (PRS) were constructed for six psychiatric disorders/traits and tested for association with PND.

Results

No variants were significantly associated with PND after genome-wide correction. Genes expressed in reproductive tissues were significantly associated with PND but no evidence of enrichment in the brain was found. Eleven traits were significantly genetically correlated with both PND and MDD. Regression analysis found that the PRS for all six disorders/traits were significantly associated with PND.

Conclusion

The r_G values for PND and MDD supports the hypothesis that much of the genetic risk is shared between the two disorders. For MDD, gene expression was significantly enriched in the nervous system and brain tissue. For PND, gene expression significantly enriched in the reproductive system may support its conceptualisation as a distinct form of depression.

Genome-wide chromatin accessibility in the developing ovary of the Black Tiger Prawn

Dr James Kijas

Biography:

Dr Kijas is a genome scientist who completed his PhD in plant genetics (1996) and has since built a research career focused on the genetics and genomics of domesticated animals. He joined CSIRO in 2003 and was instrumental in the development and delivery of a range of genomic tools for sheep. In 2016 he shifted direction to focus on aquaculture species and built a research focus on Atlantic salmon. The last two years has seen his research focus widen to include development of an applied breeding program for barramundi, functional genomics of crustaceans and SNP tool development in abalone.

Black Tiger prawn (*Penaeus monodon*) is a major aquaculture production species farmed locally in Queensland and interest is returning in its production globally. Breeding programs are relatively new in comparison to terrestrial livestock species, and there is a lag in the development of the genomic tools needed to implement genomics into applied breeding and understand the biology of key traits. Draft genome assemblies are now emerging, prompting us generate the datasets needed to construct a first pass functional annotation. Multiple tissue types were collected from farmed animals, with a particular focus on ovary at different stages of maturation prior to spawning as reproductive performance remains an issue in some breeding scenarios. RNA-Seq and ATAC-seq were performed to collect both transcriptomes and chromatin accessibility data to identify gene regulatory elements. Analysis of both datatypes against the available reference assembly provides a first pass functional annotation. We report the tissue specific distribution and number of putative cis regulatory elements, and their proximity to protein coding genes. We also present the association of chromatin accessibility status and gene expression, which to our knowledge is yet to be assessed in a crustacean. The findings represent the first step towards enabling biology driven genomic selection, and provides baseline information on reproductive performance in this important aquaculture species.

Heritability estimates in salmon using different degrees of relatedness

Mr Panagiotis Kokkinias¹, Dr Pau Navarro², Professor Ross D. Houston¹, Dr Alastair Hamilton³, Dr Ricardo Pong-Wong¹, Professor Chris S. Haley^{1,2}

¹The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, ²MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, ³Hendrix Genetics Aquaculture B.V.

Biography:

I am a PhD student at The Roslin Institute (University of Edinburgh), working in the optimisation of genomic breeding and the determination of the genetic architecture of complex traits of farmed Atlantic salmon. I am most interested in the field of population and quantitative genetics and their applications in aquaculture. In particular, selective breeding and applications of genetic markers are very fascinating areas for me. Moreover, I have work experience as a trainee assistant manager in a salmon hatchery and in various research projects related with fish developmental stages, nutritional requirements and deformities.

Accurate heritability estimations can provide essential knowledge for the design of an Atlantic salmon breeding program. Thus, we estimate the heritability, based on both closely and distantly related individuals, by estimating pedigree-based and SNP-based heritability simultaneously in a single model. This allows us to estimate the proportion of genetic variance that can be captured using genomic data from single nucleotide polymorphisms (SNPs) in distantly related individuals versus that which is only associated with close genetic relationships within a pedigree. Similar studies in humans show that the SNP-based heritability explains around one-half of the total heritability. Heritability estimates for the SNP and pedigree-based components can enhance our understanding of the variation that can be explained by SNPs at a population-wide level.

Our analysis examined multiple quantitative phenotypes in various populations of salmon, with 11K and 33K SNP densities. The results show that the SNP-based relationships explain the majority of the genetic variance (90%), where in human studies explain around 50% of the total genetic variance. Despite the lower number of genotyped SNPs, compared the human studies, the genome coverage of the SNP data is higher, potentially due to stronger pedigree structure and greater linkage disequilibrium in the salmon populations. Different thresholds of relatedness were used to create pedigree-based matrices and estimate SNP-based heritability. Changing the thresholds had no effect on the estimates of variance components. Finally, the higher SNP density resulted in greater heritability and genetic variance estimates potentially as higher linkage disequilibrium captures the effect of more causative variants.

Trait stability and broad adaptation over forty years of selection in a global wheat breeding program

Dr Margaret Krause

Biography:

Margaret Krause is a postdoctoral fellow working with the Global Wheat Program at the International Maize and Wheat Improvement Center (CIMMYT) in Mexico. She completed a PhD in Plant Breeding at Cornell University in 2019. She was awarded a Postdoctoral Research Fellowship in Biology from the U.S. National Science Foundation to work with Dr. Matthew Reynolds of CIMMYT and Dr. Jesse Poland of Kansas State University to leverage forty years of CIMMYT's international yield trial records to explore trait stability and broad adaptation in wheat as well as the influence of artificial selection on the genome over time.

Crop varieties that are able to maintain consistent yields across a range of environmental stresses are of value to farmers, particularly those who are resource-poor with limited ability to recover from crop losses. Understanding the genetic basis of trait stability across environmental gradients will enhance our ability to develop broadly adapted, climate-resilient crop varieties. However, trait stability is expensive and laborious to measure due to the extensive multi-environment trials required to produce reliable estimates. Since 1979, the International Maize and Wheat Improvement Center (CIMMYT) in Mexico has sent wheat breeding material to collaborators worldwide for evaluation each year. Over forty years, CIMMYT has amassed millions of agronomic data points from more than 500 locations ranging from Norway to New Zealand and spanning over 3,500 m in elevation. These trials provide a rare opportunity to improve our understanding of the genetic controls of trait stability in wheat and inform selection strategies. In addition, coupled with whole genome resequencing of the breeding program's founder lines, this dataset enables an exploration of the effects of artificial selection for broad adaptation on the wheat genome over time.

GWAS of transferrin N-glycans: one step closer to understanding the genetics of protein glycosylation

Arianna Landini¹, Pau Navarro², Irena Trbojevic-Akmacic³, Frano Vuckovic³, Caroline Hayward², Gordan Lauc^{3,4}, Lucija Klaric², James F. Wilson^{1,2}

¹Centre for Global Health Research, Usher Institute, University of Edinburgh, ²MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, ³Genos Glycoscience Research Laboratory, ⁴Faculty of Pharmacy and Biochemistry, University of Zagreb

Biography:

Arianna Landini is a final year PhD student at University of Edinburgh and a Marie Skłodowska-Curie Early Stage Researcher part of the IMforFUTURE project. She is interested in human population genetics, evolutionary genetics and genetic architecture of complex traits. Her current research goal is to investigate the genetic regulation of mainly N-glycomic traits, but also other omics (e.g. lipidomics), with a focus on low frequency and rare genetic variants. Omics data are in fact extensively employed in genetic association studies as 'proxies' for complex traits and diseases. While biochemical pathways involving omics are often well characterised by in-vitro and in-vivo experiments, there is often a lack of understanding about their genetic regulation.

Glycomics, studying the collection of glycans in biological systems, is an emerging field among omics data. Despite glycans being involved in the aging process and in a wide variety of diseases (including Parkinson's disease, rheumatoid arthritis, Crohn's disease, type 2 diabetes, cancer), genetic regulation of glycosylation is yet not fully understood.

To help address this knowledge gap, we performed the first study integrating genomics and glycomics of the blood plasma glycoprotein transferrin.

We investigated the genetic regulation of transferrin N-glycosylation and compared results to state-of-the-art knowledge about genetic regulation of immunoglobulin G (IgG) N-glycosylation.

We performed a genome-wide association study of 35 UPLC-quantified transferrin N-glycan traits in two genetically isolated cohorts (discovery N=948, replication N=959) using Haplotype Reference Consortium imputed genotypes.

We identified 10 genome-wide significant loci (p -value $< 5 \times 10^{-8}/35$ traits), mapping in genes encoding glycosyltransferases (MGAT5, FUT6, FUT8, ST3GAL4, B3GAT1), genes potentially related to the glycosylation process (MSR1, TUSC3) and transferrin (TF), or previously associated with N-glycosylation of IgG (NXPE1, NXPE4). 3 of these loci (TF, MSR1, TUSC3) have not previously been associated with glycosylation and 6 replicated (MGAT5, TF, ST3GAL4, B3GAT1, FUT8, FUT6).

We suggest that while some genes are responsible for regulating the N-glycosylation of both transferrin and IgG, others instead appear to control N-glycosylation of only one of these glycoproteins. Our findings shed light on the complex mechanisms regulating protein N-glycosylation and corroborate the notion that while some genes specifically glycosylate a restricted number of proteins, others instead affect the N-glycosylation of multiple proteins.

Improving Phytophthora Root Rot Resistance in Chickpea using Genomic Selection

Dr Yongle Li

Biography:

Yongle (Leo) Li received his PhD from the Technical University of Munich in 2012 in the area of quantitative genetics and plant molecular breeding. After finishing his PhD, he joined the Global Wheat Program of CIMMYT in Mexico as a post-doctoral fellow implementing genomic selection method into the wheat breeding program. In 2013 he took up another post-doctoral fellow position at the Waite campus of the University of Adelaide to develop and implement molecular breeding tools to assist chickpea and oat breeding. His research interests are in the fields of quantitative and population genetics.

Phytophthora root rot (PRR) is one of the major fungal diseases of chickpea in NSW and Queensland which are the major chickpea production area in Australia. We have identified more than 10 QTLs for PRR resistance using three RIL mapping populations. Subsequently, we have developed and validated nine KASP markers flanking three major QTL. The KASP markers were used to screen two F2 populations and ~500 advanced chickpea lines for recurrent selection.

Although we have successfully implemented marker-assisted selection (MAS) in the project, we noticed many QTL have relatively small effect size which limited the efficacy of MAS. Therefore, we also investigated the value of an emerging alternative method called genomic selection (GS). GS used all marker across the genome to estimate the breeding value of individual plants and thus shift the focus of marker identification to parent selection. This approach is more relevant to breeding programs as it can select the best parents for crossing, reduce the cost and time of a breeding cycle.

We used the three RIL mapping population to predict the PRR resistance of a set of chickpea germplasm with divers genetic background and the result was promising with a prediction accuracy of up to 0.7. The work is still in progress with the aims of assisting parent selection from F2 populations and advanced lines selection with the purpose of releasing new varieties.

Genomic Selection in Chickpea using Whole Genome Re-sequencing Data

Dr Yongle Li

Biography:

Yongle (Leo) Li received his PhD from the Technical University of Munich in 2012 in the area of quantitative genetics and plant molecular breeding. After finishing his PhD, he joined the Global Wheat Program of CIMMYT in Mexico as a post-doctoral fellow implementing genomic selection method into the wheat breeding program. In 2013 he took up another post-doctoral fellow position at the Waite campus of the University of Adelaide to develop and implement molecular breeding tools to assist chickpea and oat breeding. His research interests are in the fields of quantitative and population genetics.

Chickpea production in Australia has increased dramatically in the recent years, mainly due to the fast growing demand from the Indian subcontinent. Achieving yield potential in chickpea is limited by many yield constraints such as biotic and abiotic stresses. Phytophthora root rot (PRR) is one of the major root diseases of chickpea in NSW and QLD which are the major chickpea production areas in Australia.

We re-sequenced 310 chickpea advanced breeding lines using Illumina next-generation sequencing technology. More than half million SNPs were discovered with a genome coverage of 5-10X. Analysis of population structure revealed a distinct group of 75 breeding lines with many unique alleles difference from Australian varieties released recently. Genome-wide association studies (GWAS) identified several SNPs significantly associated with PRR resistance and grain yield evaluated in seven field locations. Reduced level of nucleotide diversity and the long extent level of linkage disequilibrium suggested some regions may have gone through selective sweeps probably caused by selective breeding of PRR. We further investigated the effect of genotype by environment interaction on genomic prediction, which is another approach to deal with complex traits. We found that the training set had better prediction accuracy if phenotyped in the environments relevant to the targeted environments. We also investigated the effect of SNP function on prediction accuracy using different subsets of SNPs based on their annotation such as SNP located in regulatory, exon, and alternative splice site regions. These results have important implication for implementing genomic selection into plant breeding programs.

An atlas of pleiotropy for the human genome

Ms Ting Li

Biography:

Ting is a PhD student from Biostatistic group in Sun Yat-sen University, supervised by Prof. Xia Shen after achieving her BSc degree in 2018. Her research interest is in developing novel statistical method especially mixture model contributed to quantitative genetics

Pleiotropy describes the shared genetic basis across complex traits. As a ubiquitous phenomenon across the genome, the concept of pleiotropy has been studied for over 100 years. Quantifying the overall magnitude of every single locus' genetic effect on the widely measured human phenome is of great challenge. We introduce a unified modeling technique that can consistently provide a total genetic contribution assessment (TGCA) of a gene or genetic variant, without thresholding genetic association signals. We performed a genome-wide evaluation of TGCA in the UK Biobank and highlighted top loci such as MHC of the human genome. Tissue-specificity investigation revealed that gene expressions in the brain contribute little to human complex traits variation. We verified this phenomenon by functional enrichment of TGCA around transcription start sites in different tissues. TGCA provides a genome-wide atlas resource for the overall genetic contributions in humans.

Identification of Predictor Genes of Feed Efficiency in Beef Cattle by Applying Machine Learning Methods to Multi-Tissue Transcriptome Data

Dr Yutao Li

Biography:

Dr Yutao Li is a Senior Scientist in the Livestock and Aquaculture Program at CSIRO Agriculture and Food. She received a PhD in Quantitative Genetics from the University of New South Wales in 1995. Prior to CSIRO, she worked as a Research Officer at the Sheep Industry Branch of the Western Australian Department of Agriculture, and a Quantitative Geneticist at ForBio Research Pty Ltd. Currently, her major research interests have been: 1) genomic selection and prediction; and 2) developing and applying machine learning methods for predicting future individual performance using high dimensional genetics, genomics and sensor datasets.

The rapid development of high throughput technologies boosted the generation of large RNA-seq datasets aiming to investigate predictor genes of economically important traits, such as feed efficiency (FE) in beef cattle. In this sense, machine learning (ML) methods are arising as promising accurate predictors of superior animals. In this study, we used RNA-seq data of skeletal muscle from 9 high FE and 9 low FE Nellore bulls to evaluate the prediction accuracies of four analytical methods in classifying animals as either high or low FE. Analysis included three ML methods (Random Forests (RF), Supporting Vector Machine (SVM) and Extreme Gradient Boost (XgBoost)) and the conventional edgeR. RF, Xgboost, and edgeR were used to identify potential predictor genes individually, then the reliability of different combinations of top-ranking predictor genes (sorted by Gini, Gain, and p-value for RF, Xgboost, and edgeR, respectively) was verified by SVM. Furthermore, a method that combined RF and Xgboost (RX) was also explored. Firstly, RF was applied to pick up positive genes (Gini > 0) which were processed through Xgboost to identify molecular-predictor genes. Overall, the RX was found to outperform other ML methods in isolation. A set of 39 potential predictor genes were identified in skeletal muscle. Further analyses are being conducted to explore predictor genes in the other four tissues (i.e., adrenal gland, hypothalamus, liver, and pituitary).

Utility of Anthesis-Silk Interval Information to Improve Maize Grain Yield Predictability under Nitrogen Limited Conditions

Dr Dayane Cristina Lima

Biography:

Dayane Lima is a postdoctoral research associate at the University of Wisconsin - Madison under the guidance of Dr. Shawn Kaeppeler and Dr. Natalia de Leon. She graduated with a B.S. in Agronomy from the Federal University of Lavras. There, she concluded her master's in Genetics and Plant Breeding. In recognition of her outstanding performance, she transitioned direct and early to the PhD, in which part of it was at West Virginia University. Dayane's current research includes abiotic stress tolerance in maize and genomics-assisted breeding.

Lower grain production is expected under nitrogen (N) deficient conditions. Delayed silking, measured as the anthesis-silk interval (ASI - the period between pollen shed and silking), is a good indicator of response to N and water stress in maize. The aim of this study was to evaluate how ASI is affected by N availability and to study the utility of ASI for the genomic prediction of yield.

A set of 400 hybrids derived from 13 biparental populations were evaluated under optimal N (HN) and low N (LN) availability in Hancock, Wisconsin in 2018 and 2019. The experimental design was a complete randomized block with two replications. Anthesis and silk time in growing degree days (GDD), and grain yield (bu/a) were measured. A set of 20,981 SNPs was used for genomic prediction, in which GBLUP models were employed. Prediction ability was calculated through 5-fold cross-validation as the correlation between genomic predictions for ASI and adjusted means of yield.

For the first year of evaluation, under HN conditions, average ASI and grain yield were 46.6 GDD and 140.2 bu/a, respectively. Limiting N significantly increased ASI (54.01%), and reduced grain yield (26.7%) compared to HN. The correlation between these two traits was -0.53. The accuracy of predictions ranged from 0.06 to 0.4, depending on the treatment predicted.

ASI was a good indicator of stress, however, it did not improve yield predictability. We expect a higher accuracy with the inclusion of a second year of data, which will be part of the poster presentation.

The genetic architecture of 25 hydroxyvitamin D concentration

Dr Tian Lin

Biography:

Tian is working as a research assistant in PCTG group with Professor Wray. She received her PhD in Plant Biology from Iowa State University in 2014.

Vitamin D deficiency is associated with a range of adverse health outcomes. It is highly impacted by environmental factors (e.g. season, skin covering, outdoor behaviour), but family and genome-wide association studies have shown that genetic risk factors also have an impact. Here, we used data from the UK Biobank to describe the genetic architecture of 25 hydroxyvitamin D (25OHD), the main metabolite used to assess vitamin D status. Our sample included a set of 58,738 close ($r > 0.2$) relatives from whom we estimated the heritability of 25OHD to be 0.32 (s.e. = 0.01). From a set of 50,000 unrelated individuals, and using a model that fit a single genomic relationship matrix constructed from all single nucleotide polymorphisms (SNPs), the SNP-based heritability estimate (h^2_{SNP}) was 0.13 (s.e. = 0.01). Bivariate GREML showed that h^2_{SNP} assessed in summer was significantly higher ($P = 1.5 \times 10^{-3}$) than in winter (0.19, s.e. = 0.02 vs. 0.10, s.e. = 0.02). The estimates of heritability (both family and SNP-based) were hardly impacted with BMI included as covariate. Using SBayesS, we estimated that about 0.8% of the ~1.1 million HapMap3 panel common SNPs affect variation in 25OHD, a lower polygenicity estimate than that of most complex traits. Lastly, the SBayesS S parameter, which represents the relationship between minor allele frequency and effect size, was -0.78 (s.e. = 0.04), providing evidence of negative selection on the genetic variants associated with 25OHD levels. These findings provide new insights into the genetic basis underlying 25OHD status.

Estimating population abundance of a critically endangered species via close-kin mark-recapture

Dr Luke Lloyd-Jones¹, Dr Mark V Bravington¹, Mrs Emma Lawrence¹, Dr Kyle N Armstrong², Dr David A Westcott³

¹Data61, CSIRO, ²University of Adelaide, ³Land and Water, CSIRO

Biography:

Luke is a statistician currently applying his skills at the interface of methodology development, ecology and genetics. Luke currently works on a broad range of Australian ecological problems in the Great Barrier Reef, Murray-Darling Basin and on close-kin mark-recapture – a powerful new technique for estimating population size and rates of change for wild species. Luke Lloyd-Jones is a research scientist at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), which is Australia's national science agency. Before joining CSIRO, Luke worked on exciting problems in human quantitative genetics within the group of Professors Peter Visscher, Naomi Wray and Jian Yang.

Mark-recapture (MR) is a powerful methodological framework for studying wild populations. The recent close-kin MR (CKMR) methodology, utilises modern genotyping technology to recover the genetic tag of an individual's closely-related kin, rather than the individual. Given the observed kin, classical MR demographic models can then be applied via a likelihood that incorporates species-specific kin-recapture probabilities. CKMR greatly generalises the application space of MR and is highly useful for populations that are difficult to survey due to high-mobility, habitat access and knowledge of a species' total range. The critically endangered Christmas Island flying fox (*Pteropus natalis*) is one such species and requires urgent knowledge of population abundance and rate of population change. We use a species-specific set of approximately 1,500 single nucleotide polymorphisms (SNPs) derived from diversity arrays technology sequencing to infer close-kin status between pairs from a sample of ~500 Christmas Island flying foxes. Given these kin, we use the CKMR methodology to estimate population abundance and the retrospective rate of change of this species, which are essential to the future management of this critically endangered species.

Effects of population parameters and genetic models on power and accuracy of GWAS

Mr Zhi Loh

Biography:

My name is Zhi Loh and I have received a Bachelor in Natural Environment and Wilderness Studies from University of Tasmania in 2018. I then received a Bachelor in Science with Honours (majoring in genetics) from University of New England in 2019. Currently, I am studying in the second year in Doctor of Philosophy (majoring in genetics) at University of New England. Under supervision from Julius van der Werf and Sam Clark, my current focus is in designing an algorithm for estimation of breeding values of animals using genomic data, while incorporating non-additive genetic components and inbreeding.

Genome wide association studies (GWAS) are commonly used for the detection of associations between genotypes and phenotypes. There is often a debate regarding whether associated regions detected using GWAS are true causative regions. In this study, factors that affect detection of quantitative trait loci (QTL) were evaluated, with special emphasis on the effects of selection, demography, data size, QTL effect size, QTL number, allele frequency and non-additive genetic components such as dominance and epistasis. A population was simulated using a gene drop algorithm with selection. We found that the detection rate and the accuracy of estimated effects of QTL is significantly affected by selection. Strong selection process increases number of false positives, and reduces detection rate. The QTL detection rate and accuracy of estimated effect sizes decrease with small effective population size and low heritability, which is consistent with previous studies. Small sample sizes reduce the number of detected QTLs, with minimum required sample size of 800. Increasing the number of QTL in a trait reduced detection rate and accuracy, as QTL effects are smaller. QTL with intermediate allele frequency had higher detection rate and accuracy. Epistasis had increased false positive rate and biased the estimated QTL effect sizes. These findings are important and useful when interpreting results from GWAS studies that were conducted on populations that have undergone intense selection process or with small effective population sizes.

DNA Methylation Risk Score for Pancreatic Cancer

Dr Evangelina López De Maturana^{1,2}, Mrs L Alonso^{1,2}, Mr C Lumbreras¹, Dr N Malats^{1,2}

¹Genetic and Molecular Epidemiology Group, CNIO, ²CIBERONC

Biography:

Evangelina López de Maturana works as staff scientist in the Genetic and Molecular and Epidemiology group, at the Spanish National Cancer Research Centre. In 2007, she got her PhD degree with the Europe and Doctorate Mention in the quantitative genetics field. Since 2010, Evangelina has focused her research on the Genetic Epidemiology field, particularly on the application of state-of-the-art statistical methods to omics data to deal with large p small n problem in order to identify individuals at high risk of developing either pancreatic or bladder cancer. She has also studied the ability of alternative biomarkers to predict these cancers.

Background: Pancreatic cancer (PC) has a bad prognosis, mainly because of its late diagnosis and aggressive nature. Discovering new biomarkers to define high-risk populations is a must towards an early diagnosis. DNA methylation (DNAm) in peripheral blood has been barely explored as PC-associated marker and only at the individual CpG level, not capturing the whole DNAm variability. We aimed to build a whole-methylome risk score (WMRS) based on epigenome-wide CpG aggregation associated with PC risk.

Methods: Epidemiological data and leukocyte-DNA from 338 cases and 285 controls included in the PanGenEU study were used. DNAm was determined with the Infinium MethylationEPIC array. WMRS estimates were obtained using a Bayesian kernel-based regression adjusted for immune cell composition, age, gender and region of recruitment. Kernel was based on the pairwise similarity of epigenome profiles and further explored using clustering techniques. The predictive ability of WMRS was computed as the averaged area under the curve (AUC) in a 10-fold cross-validation scenario. Finally, we associated the WMRS with potential PC risk factors.

Results: We identified four clusters of individuals based on their DNAm similarity, mainly driven by immune cells composition. Epigenetic variance explained ~59% of the phenotypic variance. The predictive ability of WMRS was high (AUC=0.70±0.05). WMRS associated with a higher risk of being diabetic, smoker, and having family history of PC, though with a lower risk of suffering from nasal allergies.

Conclusions: WMRS may represent a useful tool to identify individuals at high-risk of PC.

Minimum number of generations to start detecting purging

Dr Eugenio López-Cortegano

Biography:

Eugenio López-Cortegano is a postdoc in Peter Keightley's team at the University of Edinburgh. There, he investigates the nature of de novo mutation and its evolutionary impact, as well as the genomic factors that influence mutability. He also maintains active research and collaborations regarding the development of quantitative models and methods for the study of small and fragmented populations.

Minimum viable population (MVP) sizes have been proposed to determine whether a given endangered population is likely to survive in the short and long term (e.g. the 50/500 rule). However, these recommendations only take inbreeding depression into consideration, neglecting the chance of selection to purge deleterious alleles. Under genetic purging, small populations are expected to reduce their inbreeding load, making them possible to survive even with critically small effective population sizes, yet purging is also challenging to detect in practice, as it is expected to be less efficient under faster inbreeding, and more delayed in time for larger populations. Here I propose a minimum number of generations to start detecting purging, based on the maximum rate of increase of ancestral inbreeding, given its close relation to purging. This number can be approximated as $t \approx \sqrt{2N}$, indicating that purging should become relevant faster for smaller populations, but that a temporal dimension is indeed required to start detecting it (e.g. no less than 10 generations with $N=50$). This time point is compared to predictions from a different inbreeding-purging model, showing a good fit to the time point for fitness partial recovery under inbreeding, if purging is relevant. In consequence, conservation policies should take into consideration not only MVP sizes, but should also that a number of generations is needed to evaluate the possible consequences of purging, which can make a difference in terms of how likely small inbred populations are to survive.

Convergence within divergence: insights into wheat adaptation from Triticum population sequencing

Dr Fei Lu

Biography:

Fei Lu and his research group combine multiple approaches from genomics, quantitative genetics, evolution, and computer science to dissect agronomic traits of wheat, especially yield and its components (e.g. heterosis, biomass, abiotic stress resistance). They seek to identify genetic variants shaping yield-associated traits and investigate how these variants interact with different environments. Fei's group is developing functional genome prediction approaches to pinpoint causal variants of these traits. The successful development of the set of approaches will tremendously expand the application of genomic editing technologies in the improvement of quantitative traits.

Bread wheat expanded its habitats from a core area of the Fertile Crescent to global environments within ~10,000 years. Genetic mechanisms of this remarkable evolutionary success are not well understood. By whole-genome sequencing of populations from 25 subspecies within genera *Triticum* and *Aegilops*, we identified composite introgression from wild populations contributing to a substantial portion (4-32%) of the bread wheat genome, which increased the genetic diversity of bread wheat and allowed its divergent adaptation. Meanwhile, convergent adaptation to human selection showed 2- to 16-fold enrichment relative to random expectation — a certain set of genes were repeatedly selected in *Triticum* species despite their drastic differences in ploidy levels and growing zones, indicating the important role of evolutionary constraints in shaping the adaptation landscape of bread wheat. These results showed the genetic necessities of wheat as a global crop and provided new perspectives on transferring adaptation success across species for crop improvement.

Dissecting genetic pleiotropy between hyperuricemia and chronic kidney disease using local Bayesian regression methods

Ms Alexa Lupi

Biography:

Alexa Lupi is a 3rd year PhD student in Biostatistics at Michigan State University (MSU) in East Lansing, MI, US. She is interested in statistical genetics, working with high dimensional data, and epidemiological studies. She is a member of the Quantgen research group at MSU, headed by Drs. Ana Vazquez and Gustavo de los Campos.

Chronic kidney disease and hyperuricemia are closely related comorbidities, yet their shared genetic variants and regions of pleiotropy are not well-defined. These regions can be identified through obtaining genetic covariance estimates, but existing methods to obtain these estimates, like multi-trait models, are not currently computationally feasible with a sample size as large as that of the UK Biobank (over 300,000 unrelated Caucasians). We implemented a series of local Bayesian regressions across the entire genome to obtain genetic covariance point estimates between chronic kidney disease and hyperuricemia within SNP segments in local linkage disequilibrium. We applied resampling methods to obtain interval estimates for segments in a filtered set of genomic regions (about 3.6% of the segments) and found 267 SNP segments with statistically significant covariance estimates (17 positive, 250 negative), implicating 188 distinct shared loci. Some of these significant shared loci are validated with consistent directionality to the covariance estimates we found, including all 9 loci identified as shared by Johnson et al. 2018, and 24 out of the 35 shared loci identified by Leask et al. (accepted). Numerous novel shared loci were also identified, such as THBS3/MTX1/GBAP1, GLI2, SLC7A9/CEP89, CYP24A1, KCNS3, CHD9, ARL15, PAX8, and IGF1R. To examine potential biological mechanisms for these shared loci, we performed colocalization analyses.

What if all variants are rare?

Dr Danilo Lyra

Biography:

Danilo is a crop quantitative geneticist at Rothamsted Research. He holds an M.S. and Ph.D. degree in Genetics and Plant Breeding having an extensive background in potato, maize, and wheat breeding/genomics. Danilo has experience in modern statistical genomics approaches where he focuses on understanding the genetic architecture of quantitative traits. Currently, he is working on three projects with extensive collaborations in hexaploid wheat: 1) combining high-throughput plant phenotyping and molecular genetics for trait dissection; 2) gene-based mapping of trehalose biosynthetic pathway genes for source- and sink-related traits; 3) development of a quantitative genetics pipeline for the wheat Cadenza TILLING population.

The relative contribution of rare variants to quantitative genetic variation is largely unclear, despite recent advances in genomics. One reason for this is that natural populations tend to violate the assumptions of genome-wide association and genomic heritability models, with population structure and cryptic relatedness being particularly problematic. Another complication is that the contributions of common and rare variants are difficult to disentangle because of linkage disequilibrium. To eliminate these limitations, we performed quantitative genomic analyses in an experimental TILLING population of wheat (~1100) in which all chemically induced mutations were either rare or unique, starting with a (mostly) uniform genomic background. Using exome re-sequencing data, we partitioned the complete set of ~7M mutations into non-synonymous, synonymous, and non-coding sets, and then conducted association and variance component analyses for three phenotypic traits (i.e. plant height, flowering/heading time, and senescence). By scanning around 77K genes using several analytical approaches developed specifically for rare variants, we identified significant associations for all phenotypes, with most signals coming from non-synonymous and synonymous substitutions. A few well-known genes (e.g. FLOWERING LOCUS T), as well as many others were detected. In addition, up to a third of the variance for all traits could be explained by the cumulative effects of all mutations. We therefore expect that this approach will lead to the identification of candidate genes across a range of phenotypic traits and contribute to the development of novel analytical methodology, ultimately enhancing our understanding of the importance of rare variants in both natural and experimental populations.

Increasing the accuracy of genomic prediction for crossbred livestock: examples from dairy cattle

Dr Iona Macleod¹, Dr Majid Khansefid¹, Dr Mekonnen Haile-Mariam¹, Dr Ruidong Xiang^{1,4}, Dr Sunduimijid Bolormaa¹, Dr Gerben de Jong², Ms Erin O'Conner³, Dr Chris Schrooten², Prof Jennie Pryce^{1,5}, Prof Michael Goddard^{1,4}, Dr Hans Daetwyler^{1,5}

¹Agriculture Victoria, ²CRV, ³CRV Ambreed, ⁴University of Melbourne, ⁵LaTrobe University

Biography:

Iona MacLeod is a senior research scientist with Agriculture Victoria Research based in Melbourne, Australia, at the Centre for AgriBioscience. After a varied career, Iona completed a PhD in 2013 in quantitative and population genomics and currently works in the field of quantitative livestock genomics.

There is increasing interest in generating accurate genomic predictions for crossbreds because various livestock production systems use crossbreds as breeding stock and for production. We investigated the accuracy and bias of genomic prediction for crossbred dairy cattle compared to purebreds using several approaches. The traits tested were milk, fat and protein yields. Genomic prediction was implemented with a Bayesian method and a custom set of 46,000 sequence genotypes selected from a multi-omics analysis of whole genome sequence. With these tools we explored the use of different multi-breed training populations: Holstein, Jersey and their crossbreds, where the latter included more complex crosses than just F1. The multi-breed reference populations included: a Holstein dominated set of over 36,000 pure and crossbred animals, and a set of over 15,000 pure and crossbred animals with equalised Holstein and Jersey breed proportions. In addition, we compared the accuracy and bias of genomic prediction using the custom set of sequence genotypes versus standard commercial SNP chip genotypes. The results indicated that it was preferable to equalise breed proportions in the multi-breed training approach rather than have a larger multi-breed set that was dominated by one of the breeds represented in the crossbreds. It was also advantageous to include crossbreds in the multi-breed training population. The selected set of custom sequence variants generally outperformed the standard commercial SNP array, highlighting the potential value of creating custom SNP arrays where variants are selected based on a comprehensive multi-omics data analysis.

Estimation of macro- and micro-environmental sensitivity of genotypes using unbalanced data

Miss Mette Dam Madsen¹, Professor Julius van der Werf¹, Dr. Vinzent Börner², Dr. Sam Clark¹

¹School of Environmental and Rural Science, University Of New England, ²Animal Genetic and Breeding Unit, University Of New England

Biography:

Mette Dam Madsen is a PhD-student at University of New England, Australia, where she is studying environmental sensitivity in livestock, with a special focus on micro-environmental sensitivity. She holds a MSc in biology with speciality in quantitative genetics. Previous work includes modelling macro-environmental sensitivity for growth rate in Danish pigs (master thesis, Aarhus University, Denmark) and modelling group recorded feed intake for feed efficiency in Danish mink (research assistant Aarhus University, Denmark).

Environmental sensitivity describes the dependency of an individual's performance on its environment. If environments can be defined, they are macro-environments, whereas undefinable environments are micro-environments. Previous studies estimating macro- and micro-environmental sensitivity have mostly been using balanced half sib designs, however, in extensive breeding programs field data can be very unbalanced. The aim of this study was to investigate the consequences of an unbalanced half sib design to estimate variance components for macro- and micro-environmental sensitivity. Stochastic simulation of phenotypes included polygenic, macro- and micro-environmental genetic factors, a random herd effect and random environmental effects. The simulated pedigree contained 100 sires with 20, 50, 100 or 200 offspring each (balanced data) or on average (unbalanced). Analysis using a reaction norm model fitting all simulated components was performed in ASReml4.1. Preliminary results, averaged over 20 replicates, showed there was no difference in prediction accuracy of breeding values for macro- and micro-environmental sensitivity between even and uneven use of sires. There was no significant difference between the genetic variance estimated for polygenic, macro- and micro-environmental sensitivity when using balanced versus unbalanced data. Genetic correlations between the polygenic and macro-environmental sensitivity was underestimated when sires had 20 offspring each (balanced) and overestimated when sires had 20 offspring on average (unbalanced). This indicates that uneven use of sires with at least 50 offspring on average does not affect estimates of variance components or breeding values for macro- and micro-environmental sensitivity compared to even use of sires, however more extreme scenarios could produce different results.

Quantifying selection and adaptation on polygenic traits

Dr Medmat Mahmoud

Biography:

A postdoctoral fellow at University of Göttingen, Center for Integrated Breeding Research (CiBreed). My Ph.D. was in Quantitative Genetics from The University of Giessen (Germany), and my thesis was on “Genetic epidemiology of bovine infectious diseases”.

We developed an R package, “Ghat”, that identifies selection on complex traits that are controlled by large numbers of QTLs. This package tests the significance and direction of selection by calculating a composite statistic denoted \hat{G} (Ghat), which is the summation of the estimated effect of every SNP scored multiplied by its effect size. SNP effects are estimated using shrinkage-based genomic prediction estimators. Then, the observed \hat{G} value is compared to an empirical null distribution to assess significance. We applied Ghat to different traits from different species (plants animals and humans) and observed significant increase in the power compared to locus-specific tests for selection. Theory and simulations show that the boost in power increases with increasing the number of QTLs that control the traits, making Ghat appealing for adaptation- and selection-tests in complex traits and when there is a small sample size. The Ghat package is now accessible through CRAN and on GitHub.

Genetic parameters, phenotypic and genetic correlations for female fertility traits derived from AI service records in South African Holstein cattle

Ms Ramadimetje D Kgari^{1,3}, Prof Kennedy Dzama³, Dr Carel J Muller³, **Dr Mahlako Linky Makgahlela^{1,2}**

¹Agricultural Research Council - Animal Production, ²University of the Free State, Department of Animal, Wildlife and Grassland Sciences, ³University of Stellenbosch, Department of Animal Sciences

Productivity and viability of dairy farms depend largely on the soundness of their reproductive performance. Economic losses incurred from poor fertility as a consequence of selection pressure on production heightened inclusion of fertility into dairy cattle breeding worldwide from as early as the 1990s. In South Africa, only age at first calving (AFC) and calving interval (CI) are evaluated for genetic improvement of female fertility. AI service records may not be routinely recorded but are available on farm, for possible inclusion in breeding programmes. This study aims to estimate genetic parameters for age at first service (AFS), number of services per conception for heifers (SPCh) and cows (SPC), calving to first service (CFS), number of days open (DO) and binary traits indicating whether cows were inseminated 80 days' postpartum (FS80d) and pregnancy confirmations within 100 (PD100d) or 200 (PD200d) days postpartum. Data from on-farm management systems included 64464 records from 18 South African (SA) Holstein herds. Genetic parameters and correlations were estimated using THRGIBBSF90 and POSTGIBBSF90 of BLUPf90 family of programs. Heritabilities were low (AFS=0.02±0.04) to moderate (PD200d=0.24±0.00), indicating potential for these traits to respond to selection. The genetic correlations were generally high and positive, with the highest correlations between AFS and SPCh (0.73±0.00), and CFS and SPC (0.90±0.01). The latter indicating that genetic improvement in CFS is coupled with improvement in SPC. The studied traits, including AFC and ICP, could enhance genetic improvement of female fertility in SA Holstein dairy cattle, and are less prone to management biases.

Genome wide association study of methane traits in Danish Holstein

Dr Coralia I.V. Manzanilla-Pech

Biography:

Coralia Manzanilla-Pech is a Postdoctoral Researcher at Aarhus University (Denmark). Currently, she is working mainly on genetics and genomics of methane and feed efficiency in dairy cattle. She finished her PhD in 2017 in Animal Breeding and Genetics at Wageningen University.

Selecting for lower methane emitting animals is one of the best approaches to reduce CH₄ given that genetic progress is permanent and cumulative over generations. For this, identification of the genomic architecture of CH₄ traits are required, as well as the percentage of variance explained by the number of markers in each trait. Therefore, the objectives of this study were to perform a GWAS to identify genes associated with several CH₄ traits in Danish Holstein cattle including CH₄cc (methane concentration), CH₄ g/d, residual methane (RMet, CH₄ regressed on MBW and ECM), methane intensity (MI₁=CH₄/ECM and MI₂=CH₄/BW) to determine if there are genes in common controlling these methane traits. Secondly, the aim was to calculate for each trait the percentage of genetic variance explained by windows of adjacent SNP. Approximately 2,000 cows with genotypes (50K Illumina Bovine Chip) and repeated records (7,227 phenotypes) of CH₄ were analyzed. Strong associations with CH₄cc and CH₄ g/d were found on chromosomes 13 and 26, whereas, for RMet and MI₁ and MI₂ chromosomes 2 and 4. When using windows of 30 to 100 adjacent SNP up to 2.5% of the genetic variance could be explained for CH₄cc and up to 4% of the genetic variance for CH₄ g/d. Based on our results, we conclude that either CH₄cc or CH₄ g/d are feasible traits to select for lower emitting animals. The information resulting from this study could be used as extra information in the genomic prediction of CH₄.

A gene network driving genetic variation in sheep cheese quality

Mr Hector Marina

Biography:

Hector Marina has a B.Sc. in Veterinary Science (2016) from the University of León, Spain. Currently, he is doing his Ph.D. in the Animal Production Department of the University of León under the supervision of Dr. Juan José Arranz. The focus of his study was the detection of the genetic basis of economically important traits in Spanish sheep breeds using genomic tools. During his Ph.D. studies, he also had the opportunity of implementing his skills on bioinformatics, genome-wide association studies (GWAS) and post-GWAS analyses by completing an internship with Dr. Toni Reverter (CSIRO, Australia).

High-quality cheese is the main product of the dairy sheep industry. Considerable genetic variation exists influencing sheep milk protein and fat composition which affect the texture and the melting properties of cheese. This work aims to identify SNPs associated with fourteen traits related to milk composition and cheese-making traits and to detect putative pleiotropic effects of the genes carrying these polymorphisms on the analysed traits. A total of 1,039 ewes from Spanish Assaf breed were genotyped with a custom array of 60k SNPs distributed across the ovine genome. The pleiotropic effects were calculated following the pipeline of Bolormaa et al. (2014). Only the SNPs with the highest pleiotropic effects and located within 20 kb from an annotated gene (Oarv3.1 assembly) were retained (12,426 SNPs). We implemented a stepwise regression approach to identify a panel of 550 SNPs associated with genes explaining over 95% of the genetic variance averaged across all traits. We used the partial correlations and information theory (PCIT) algorithm (Reverter & Chan, 2008) to infer a gene co-association network with regulatory potential for milk composition and cheese-making traits. Further work is underway to validate the regulatory potential of the inferred network in an independent population.

References

- Bolormaa, S. et al (2014). A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. *PLoS Genetics*, 10(3).
<https://doi.org/10.1371/journal.pgen.1004198>
- Reverter, A., & Chan, E.K.F. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics*, 24(21), 2491–2497.
<https://doi.org/10.1093/bioinformatics/btn482>

Integration of CAGE promoter/enhancer analysis of endometrial stromal cells with GWAS data to evaluate gene regulatory mechanisms in endometriosis

Ms Sushma Marla¹, Ms Joanna Crawford¹, Ms Sohye Yoon¹, Ms Stacey Belinda Andersen¹, Ms Jun Xu¹, Ms Sophie Shen¹, Mr Clayton Friedman¹, Dr Nathan Palpant¹, Dr Sally Mortlock¹, Dr Brett David McKinnon^{1,2}, Prof Michael David Mueller¹, Prof Grant William Montgomery¹, Dr Quan Nguyen¹

¹Institute For Molecular Bioscience, University Of Queensland, ²Department of Obstetrics & Gynaecology, Inselspital, University hospital of Bern

Biography:

Sushma Marla is a PhD candidate within the genomics of reproductive disorders group at the Institute for Molecular Bioscience, The University of Queensland, Australia. She has a bachelor's degree in Pharmaceutical sciences (2010) and master's degree in Pharmaceutical Biotechnology (2012). Her current research is involved in the genome-wide scan to identify key genes that contribute to endometriosis and also using novel bioinformatic tools to identify the genetic links which predispose women to endometriosis.

Endometriosis is a common cause of morbidity in women with unknown aetiology. One major approach to discover the fundamental mechanisms and causes of endometriosis is through identification and understanding the functions of the genetic variants responsible for the genetic component of the disease. In total, 9 Genome-wide association studies (GWAS) have been conducted for endometriosis to date, the largest of which comprised 20,933 cases and 482,225 controls of mainly European and East-Asian, ancestry identifying 27 genomic regions strongly associated with the endometriosis risk. This result represents a significant breakthrough toward understanding genetic risk factors for endometriosis. However, the significance can be realized only when regulatory effects of these DNA sequence variants in the tissues relevant to endometriosis are determined. The aim of this study is to examine the potential effects of the reported loci on regulatory genomic regions, including promoters and enhancers. Endometrium is a complex tissue and it is thought cells from the endometrium can initiate the occurrence of endometriosis. In this study, for the first time, we produced CAGE (Cap Analysis of Gene Expression) data from stromal cells of endometriosis patients and controls to examine the potential effects of the reported loci on regulatory genomic regions, including promoters and enhancers. In addition to CAGE data, we are analysing RNA-seq, ATAC-seq which together allow us to investigate the potential mechanisms for the reported loci at the genome-wide level.

Genetic predisposition for the deviation of offspring sex ratio in two varieties of Iberian pig.

Miss Melani Martín De Hijas Villalba

Biography:

Melani Martin de Hijas Villalba, a Ph.D. candidate at the Universitat Autònoma de Barcelona (Spain) under the direction of Dr. Joaquim Casellas. Specializing in quantitative genetics and animal breeding, with research topic mainly focused on pig genetics. More specifically, working on an omic analysis of reproductive traits in a diallelic cross between two Iberian pig populations.

Fisher's principle of sex ratio evolution explains why, in most species with sexual reproduction, sex ratio fluctuates around 1:1 equilibrium. Nevertheless, deviations in the sex of the offspring, both male- and female-biased, have been found in several animal populations. This study aims to find the genetic basis for the variability of sex ratio in two different varieties of Iberian pig, Entrepelado and Retinto. Genomic, pedigree and phenotypic data from 423 Entrepelado and 397 Retinto sows were analyzed. For each sow, a new phenotype was created as the across-litter weighted variance of sex ratio departure among piglets born alive. This phenotype was analyzed by an animal mixed linear model under Bayesian inference and accounting for population mean, herd-year-season and additive genetic effects. Heritabilities suggested different genetic backgrounds between Iberian pig varieties, this being larger in Retinto (0.028 (0.002-0.095)) than Entrepelado (0.048 (0.006-0.123)). Breeding values were analyzed under a lineal GWAs approach accounting for family structure, and several chromosomes contained SNP with significant estimates. Again, differences between pig varieties were revealed. Only a few SNP reached statistical significance in Retinto sows. On the other hand, the number of significant SNP was substantially bigger in Entrepelado sows, and they scattered across the whole genome. Despite differences between varieties, they both revealed genetic background for the trait and shared a significant genomic region in chromosome 6, with outstanding potential implications for the Iberian pig industry.

Single-cell analysis of the genetic effect on the differentiation of cardiomyocytes

Ms Maika Matsumoto

Biography:

Maika Matsumoto is a first-year PhD student at Institute for Molecular Biosciences, University of Queensland. She obtained a master's degree in Bioinformatics from the same university as she is currently studying at. In her thesis, she investigates the genetic effect on the cell development process of stem and cancer cells in in-vitro and in-vivo conditions by developing a bioinformatics analysis pipeline for single-cell and spatial transcriptomics.

The majority of genetic loci underlying common disease risk act through changing genome regulation activities, most commonly linked to expression quantitative trait loci, where gene expression is measured using bulk populations of mature cells. A crucial missing step is evidence of variation in the expression of these genes as cells progress from a pluripotent to the differentiated state. This is especially important for cardiovascular disease (CVD), as the majority of cardiac cells have limited properties for postneonatal renewal.

To investigate the dynamic changes in gene expression across the cardiac lineage, we differentiated 5 iPSC cell lines derived from healthy donors with different genetic backgrounds. Extensive single-cell transcriptomic analysis was performed on >120,000 cells isolated at four in-vitro differentiation time points, which correspond to stage-specific transitions in cardiac differentiation.

By canonical correlation analysis and graph-based clustering, we removed the batch effects and identified the distinct cell subpopulations comprised of 14 cell types: 1 (Day 0), 4 (Day 2), 5 (Day 5) and 4 (Day 15). These cell types were aligned into cell developmental lineages, which were reconstructed by trajectory analysis based on gene expression.

Furthermore, we identified the gene modules with > 400 genes that were correlated with the differentiation time. Four out of these genes were associated with CVD and matched with the sample genotypes. The expression pattern of the CVD-associated genes during differentiation differed between cell lines.

This study improves our understanding of the genetic effects underlying cardiac development and provides insights into the new treatment strategy for cardiovascular disease.

Overfitting in polygenic risk score analyses: Exploring the impact of sample overlap and first degree relatives

Professor Sarah Medland

Biography:

Professor Sarah Medland is a Psychiatric and Statistical Geneticist working on Neuro-Imaging genetics and Mental Health.

Independence of the discovery and target samples is essential to avoid estimation bias due to over-fitting within polygenic risk score analyses. Although this is a well-known methodological issue, in recent years perceptions have arisen that this may not matter with biobank type discovery samples when the overlap is very small and few authors have considered the impact of relatedness across the discovery and target samples. To empirically examine these issues a discovery sample of ~340,000 individuals were extracted from the UK Biobank (app. 25331) and GWAS were conducted for a continuous (height) and a binary trait (day-time sleepiness) and for random continuous and binary traits. Polygenic risk scores were calculated and polygenic risk score analyses were conducted using target samples comprising randomly selected baseline samples of 2,000, 5,000 and 10,000 individuals who were not in the discovery sample. To examine over-fitting due to non-independence, the target samples were spiked with 5, 10, 50, 100 and 200 individuals who had been part of the discovery sample or their first degree relatives. Sample size was maintained between baseline and spiked samples and 1,000 replicates were performed. As expected degree of over-fitting increased with degree of overlap, decreased with total target size and also arose due to non-independence due to relatives. Over-fitting was present with as few as 5 overlapping individuals (0.001% of the initial discovery sample) confirming the importance of independence for accurate estimates of variance explained. The impact of non-independent controls (in the absence of overlapping cases) will also be presented.

WOMBAT: A tool for estimation of genetic parameters -- highlights and updates

Professor Karin Meyer

Biography:

Researcher in animal breeding

WOMBAT is a freely available software package for linear mixed model analyses in quantitative genetics, distributed since 2006. Its main focus is the estimation of covariance components and genetic parameters via restricted maximum likelihood (REML) and the prediction of individuals' genetic merit. It predominantly targets models and tasks common to animal breeding applications, but is well suited to and used in other areas of quantitative genetics.

We highlight its key characteristics and capabilities, some unique features and selected, recent extensions to better accommodate analyses utilising genomic information. Specific topics considered are parsimonious models for multivariate analyses, REML estimation subject to a penalty on the likelihood to reduce sampling variation, pooling of results from analyses by parts, construction of relationship matrices and their inverses, genome wide association screens and so-called single step analyses. In addition we outline computational strategies employed and the scope for large scale analyses with modern hardware.

WOMBAT is distributed as executable programs for either Linux or Windows operating systems. These are accompanied by a detailed user manual and extensive suite of worked examples. All material is available for download from <http://didgeridoo.une.edu.au/km/wombat.php>

Across-breed genomic prediction in Australian sheep breeds using selected genetic markers from whole genome sequence data

Dr Nasir Moghaddar¹, Dr Iona M MacLeod², Dr Sunduimijid Bolormaa², Prof Andrew A Swan³, Prof Daniel J Brown³, Prof Julius HJ van der Werf¹

¹University of New England, ²Agriculture Victoria Research, AgriBio Centre, ³Animal Genetics and Breeding Unit (AGBU)

Across-breed genomic prediction using moderate density single nucleotide polymorphism (SNP) markers has been associated with low accuracy, more particularly for genetically distant breeds. Whole genome sequence (WGS) data provides information on genetic markers at or in high linkage disequilibrium with causative markers underlying the genetic variation of a polygenic trait. In this study, the accuracy of across-breed genomic prediction was compared using either 50k SNP genotypes or 50k SNP genotypes plus 2,590 selected markers from WGS data. Genomic prediction was based on genomic best linear unbiased prediction (GBLUP) and considering 350 independent purebred Border Leicester (BL) animals as test set. Reference sets for genomic prediction for different traits were: a) 1,764 to 1,930 purebred BL, b) 2,382 to 3,635 purebred Merinos or c) 13,074 to 15,864 mixed purebred and crossbred animals other than BL breed. The studied traits were weight, scanned fat and scanned eye muscle depth measured at post weaning age. Selection of genetic markers was based on results from a multi-trait genome-wide association study using WGS, performed on a large multi-breed sheep data set. Accuracy of within breed genomic prediction from a small available BL reference set (1,553 animals) was on average 0.28 which improved to 0.32 by using selected WGS markers combined with 50k SNP genotypes. Across-breed prediction accuracy was on average 0.10 which increased to 0.19 by integrating selected WGS markers with 50k SNP genotypes. Bias of genomic breeding values tended to be lower based on prediction from 50k SNP genotypes integrated with selected WGS markers.

Effect of preselection of whole-genome sequence variants for kinship estimation

Mr Eduard Molinero

Biography:

During my Biotechnology studies in Universitat de Lleida, Spain (UdL), I performed an internship on the production of transgenic rice plants to improve its β -carotene content. That internship motivated me into searching for alternative ways to obtain comparable results that are in line with the current European Union legislation. Therefore, I went to Wageningen University and Research, Netherlands (WUR), to increase my knowledge on molecular breeding. Thereafter, I joined the Animal breeding group in UdL to develop my PhD studies working on whole-genome sequencing for the discovery of genetic variants associated with meat quality and fat metabolism in pig.

Population structure can be accounted for in genetic evaluations and genome-wide association studies by fitting a genomic kinship matrix in mixed models. Kinship matrices are typically calculated from the genotypes available from marker arrays. As whole-genome sequence data becomes more affordable, the number of available variants increases from tens or hundreds of thousands to millions, albeit genotype uncertainty can be higher at low and moderate sequencing coverage. Computational requirements for calculating the kinship matrix do not scale well to such volume of data and therefore there is a need to preselect variants. We used data from 146 pigs sequenced at 8.0x. The genotype uncertainty that arises from sequencing produced a downwards bias in kinship estimations and therefore we limited our study to the subset of 250k variants with the greatest genotype certainty. Subsets of variants were preselected based on linkage disequilibrium (to retain 10k, 25k, 50k or 100k variants) and minor allele frequency (MAF>0.00, 0.05, or 0.25). The kinship matrix with each variant set was calculated using standardized genotypes and compared to a pedigree-based matrix. Including variants with lower MAF or using fewer variants resulted in kinship estimates that were biased downwards. As a consequence, the correlation between variant- and pedigree-based kinship estimations of relatives ranged from 0.82 (10k, MAF>0.00) to 0.86 (100k, MAF>0.25). The latter was greater than the correlation when using all variants for estimating kinship (0.85). As long as genotypes with high certainty are used, kinship estimation seems quite robust to the preselection of whole-genome sequence variants.

Identifying associations between tissue and cell specific genetic regulatory mechanisms and risk of female reproductive disorders.

Dr Sally Mortlock

Biography:

Dr Sally Mortlock is a Postdoctoral Researcher and lead genetic statistician within the Genomics of Reproduction Disorders group at the Institute for Molecular Bioscience, UQ. She was awarded her PhD in 2017 at the University of Sydney having completed a project in Canine Cancer Genomics. She has been heavily involved in managing and analysing the largest omic datasets for endometrium including genomic, transcriptomic and methylation datasets for both Australian and international endometriosis cohorts. Her studies have expanded knowledge of the genetic and epigenetic regulation in endometrium generating important data resources to identify gene targets regulating female reproductive traits and diseases.

Reproductive conditions account for the largest female health expenditure in Australia. One disease endometriosis, affects 10% of reproductive aged women and costs \$7.4 billion annually. GWAS have identified 27 endometriosis risk loci, although target genes and genetic mechanism's behind the disease remain to be identified. This project aims to integrate omic data from endometrium to better understand mechanisms underlying gene regulation in endometrium and investigate associations with reproductive traits and diseases. We analysed RNA-sequencing and methylation data from endometrial samples with genotype data to map quantitative trait loci for expression (eQTL) and methylation (mQTL) loci. We observed 444 sentinel cis-eQTLs and 4546 sentinel cis-mQTLs in endometrial tissue. Summary-data-based Mendelian Randomisation analyses identified putative functional genes associated with reproductive traits and diseases including endometriosis. Genetic effects on gene expression in endometrium were highly correlated with reproductive tissues, supporting evidence that genetic regulation of gene expression is shared between biologically similar tissues and cell types. Genetic regulation of gene expression specific to different endometrial cell types may be important in disease pathogenesis and we will use advanced statistical frameworks to estimate cell proportions in endometrial samples and consecutive deconvolution of cell type eQTLs. We are also integrating genetic effects with data on somatic mutations reported recently in the epithelial cells of endometriosis lesions and eutopic endometrium. Genetic effects on transcription and methylation in endometrium and its constituting cell types provides a platform to better understand of genetic effects on endometrial-related pathologies.

Determining relatedness between South African and Australian Merino populations according to a genomic relationship matrix and its principal components

Mr Cornelius Nel

Biography:

Cornelius is a PhD candidate in the department of animal science at Stellenbosch University, South Africa. His research revolves around implementing genomic methods to enhance the prediction of South African wool sheep, particularly Merinos. Particular interests are the genetic improvement of fitness traits and improving environmental sensitivity of animals reared in extensive and varying conditions.

Genomic prediction of breeding value has been implemented in the genetic evaluation of Australian Merinos but not for South African (SA) Merinos. Combining genomic resources between countries could benefit genetic evaluations in SA, depending on the level of relatedness between reference populations in the two countries. The aim of this study was to use genomic information to determine the relatedness of Merino populations between countries. SA Merinos were grouped into research flocks: Elsenburg (400), Grootfontein (115), Cradock (127) and commercial 'Industry' samples (41). The Australian Merino samples were grouped by strain: Ultrafine (270), Fine Medium 1 (224), Fine Medium 2 (133) and Strong (291). The only recent across-country links are in the Cradock flock where 4 Australian sires were used in 1988, 2 in 1996 and another 2 Ultrafine rams in 2002-2003. Internal relationships for South African flocks were low, but positive, whereas the Australian lines were negatively related. The highest level of relatedness was observed between the Cradock flock and the Ultrafine strain in Australia. A visual analysis of the first two principal components showed discernable clusters by population group, with more distant separations for the Fine Medium 1 and Strong lines. The Cradock flock plotted close to the Fine Medium 2 and Ultrafine bloodlines. The separation of Merinos by country and subpopulation contributed to knowledge of within-breed genetic variation. It was clear from the Cradock – Ultrafine comparison that relatively few genetic links could be effective in increasing overall relatedness.

Spatial understandings of disease in single cells and intact tissue sections through machine-learning integration of imaging and sequencing data

Dr Quan Nguyen¹, Mr Duy Pham¹, Mr Xiao Tan¹, Mr Andrew Su¹, Mr Minh Tran¹, Dr Pui Yeng Lam¹

¹*Institute For Molecular Bioscience*

Biography:

Dr Quan Nguyen is a Group Leader at the Institute for Molecular Bioscience, The University of Queensland. He is leading the Genomics and Machine Learning (GML) Lab. GML aims to study cancer tissue at cellular resolution and in physiological context. The group generates single-cell, spatial transcriptomics, and imaging data and develops computational-statistical methods to analyse these multimodal datatypes. By systematically identifying cell types and characterizing their spatial organisation and cell-cell interactions, the group aims to find gene-regulation mechanisms underlying cell-type specific changes in response to treatment or gene perturbation. In the past three years, Dr Nguyen has published in top tier journals and open-source software and web applications.

Single-cell sequencing technologies have proven to be the method of choice to study cell-type-specific biological processes, but spatial information is missing from single-cell sequencing data due to the dissociation of cells from their original tissues. Spatial transcriptome (ST) technologies are emerging as the next-generation technologies of single-cell RNA sequencing. For the same tissue section, ST adds a spatial dimension to high-resolution and transcriptome-wide gene expression data and also adds tissue-morphology imaging data. We hypothesise that integrating expression data with tissue imaging data would provide a comprehensive understanding of complex biological processes happening in situ between cell types within physiological tissue.

We developed machine-learning and deep learning methods, implemented in a spatial analysis software suite that include three programs: HEMnet, SpaCell and stLearn. HEMnet implements imaging analysis methods to transfer molecular labelling from immunostaining data to histological images. SpaCell uses spatial transcriptomics molecular labels, with thousands of times more genes, to predict multiple cell types in a tumour tissue. SpaCell is a deep learning software program that allows us to integrate millions of pixel intensity values with thousands of gene expression measurements from spatially-barcoded spots in a tissue. stLearn implements novel methods to incorporate spatial distance into the analysis to identify cell types, microenvironments and cell transitions within the physiological tissue.

Here, we will demonstrate the vast potential of integrating spatial and imaging data into single-cell data for discovering complex biological processes in cancer and neuronal diseases.

Genetic enhancement of striped catfish (*Pangasianodon hypophthalmus*) using quantitative genetic and genomic approaches

Mr Vu Thanh Nguyen

Biography:

Vu Thanh Nguyen is pursuing a Ph.D degree at the University of the Sunshine Coast from 2018 to understand genetic and genomic basis of striped catfish resistance to Edwardsiella ictaluri bacterial pathogen which caused the most morbidity and mortality in pangasius farming in Vietnam. He has been working in aquaculture in Vietnam for over 15 years, with profound contributions on genetic improvement breeding programs, i.e. tilapia, giant freshwater prawn and striped catfish. His interests are to utilize genetic and/or genomic information to improve disease resistance, growth performance as well as traits of commercial importance in aquatic species throughout selective breeding.

Genetic improvement has been conducted for 20 years (from 2001) to enhance productivity and disease resistance of striped catfish (*Pangasianodon hypophthalmus*) - a freshwater fish species that contributes about 1% GDP of Vietnam (valuing at US\$2.5 billion in 2019). The selection program enhanced growth performance by 9.3% per generation and achieved correlated increases in carcass traits (fillet weight) and survival by 8.5% and 7.4%, respectively. There are still remarkable genetic variations in these traits to secure long-term response to future selection in this population. The breeding objectives for *P. hypophthalmus* have been broadened by including disease resistance to *Edwardsiella ictaluri*, a devastating disease in catfish farming. Our quantitative genetic analysis showed that there are heritable genetic components in the resistance ($h^2 = 0.10 - 0.35$) and the genetic correlations of *E. ictaluri* resistance with body weight and survival were all positive (0.03 – 0.53). These results suggest that both growth and disease resistance can be improved simultaneously in this population. Despite the significant gains achieved for these traits, the breeding program for striped catfish has been devised by including new genomic information. Our population genomic analysis showed there is significant genetic diversity in the selected population and those in the lower Mekong river. Additionally, preliminary genomic prediction analysis suggests possibilities for using genomic selection to improve disease resistant, tolerant or resilient traits. With the applications of quantitative genetic and genomic approaches, we have successfully developed new genetic lines of striped catfish to sustain future development of the sector world-wide.

Comparison of risk prediction methods

Dr Guiyan Ni

Biography:

Guiyan Ni is a PostDoc working for Prof. Naomi Wray and Prof. Peter Visscher at the University of Queensland. She is primarily interested in statistical and computational methods for understanding genetic risk factors for common diseases with a particular focus on psychiatry disorders.

Aims: Polygenic risk scores (PRS) can potentially predict the risk of developing complex diseases. There are many methods to select and weight SNPs and the optimum is likely trait dependent reflecting different genetic architectures.

Methods: We compare prediction accuracy of P+T, LDpred, LDpred-funct, SBLUP and SBayesR when applied to two psychiatric disorders (schizophrenia and major depressive disorder) and two other diseases (Alzheimer's disease, type II diabetes). All these methods use GWAS summary data to build PRS in an independent sample. We first use P+T where independent genetic variants are selected under varied p-value thresholds and the linkage disequilibrium clumping thresholds are arbitrary. We then apply SBLUP which is a summary-based BLUP method and two methods in Bayesian framework (LDpred and SBayesR). LDpred-funct can use functional annotation information.

Results: We estimate the mean prediction accuracy using Nagelkerke's R^2 across 26 MDD cohorts and the standard error by leave one cohort out. LDpred, LDpred-funct and SBLUP are concordant with P+T, which are 0.039, 0.038, and 0.038 versus 0.034 (SE 0.018). SBayesR provides the best estimate of 0.041 (SE 0.020), improving the accuracy by 21% compared to P+T.

Conclusions: For MDD, although the prediction accuracies of different methods did not differ significantly, SBayesR offered the highest prediction accuracy. With a better tuning of the prior in SBayesR, the accuracy could be increased further.

A PsycheMERGE Phenome-wide Association study of genetic risk for Attention Deficit Hyperactivity Disorder and Autism

Dr Maria Niarchou

Biography:

Dr. Niarchou is a Research Instructor at Vanderbilt University Medical Center. Dr. Niarchou received a B.Sc in Psychology from the National and Kapodistrian University of Athens (Greece), an M.Sc. in Neuropsychology from the University of Bristol (UK), and a Ph.D in Psychology from Cardiff University (UK). Her research focus during her Ph.D and post-doctoral work has been on examining the role of Copy Number Variants in the development of psychiatric problems in children and adolescents. During the past two years, Dr. Niarchou's research interests have expanded to also examine how common genetic variation can explain risk for psychiatric disorders.

Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorders (ASD) are associated with risk for many behavioral and physical problems including substance use disorder, learning disability, and anxiety disorders. We generated polygenic risk scores (ADHD-PRS and ASD-PRS) trained on summary statistics from the latest Genome-Wide Association Studies of ADHD, and ASD for 48659 individuals in the VUMC Biobank (BioVU). Clinical diagnoses from the VUMC electronic health record (EHR) are available on all BioVU participants. The ADHD-PRS and ASD-PRS were significantly associated with the clinical diagnosis of ADHD (Odds Ratio (OR))=1.22, $p=3.6 \times 10^{-9}$) and ASD (OR=1.30, $p=2.1 \times 10^{-7}$), respectively. Both PRSs were associated with many other phenotypes. ADHD-PRS was most strongly associated with tobacco use disorder (OR=1.25, $p=2.99 \times 10^{-49}$), Type 2 diabetes (OR=1.11, $p = 7.6 \times 10^{-16}$), posttraumatic stress disorder (OR=1.3, $p = 3.9 \times 10^{-15}$), and obesity (OR=1.17, $p = 1.03 \times 10^{-13}$), while ASD-PRS was most strongly associated with mood disorders (OR=1.08, $p=7.0 \times 10^{-10}$), and breast cancer (OR=1.12, $p=1.9 \times 10^{-06}$). Finally, we examined how the PRSs are related to the medical phenome during different developmental periods. As expected, from the ages of 0 to 11, ADHD and ASD were the strongest associations. Among 12 – 25 year-olds we began to identify associations between ADHD-PRSs and substance addiction, pregnancy complications, posttraumatic stress disorder, and in later ages we found associations with chronic health conditions such as type 2 diabetes. At the time of the conference, we will present results from analyses across the PsycheMERGE network.

Sustainable production: genomic predictions for selecting locally adapted beef cattle

Ms Sara Nilson

Biography:

Sara Nilson is currently a Ph.D. student in the Division of Animal Science at the University of Missouri working towards a degree in genetics with a specialization in bioinformatics. Thus far, research projects include identifying local adaptation in USA beef cattle breeds and developing eco-region specific genomic predictions, and population genomics of worldwide random bred cats. Previous work includes transcriptome analysis of persistently infected beef cattle with BVDV while at the University of Nebraska for her MS. Her interests are in ecology, evolution, and disease genomics with an appreciation for one health approaches and a passion for conservation.

To feed ~10 billion people by 2050, beef production needs to increase. To sustainably increase production, environment-specific genomic predictions were formulated to improve selection of environmentally adapted beef cattle. To classify environments, K-means clustering was performed with 30-year normals of precipitation, temperature, and elevation resulting in 9 environmental regions for the United States. This multivariate approach additionally captures local effects including but not limited to pathogens, forages, and humidity. Simmental, Red Angus, and Gelbvieh breeds were utilized with over 17,000, 15,000, and 12,000, individuals respectively. Phenotypes of birth weight, weaning weight, and yearling weight were chosen due to the large number of available records. Genotypes were phased and imputed up to ~850,000 SNPs and SNPs with a minor allele frequency >0.01 were retained. In BLUPF90, bivariate models were analyzed for each region combination, phenotype, and breed, e.g. Region 3 and Region 8 for birth weight for Simmental. Regions with less than 1,000 animals were not included. Multivariate models of all regions were analyzed for each phenotype and breed. Genetic correlations indicate substantial GxE effects among regions, reflected by different allele substitution effects between regions. Environment-specific heritabilities were calculated. Breeding values were estimated (EBV) for every animal for each region and phenotype. Accuracy was defined as the correlation between EBVs of the full dataset and EBVs when the youngest 25% of animals were excluded. Animals significantly reranked between regions. These predictions will serve as valuable tools to identify exceptional animals that will perform in their environment before breeding or purchasing an animal.

High-definition likelihood inference of genetic correlations across human complex traits

Mr Zheng Ning

Biography:

Zheng Ning did his Bachelor's degree in mathematics from Zhejiang University and his Master's degree in statistics from Uppsala University. In 2015, Zheng Ning became a Ph.D. student at the Department of Epidemiology and Biostatistics in Karolinska Institutet. In September 2020, Zheng successfully defended his thesis entitled "Novel statistical methods for genome-wide association summary statistics".

Genetic correlation is a central parameter for understanding shared genetic architecture between complex traits. By using summary statistics from genome-wide association studies (GWAS), linkage disequilibrium score regression (LDSC) was developed for unbiased estimation of genetic correlations. Although easy to use, LDSC only partially utilizes LD information. By fully accounting for LD across the genome, we develop a high-definition likelihood (HDL) method to improve precision in genetic correlation estimation. Compared to LDSC, HDL reduces the variance of genetic correlation estimates by about 60%, equivalent to a 2.5-fold increase in sample size. We apply HDL and LDSC to estimate 435 genetic correlations among 30 behavioral and disease-related phenotypes measured in the UK Biobank (UKBB). In addition to 154 significant genetic correlations observed for both methods, HDL identified another 57 significant genetic correlations, compared to only another 2 significant genetic correlations identified by LDSC. HDL brings more power to genomic analyses and better reveals the underlying connections across human complex traits.

Genetic Architecture and Gene Mapping of Cyanogenic Glucoside in Cassava

Mr Alex Chukwuka Ogbonna

Biography:

Alex Ogbonna is a PhD student at Cornell University working in the lab of Dr. Lukas Mueller. Upon completing a BS in Chemistry and MS in Analytical Chemistry, he worked with a cassava molecular breeding team at National Root Crops Research Institute Nigeria and later moved to Boyce Thompson Institute as a web database developer. There he developed database tools to assist breeders of cassava, banana, yam, sweet potato, maize, and solanaceous crops. His current research is focused on the impact of modern breeding on key traits architecture evolution through domestication and its implication for cassava breeding approach.

Cassava is a root crop originating from South America and a major staple crop in the Tropics, including Africa. In this study, we investigated the genetic architecture of Hydrogen Cyanide (HCN), a major component of tuber quality in both continents. HCN is involved in cassava plant defense against herbivores but a toxic compound upon tissue disruption. We genotyped 3,354 landraces and modern breeding lines originating from 26 different Brazilian states and phenotypically, 1,389 individuals were characterized across multi-year trials. All plant material was subjected to high density genotyping using Genotyping-By-Sequencing (GBS) and 27,000 Single Nucleotide Polymorphic (SNP) markers were selected. We performed the first association mapping (GWAS) to characterize the genetic architecture and gene mapping of HCN. Field experiment revealed phenotypic heritability of 0.74 (with SNP based heritability, 0.41) for HCN. Association studies revealed two major loci contributing 7 and 30% of the marker based estimated variance and indicated the presence of genes encoding for an ATPase and MATE protein respectively. We developed and validated sets of diagnostic markers for breeding applications and investigated evidence for domestication in HCN. Our findings were further validated in an African population and provides future resources for genetic studies of cyanide in cassava.

Leveraging polygenic risk scores to target genetic modifiers in families with clinically heterogeneous epilepsy

Ms Karen Oliver

Biography:

Karen Oliver is a PhD candidate from the Walter and Eliza Hall Institute of Medical Research and Research Fellow at the University of Melbourne. Karen received her Master of Science in Bioinformatics with first class honours in 2015 and has worked in the epilepsy clinical genetics field for over 13 years. She is a member of the International League Against Epilepsy Consortium on Complex Epilepsies and has co-authored 36 peer-reviewed papers. Her current major interest, and a large focus of her PhD research, is the use of polygenic risk scores in the epilepsies.

The developmental and epileptic encephalopathies (DEEs) are a heterogeneous group of rare disorders, characterised by early-onset seizures, developmental delay and regression. Whilst largely conceptualised as sporadic, de novo dominant disorders, we not infrequently observe DEE cases with a family history of mild epilepsy. We hypothesise that epilepsy risk, due to common genetic variation, is enriched in familial versus non-familial epilepsy cases and modifies the effect of rare monogenic variants segregating in phenotypically heterogeneous families.

To capture common epilepsy risk in patients diagnosed with DEE (n=90 unrelated), we calculated polygenic risk scores (PRS) using 11 genome-wide significant SNPs from the largest epilepsy genome-wide association study to date. The DEE cohort was divided into those with an affected first degree relative (n=50; “familial”) versus those without (n=40; “non-familial”). Using linear regression, we compared scores for the two groups (with sex and the first three ancestry principle components as covariates). We found significant PRS enrichment for familial versus non-familial DEE cases (p -value = 0.02).

One familial DEE case was known to have a maternally-inherited, pathogenic variant in CHD2. The proband’s mother also had epilepsy, but her phenotype was milder. We calculated the mother’s common epilepsy risk to be much lower than her more severely affected child’s (standardised PRS; 0.89 versus 3.08 respectively).

Our preliminary data, though on small numbers, are consistent with our hypotheses being supported. Subsequent analyses, have the potential to help clarify the interplay between rare and common variants and provide a future pathway for targeting important phenotypic modifiers.

Identification of genetic variants linking dairy fertility and milk production traits

Dr Ee Cheng Ooi^{1,2}, Professor Michael E Goddard^{1,2}, Professor Jennie E Pryce^{1,3}

¹Agriculture Victoria, ²University of Melbourne, ³La Trobe University

Biography:

Dr Ee Cheng Ooi is a dairy veterinarian with a strong focus on improving herd fertility. She has worked in a variety of roles within the Australian dairy industry, including as a clinician and herd consultant, extension officer and science communicator, technical lead and now as a dairy scientist. She is currently completing her PhD looking at all aspects of improving fertility through genetic selection, including farmer attitudes towards the Australian fertility estimated breeding value (EBV), validation of this EBV using herd records, and now improvement of the EBV using genomic approaches.

Fertility in dairy cattle has declined over the last fifty years as an unintended consequence of selection for high milk yield. Lactation is obviously contingent on parturition – this makes fertility a key driver of profitability, particularly on pasture-based dairy farms. The ideal cow does not only conceive – she does it at the right time, on the first attempt, and achieves and maintains pregnancy despite producing more than 60 litres of milk per day.

The exact physiological mechanisms linking fertility and milk production are still uncertain, despite significant research efforts. Results from observational studies and in vivo experimentation have been equivocal – largely because nutrition, health, management interventions and environmental factors all combine to confound analysis of herd reproductive performance.

Advances in genomics allow a direct approach to testing hypotheses. However, from a genetic perspective, fertility is a complex trait composed of successive biological events, with phenotypes that are difficult to measure. In this study, the use of a genome-wide association study incorporating large multi-breed reference population and a subset of variants which have been pre-selected for significance gives us significantly more power to identify variants of interest. It also allows us to identify variant clusters that have similar effects on multiple traits possibly indicating a common physiological pathway.

This study aims to uncover the physiological mechanisms underlying milk production and fertility, which may assist herd managers in uncoupling these traits to breed cattle that are both productive and highly fertile.

Identification of a functional sequence variant in the MYH3 locus affecting muscle fiber-type composition and intramuscular fat content in pigs

Dr Hee-bok Park

Biography:

Research Interests: Genetics and genomics, Bioinformatics, Animal Breeding

Muscle fiber composition and intramuscular fat (IMF) contents are complex quantitative traits of critical economic importance in pork quality. However, the genetic factors underlying the two traits remain to be elucidated. To identify functional sequence variants (FSVs) that affects the two meat quality traits in the loin muscle in pigs, we conducted genome-wide association study (GWAS) and high-resolution mapping analysis using two independent intercrosses between two Western breeds (Landrace and Duroc) and Korean native pigs (KNPs). Based on a joint linkage and association analysis followed by GWAS using the Porcine60Kbeadchip, we detected a 488.1-kb region on porcine chromosome 12 that influences both reddish meat color (a^*) and IMF content. In this critical region, only the MYH3, encoding myosin heavy chain 3, was revealed to be preferentially overexpressed in the skeletal muscle in KNP compared to Landrace. Additionally, we generated MYH3-transgenic mice, demonstrating that this gene regulates both myofiber-type specification and adipogenesis in skeletal muscle. Subsequently, we detected a structural variant (6-bp deletion) in the promotor/regulatory region of MYH3 for which Q allele carriers exhibited significantly higher values of IMF contents and a^* than q allele carriers. Furthermore, we performed chromatin immunoprecipitation and co-transfection assays to show that the 6-bp deletion FSV in the 5'-UTR of MYH3 for which it abrogated binding of the myogenic regulatory factors. In conclusion, we discovered a FSV in the porcine MYH3 that can provide novel insights into the genetic basis of the myogenesis and lipidogenesis in mammalian skeletal muscle and improve meat quality in pigs.

Increasing SNP calling accuracy, coverage and read-depth in sequence data by the use of haplotype blocks

Mr Torsten Pook¹, Dr Adnane Nemri², Dr Eric Gerardo Gonzalez Segovia³, Prof Dr Henner Simianer¹, Prof Dr Chris-Carolin Schoen³

¹University of Goettingen, ²KWS SAAT SE, ³Technical University of Munich

Biography:

Torsten Pook, 28, is a PostDoc in the Animal Breeding and Genetics Group at the University of Goettingen. He studied Business Mathematics at the University of Mannheim and wrote his master thesis on "Forecasts of daily shipments in logistics" in cooperation with DB Schenker. He is working in quantitative genetics since his PhD on the topic of "Methods and software to enhance statistical analysis in large scale problems in breeding and quantitative genetics". His research interests are the simulation of breeding programs (software: MoBPS), genomic prediction, improving of imputation and the creating of haplotype blocks and libraries (software: HaploBlocker).

Even though the price of sequencing has gone down substantially in recent years, the generation of sequence data with high read-depth is still costly. Here, we propose a new strategy to merge reads of individuals, and their respective haplotypes, that are in identity-by-descent (IBD) and thereby generate data with high read-depth by collecting much cheaper low read-depth data.

For this, we execute a pipeline of first generating a haplotype library in the software HaploBlocker, as the haplotype blocks obtained here are group-wise IBD. The tool can be applied efficiently to large-scale datasets. Allelic variants for each marker are then called by using the reads of all haplotypes that stem from the same haplotype block. To still be able to call recent mutations, a higher weight is given to the reads that stem from the haplotype itself.

We evaluated our method on a dataset containing 340 doubled haploid lines of the European maize landrace Petkuser Ferdinand Rot that was sequenced with 0.5X read-depth. Our pipeline lead to an increase of the average read-depth to 66.6X (median: 43X) with 91.9% of all variants being called. Further, 99.8% of all called variants are in concordance with the genotype calls from the Affymetrix Axiom Maize Genotyping Array. In contrast, only 39.4% of all variants were called in the original data with 0.5X read-depth and concordance rates to the SNP-array were lower (98.9%). These results underpin the power of our new strategy to increase SNP calling accuracy, coverage and read-depth at the same time.

A Two-Part Strategy for using Genomic Selection in Hybrid Crop Breeding Programs

Mr Owen Powell

Biography:

My research interests centre on using quantitative genetics to drive genetic gain and efficiency in plant and animal breeding programmes. My previous work has focused on using simulations to demonstrate the potential of genomic prediction to exploit synergies between plant and animal breeding.

Stochastic simulations were used to quantify the impact of new genomic breeding strategies in a wide variety of settings; from low to middle-income (LMIC) dairy cattle breeding programs to large, well-funded maize breeding programs. My current position focuses on the development of genomic prediction methods that combine biological, environmental and management information under a unifying framework.

Overall food production will need to increase by between 70-100% in the next 30 years. Enhancing the efficiency of breeding programs is one route for achieving this. This study performed stochastic simulations to quantify the potential of a two-part strategy in combination with two crossing schemes to increase the rate of genetic gain in hybrid crop breeding programs. The two crossing schemes were: (i) a circular crossing scheme, and (ii) a maximum avoidance of inbreeding crossing scheme. The results from this study show that the implementation of genomic selection increased the rate of genetic gain and that the two-part hybrid crop breeding program generated the most genetic gain. This study also shows that the maximum avoidance of inbreeding crossing scheme increased long-term genetic gain in two-part hybrid crop breeding programs completing multiple selection cycles per year, as a result of maintaining higher levels of genetic variance over time. The flexibility of the two-part strategy offers further opportunities to integrate new technologies to further increase genetic gain in hybrid crop breeding programs, such as the use of outbred training populations. However, the practical implementation of the two-part strategy will require the development of bespoke transition strategies to fundamentally change the data, logistics, and infrastructure that underpin hybrid crop breeding programs.

Selection for A2 β -casein alleles in Australian dairy populations

Professor Jennie Pryce

Biography:

Professor Jennie Pryce is a geneticist well known for her work in the area of dairy genetics and genomics. Jennie is principal research scientist for Agriculture Victoria and Professor of animal genetics at La Trobe University. Her main interests are genetic improvement of functional traits in dairy cows (especially female fertility and feed conversion efficiency), optimization of breeding scheme design under genomic selection, development of dairy selection, and improvement of accuracy of existing and new breeding values.

Milk products from cows that are homozygous for the A2 β -casein allele is marketed in several countries by the A2 Milk Company. Breeds such as Guernsey have been promoted as being higher in the A2 allele than other breeds. However, the alleles present at the β -casein locus of genotyped sires of all breeds are published by DataGene and available to farmers when making selection decisions alongside estimated breeding values and indices. In this study we assessed the frequency of A1 and A2 alleles using low density genotypes (mainly <10k) of more than 30,000 cows born in Australia imputed to full sequence using a reference population of >2700 sequenced key ancestor *Bos taurus* sires from the 1000 Bull Genomes Project. The A2 mutation was identified and validated in imputed sequence data using the β -casein genotypes of known sires. The cows used in this study were predominantly from the national Genomic Information Nucleus (Ginfo), where entire herds are genotyped and selected on the completeness of their data recording. Our results show that over the last 25 years there has been intense selection for the A2 allele, especially in the Holstein breed. The frequency of A2 allele in Jersey animals was higher than Holstein in the late 1990s, when ~50% of Jersey cows were A2/A2, while only ~20% of Holstein cows were A2/A2. For Holstein bulls and cows born in 2010-2012, ~45% and ~60% were A2/A2, respectively. To conclude, interest in A2 milk has resulted in a trend to select for A2/A2 bulls.

Impact of artificial selection on haplotype diversity in hexaploid wheat

Mr Jesus Quiroz-Chavez

Biography:

BSc: Biological Sciences

MSc: Agricultural Biotechnology

PhD: Crop Sciences (Wheat Genetics and Genomics)

*I finished my BSc degree in Biological Sciences with a major in plant biology. I continued working in plant sciences with *Arabidopsis thaliana* and abiotic stress (NaCl and drought) tolerance during my MSc degree. After my master, I worked in the industry in a maize breeding program, *Fusarium ssp.* resistance, and Doubled Haploids. Now I am a 3rd year Ph.D. student at John Innes Centre (JIC) working in wheat genetics, genomics, and bioinformatics for yield traits.*

Improvements in crop yields and nutritional value are urgently required to ensure food security. Generating novel allelic combinations by exploiting genetic diversity available in landrace collections provides one way to advance towards this objective. However, identifying valuable alleles remains one of the biggest challenges faced by wheat breeders.

We hypothesize that modern breeding has impacted on haplotype diversity in elite cultivars. In a pilot study, we analysed exome capture sequencing from UK elite cultivars and found that they have, three times less genetic variation than the Watkins landrace collection. The coding sequence-based data underestimates the genetic variation found in non-coding regions, which are known to be important for agronomic traits. Advances in the wheat genome assembly and annotation (IWGSC, v1.0), have opened up possibilities for in-depth analysis of the genetic variation across germplasm populations. In this study, we are employing promoter-exome and whole genome skim sequencing to develop genome-enabled tools to capture novel variation, not available in reference genomes, and construct haplotypes among 100 Watkins landraces and 113 elite UK cultivars. We will assess the phenotypic values of those haplotypes to test the effect of breeding for quantitative traits in elite germplasm.

A better understanding of how breeding has shaped these haplotypes will allow to uncover alleles from landraces with unexploited potential. More important, we will be stepping towards the exploitation of haplotype-based phenotypic analysis in wheat by innovating in a reference free-alignment variation discovery.

Improvement of Genomic Prediction by Including Additive-by-Additive Epistasis. A Case Study in Advanced Wheat Breeding Lines

Mr Miguel Angel Raffo

Biography:

Miguel Angel Raffo studied Agronomy in Universidad de la República, Uruguay, where he graduated in 2016. Then he obtained his MSc degree in Plant Science at the same institution and in combination with the National Institute of Agricultural Research of Uruguay (INIA), where he worked in wheat breeding for resistance to diseases. In 2019, he started a Ph.D. at the Center of Quantitative Genetic and Genomic (QGG) at the University of Aarhus, Denmark. The focus of his Ph.D. is on the application of statistical methods for genomic selection and prediction, with an emphasis on gene-by-gene and genotype-by-environment interactions on wheat.

Epistasis is the principal non-additive genetic effect in wheat since dominance is negligible in inbred lines. Epistatic interactions are fixed in inbred lines and can be used by breeders to select cultivars based on total genetic-merit. A correct model definition for variance component estimation contributes to disentangle the genetic architecture of complex traits in wheat. We aimed to i) evaluate the performance of the extended genomic best linear unbiased prediction (EG-BLUP) and the natural and orthogonal interactions approach (NOIA) for variance components estimation in a commercial wheat-breeding population, and ii) investigate whether including epistasis in genomic prediction enhance the predictive ability (PA) for wheat-breeding lines. In total, 2060 F6-lines from NordicSeed A/S Breeding-Company were phenotyped for grain yield in multiple years/locations in Denmark and genotyped using SNP15K-Illumina-BeadChip. Four models were used to estimate variance components and heritability on plot level: i) Baseline without genomic information, ii) G-BLUP, iii) EG-BLUP including additive-by-additive epistasis, and iv) NOIA. Narrow-sense and broad-sense grain yield plot heritabilities estimated with the G-BLUP were 0.15 and 0.31, respectively. EG-BLUP and NOIA failed to achieve orthogonal partition of genetic variances. Even though NOIA removed the Hardy-Weinberg-Equilibrium assumption, both models yielded similar estimates, indicating that the lack of orthogonality can be attributed to the linkage disequilibrium. PA of G-BLUP and EG-BLUP were studied using Leave-One-Line-Out cross-validation, and EG-BLUP increased PA (16.5%) significantly. We conclude that although the variance partition into orthogonal genetic effects was not possible, the EG-BLUP can enhance predictions of total genetic-merit contributing to better product development.

Genetic improvement of pacific white shrimp *Litopenaeus vannamei*: breeding program, progress and prospects

Fernanda Raidan

Biography:

I am a scientist passionate about performing research in the areas of quantitative genetics and statistical analysis. During my PhD research (2012-2016) I started building my research portfolio expanding my scientific interests, but at the same time keeping the focus on genetics by environment interactions, using cattle or fish as a model. I developed further skills on genomics and systems biology during a postdoctoral research position at the University of Queensland (2016-2017) and CSIRO (2017-2018). I joined CSIRO Aquaculture team as a quantitative researcher in May 2018 which the work focusses are attending prawns and abalone commercial breeding programs.

Pacific white shrimp (*L. vannamei*) is the world's most important cultured species on both the Atlantic and Pacific coast of the Americas and in Asia. Viet Uc Seafood Joint Stock Company is a Vietnam-based *L. vannamei* producer, breeder and seedstock supplier that since 2011, in collaboration with CSIRO, has completed 30 cycles of selection (~ 8 generations) with the primary objective of increasing growth rate. In each cycle, most parents were selected from the cycle that most recently reach sexual maturity (~ 9 months) and remaining parents (0-20) were selected from older cycles (12-15 months of age). Each parent was mated once with one other individual. Infrastructure allowed 40 families to be produced per cycle, with families maintained in separate tanks prior to family tagging and raceway stocking at approximately 10 weeks of age. After tagging, approximately 50 individual per family were stocked at commercial densities (adjusted with a proportion of untagged and tagged animals of 9:1) in plastic lined raceways for progeny testing. Univariate restricted maximum likelihood analyses were undertaken. Heritability for harvest weight at 5 months of age was 0.52 ± 0.10 and the selection response per generation was 7.9%, while the phenotypic improvement per generation was 1.1g. The use of the conventional methods of quantitative genetics to genetically improve the Pacific white shrimp has allowed for continuous progress of great value to increase the shrimp industry's profitability. Furthermore, recent advances in such areas as DNA marker based parental assignment will improve the accuracy of genetic parameters estimation and selection.

Prediction of evolutionary constraint by genomic annotations improves prioritization of causal variants in maize

Dr Guillaume Ramstein

Biography:

Throughout my education and research, I have trained in genetics applied to plant breeding. In 2017, I obtained my PhD from the University of Wisconsin-Madison, in quantitative genetics for inference and prediction in switchgrass. Since then, I have worked as a postdoctoral research at Cornell University. I have investigated strategies to incorporate information from computational biology and bioinformatics into quantitative genetics models. The overarching goal of my research has been to detect the polymorphisms responsible for observed differences in plant populations, without the confounding effects from genetic linkage that are common in QTL mapping and genomic prediction studies.

Precise functional annotation of genetic variants could shed light onto the genetic architecture of important agronomic traits. However, current QTL mapping methods typically lack the resolution to pinpoint candidate causal loci because of confounding effects of linkage disequilibrium (LD). To address this limitation, this study aims at leveraging functional effects and evolutionary constraint to avoid statistical biases from LD and identify classes of loci most likely to impact fitness. We conducted sequence analysis and in silico mutagenesis based on neural networks to estimate effects of polymorphisms on genomic structure (GC content, transposon insertion, k-mer frequency) and protein function (predicted effects of polymorphisms on peptide sequence or protein representations). We used within-species conservation (minor allele frequency in the maize Hapmap panel) and level of conservation across plant species (Genomic Evolutionary Rate Profiling scores) to reflect evolutionary constraint at loci. Analyses of functional enrichment in hybrid maize (a diverse panel and a collection of 24 bi-parental crosses) indicated that effects of polymorphisms on genomic structure and protein function could capture effects of variants on fitness-related traits such as plant height and grain yield. Prioritization of variants based on expression effects and evolutionary constraint, as assessed in this study, holds the potential to increase robustness of genomic prediction, and enable selection against mutation load, especially in landraces and newly domesticated species.

Inbreeding and behavioural plasticity in parent-offspring interactions

Mr Tom Ratz¹, Ms Anastasia Perodaskalaki¹, Dr Jacob A Moorad¹, Dr Per T Smiseth¹

¹University Of Edinburgh

Biography:

*I am currently a PhD student in evolutionary biology at the University of Edinburgh, Scotland. I am interested in the role of social interactions in evolution. My current research focuses on parental care and parent-offspring interactions in the burying beetle *Nicrophorus vespilloides*. Prior to the PhD, I graduated and completed a master's degree in behavioural ecology at the University of Tours (France) and a master's degree in mathematical modelling in ecology at the University of Rennes (France).*

Inbreeding can have major impacts on population persistence and evolution in changing environments. The severity of inbreeding depression often varies across environments, which could reflect that inbreeding reduces phenotypic plasticity and is supported by mixed evidence developmental traits. Despite the ubiquity of behavioural plasticity, its role in the short-term response to environmental changes and key life-history events such as mating, parenting or risk avoidance, it is unclear whether inbreeding can also alter behavioural plasticity. We investigated whether inbreeding reduces behavioural plasticity in social interactions in the context of parent-offspring interactions in the burying beetle *Nicrophorus vespilloides*. To do so, we recorded how offspring adjusted their begging and association with parents in response to increasing levels of hunger, and how parents adjusted their parental care with variation in brood size. We found that inbred offspring were more responsive than outbred ones in reducing their time spent associating with a parent in response to increasing level of hunger. We found no effect of inbreeding on plasticity of offspring begging or any parental behaviour. Overall, our results show that inbreeding does not necessarily reduce behavioural plasticity in parent-offspring interactions. Conversely, we found evidence that inbreeding can lead to increased plastic response, which suggests that inbreeding might alter canalisation of behavioural traits.

Efficient computation of the additive relationship matrix and its inverse in self-breeding individuals

Ms Ines Rebollo^{1,2}, J Rosas², I. Aguilar²

¹Facultad de Agronomía, Universidad de la República, ²Instituto Nacional de Investigación Agropecuaria

Biography:

Ines Rebollo is an Agricultural Engineer from the Universidad de la República (UdeLaR, Uruguay) and a second-year Master's student in Agricultural Sciences at the same University. She is advised by Juan Rosas and Ignacio Aguilar and has a scholarship from the Instituto Nacional de Investigación Agropecuaria (INIA) for her M.Sc. studies. She is interested in plant breeding and programming.

Inbreeding increases homozygosity and therefore additive relationships within and among related individuals. The main cause of inbreeding is breeding of related individuals in which case the inbreeding coefficient (F_s) = $0.5asd$ where asd is the additive relationship among the individual's parents. An extreme case of inbreeding is the self-breeding occurring in plant inbred lines such as those generated by multiple generations of self-pollination, or by double haploid production. In the case of selfing generations, $F_s = 1 - 0.5n$ where n is the number of selfing generations. If the parents of an individual are related and then it is self-bred, both sources for inbreeding should be accounted for in the progeny and $F_s = 1 - 0.5n + 0.5n(0.5asd)$. In order to perform Best Linear Unbiased Prediction (BLUP), accurate calculation of the additive relationship coefficients matrix (A) or its inverse (A^{-1}) is needed depending on the solving algorithm, or for single-step genomic BLUP where the submatrix A_{22} is used. Current methods to calculate A accounting for selfing generations require the expansion of the pedigree, which is computationally inefficient. Furthermore, freely available algorithms for setting up A^{-1} without generating A and inverting it do not contemplate selfing generations. The objective of this work was to develop efficient methods for calculating A and A^{-1} matrices accounting for inbreeding in self-bred individuals. Existing algorithms were adapted to account for selfing generations in A that require less memory than existing methods, and algorithms for the direct construction of A^{-1} accounting for inbreeding and selfing were developed in R. These algorithms are freely available at <https://github.com/minesrebollo>. In the future, these methods will be tested in large datasets and their performance will be reported.

Keywords: self-pollination relationship matrix, computing methods, Single-Step GBLUP.

Using Pool-Sequencing to Elucidate the Genetic Background of Seizures in the Göttingen Minipig

Dr Christian Reimer

Biography:

Originating from a rural area and background in Northern Germany, Christian Reimer received a BSc and MSc in agricultural sciences from University of Goettingen. He followed up as a PhD student in the animal breeding and genetics group of Prof. Dr. Henner Simianer and graduated with a thesis entitled "Sequence-based Analyses of the Goettingen Minipig Genome". As a post-doc, he focuses on genomic evaluations based on WGS- and array derived genomic data, as well as general problems of animal breeding, such as phenotyping of complex behavior, optimization of breeding goals or evaluation of genetic disorders.

The Göttingen Minipig (GMP) is a well-established animal model species in human toxicology testing, and especially benefits from its well defined genetic background. In the past, clinical symptoms of epileptic seizures have been observed in some cases. Even though incidences are minor, every appearance within a clinical trial might compromise the respective testing scheme and is therefore highly undesired. A heritability of ~ 0.42 , estimated based on pedigree data, indicates a strong genetic basis of this disorder. We gathered blood samples of 122 GMPs including 22 case animals and conducted short-read re-sequencing on 11 DNA-pools, one made up from case animals, to a target depth of 30X. Sequence preparation, variant calling and filtering resulted in a set of roughly 16M raw variants. We calculated fixation index and the sequence depth ratio between the seizure pool and the healthy pools from the same sub-population, as well as all healthy pools, respectively. We encountered regions of notable differentiation between the seizure and the healthy animals, including a region on chromosome X, usually found to be highly conserved in the GMP. Depth analysis revealed systematic depth differences for large regions, indicating that there might be different copy numbers between the contrasting samples. Annotation of high-FST SNPs with deleterious effect on protein synthesis, predicted by SIFT, resulted in a comprehensive candidate gene set, including DSCAM and DCDC1, both known to be involved in epilepsy in humans.

ImmuneDEX: Development of a genomic prediction for immune competence for the Australian Angus cattle

Dr Antonio Reverter¹, Dr Brad Hine², Dr Laercio Porto-Neto¹, Dr Sonja Dominik², Christian Duff³, Andrew Byrne³, Peter Parnell³, Dr Aaron Ingham¹

¹CSIRO Agriculture And Food, ²CSIRO Agriculture and Food, ³Angus Australia

Biography:

After graduating in Veterinary Sciences from the Universitat Autònoma de Barcelona, Catalonia, Antonio (Toni) Reverter was awarded a scholarship and in 1990 moved to Colorado State University to pursue a double degree of MS in Statistics and a PhD in Animal Sciences. In 1995, Toni joined the Animal Genetics and Breeding Unit in Armidale, NSW, Australia. At the AGBU, Toni worked on all aspects, including computer programming and quantitative analyses related to genetic evaluation system for beef cattle. Since 2002, Toni has been based in Brisbane with CSIRO where currently he is the Team Leader of Animal Genomics.

In animal breeding and genetics, the ability to cope with disease with minimal detriment on growth and fertility phenotypes is a desired objective addressing animal production and welfare considerations. However, due to the varying and complex nature of disease, defining and objectively measuring immune competence has long eluded animal scientists. Using standard protocols, we measure both cell-mediated innate immunity (IC1, for Immune Competence 1) and antibody-mediated acquire immunity (IC2, for Immune Competence 2). When applied to a population of 3,028 Australian Angus cattle representing 195 industry sires and genotyped for 50K SNPs, we show that IC1 and IC2 are moderately heritable (ie. 0.283 and 0.226 for IC1 and IC2, respectively) with a genetic correlation of 0.201. We further show how the genomic estimated breeding values (GEBVs) for IC1 and IC2 can be combined in a single index, namely ImmuneDEX that can be used as a breeding tool to identify selection candidates. ImmuneDEX exploits the correlation between the GEBVs for IC1 and IC2 as well as the ranking of animals by either metric while giving priority to the more heritable IC1. We use simulation to test the optimality of ImmuneDEX against the average value and the average ranking of IC1 and IC2, and with varying correlations between IC1 and IC2. We use real data to explore the relationship between ImmuneDEX and a range of phenotypes including temperament, growth, feedlot performance, and carcass quality.

Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration

Dr Joana Revez

Biography:

I graduated in Biology and completed a PhD in asthma Genetics. Currently, I am a postdoc in the Program in Complex Trait Genomics group, at the University of Queensland. My recent research focuses on understanding the contribution of genetic factors to vitamin D levels, and how these differ across different ancestry backgrounds.

Vitamin D deficiency, a trait influenced both by environmental and genetic factors, is a candidate risk factor for a range of adverse health outcomes. Here, we conducted a genome-wide association study (GWAS) of 25 hydroxyvitamin D (25OHD), the main metabolite used to assess vitamin D levels, in 417,580 Europeans from the UK Biobank and conducted a suite of post-GWAS analyses to clarify the relationship between 25OHD status and health. We identified 143 independent loci in 112 1-Mb regions providing new insights into the physiology of vitamin D and implicating genes involved in (a) lipid and lipoprotein metabolism, (b) dermal tissue properties, and (c) the sulphonation and glucuronidation of 25OHD. Mendelian randomization models found no robust evidence that 25OHD concentration had causal effects on candidate phenotypes (e.g. BMI, psychiatric disorders), but many phenotypes had causal effects on 25OHD concentration. We also conducted a GWAS of 25OHD variance to identify putative gene by environment (GxE) risk loci. We identified 25 independent variants associated with 25OHD variance, including five with strong evidence of interacting with season of measurement, and at least 10 GxE candidates with yet-to-be-identified environmental risk factors. These findings provide new insights into the physiology of vitamin D and the relationship between 25OHD status and health.

The transmissibility model to account for genetic and non-genetic inheritance in livestock species

Mrs Anne Ricard¹, Dr Ingrid David¹

¹INRAE, France

Biography:

Dr Ingrid David is a veterinarian with a master's degree in epidemiology and biostatistics and a PhD in animal genetics. Since 2008, she is a researcher at INRAE in Toulouse, France, in the Animal Genetics department. Her research focuses on interaction in genetics. She is trying to improve models for genetic evaluation accounting for genotype by environment and genotype by genotype interaction.

Animal selection in livestock species is performed by selecting animals based on genetic inheritance. However, non-genetic information such as microbiota, epigenetic marks and behavior can also be inherited across generations. To better predict the transmissible potential of each animal by taking into account these diverse sources of inheritance, we propose the “transmissibility model”. Similarly to the animal model, this model uses pedigree and phenotypic information to estimate variance components, but differs by estimating the sire and dam path coefficients of inherited information from parent to offspring instead of using a set value of 0.5 (additive genetic relationship matrix). On the one hand, we demonstrated the structural and practical identifiability of the transmissibility model, and showed, on the other hand, that deciphering the different sources of inheritance to predict the transmissible potential of individuals for each type of inherited factors is challenging. By simulations, we showed that the transmissibility model provided similar results to the animal model when inheritance was of genetic origin only, but outperformed the animal model for estimating the covariances between relatives and predicting the transmissible potential when the proportion of inheritance of non-genetic origin was high or when the sire and dam path coefficients were very different. Applied to the study of residual feed intake in different pig and rabbit breeds, the transmissibility model showed similar or better fit to the data than the animal model with a lower estimated sire than dam path coefficient of transmission indicating the existence of non-genetic inheritance for such trait.

Breeding for productive, low methane dairy cows

Ms Caeli Richardson

Biography:

Originally from the city, Caeli Richardson discovered her passion for the dairy industry while working on a 40-cow herd in Canada during her Bachelor's and Master's degree at the University of Guelph.

Caeli is excited for the opportunity to gain a new perspective of dairy farming as she is now a PhD candidate at La Trobe University and Agriculture Victoria under the supervision of Prof. Jennie Pryce. Currently her research focuses on the opportunity to improve the long-term sustainability of dairy cattle by selecting for profitable, environmentally friendly cows – primarily through the implementation of methane emission traits.

Methane is a greenhouse gas of high interest to the dairy industry with 57% of Australia's dairy emissions attributed to enteric methane. It is vital for future breeding objectives to focus on improving the sustainability of dairy cattle, by reducing methane emissions, without negatively affecting economically important traits. By including a methane trait in the national dairy selection index, a favourable response in methane and productivity may be achieved simultaneously. Residual methane production (RMP) has been proposed as a possible candidate trait as it has statistically favorable properties compared to methane yield and methane intensity. However, there is not yet consensus on the most appropriate method to calculate RMP. Individual cow methane, dry matter intake (DMI) and energy corrected milk (ECM) records were obtained from 379 animals at the Ellinbank Dairy Research Institute measured using the SF6 method and an electronic feed recording system, respectively. We compared nine definitions of RMP, including genetic and phenotypic regression of methane production on a combination of DMI and ECM using phenotypes or direct genomic values. Heritability estimates were calculated using univariate models and correlations were estimated using bivariate models and fitted using a genomic relationship matrix. RMP candidate traits had low to moderate heritability estimates (0.10 to 0.21) and all definitions of RMP had high genetic correlations with one another (>0.79) and other methane candidate traits (> 0.59). The results suggest that direct selection for a RMP trait may result in indirect, favorable improvement in all other methane traits.

Disentangling group specific QTL allele effects from genetic background epistasis using admixed individuals in GWAS: an application to maize flowering

Dr Simon Rio¹, Dr Tristan Mary-Huard^{1,2}, Dr Laurence Moreau¹, Mr Cyril Bauland¹, Mrs Carine Palaffre³, Mrs Delphine Madur¹, Mrs Valérie Combes¹, Dr Alain Charcosset¹

¹GQE - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, ²MIA, INRA, AgroParisTech, Université Paris-Saclay, ³UE 0394 SMH, INRA

Biography:

Hi everyone, I am Simon Rio from the Centre for Plant Biotechnology and Genomics (CBGP) in Madrid. Last year, I graduated from the University of Paris-Saclay in France where I did my PhD thesis under the supervision of Alain Charcosset. My thesis work focused on the impact of genetic structure and admixture in maize quantitative genetics studies, using both GWAS and genomic prediction. I am now working at the CBGP with Julio Isidro Sanchez on the optimization of multi-environment experimental designs, the application of GWAS and genomic prediction in oats and the study of the domestication of the Spanish grapevine.

When handling a structured population in association mapping, group-specific allele effects may be observed at quantitative trait loci (QTLs) for several reasons: (i) a different linkage disequilibrium (LD) between SNPs and QTLs across groups, (ii) group-specific genetic mutations in QTL regions, and/or (iii) epistatic interactions between QTLs and other loci that have differentiated allele frequencies between groups. We present here a new genome-wide association (GWAS) approach to identify QTLs exhibiting such group-specific allele effects. We developed genetic materials including admixed progeny from different genetic groups with known genome-wide ancestries (local admixture). A dedicated statistical methodology was developed to analyze pure and admixed individuals jointly, allowing one to disentangle the factors causing the heterogeneity of allele effects across groups. This approach was applied to maize by developing an inbred “Flint-Dent” panel including admixed individuals that was evaluated for flowering time. Several associations were detected revealing a wide range of configurations of allele effects, both at known flowering QTLs (Vgt1, Vgt2 and Vgt3) and new loci. We found several QTLs whose effect depended on the group ancestry of alleles while others interacted with the genetic background. The existence of directional epistasis was highlighted by comparing admixed with pure individuals and was consistent with epistatic interactions identified at the level of QTLs. Our GWAS approach provides useful information on the stability of QTL effects across genetic groups and can be applied to a wide range of species.

Accounting for group-specific allele effects and admixture in genomic predictions: theory and experimental evaluation in maize

Dr Simon Rio¹, Dr Laurence Moreau¹, Dr Alain Charcosset¹, Dr Tristan Mary-Huard^{1,2}

¹Université Paris-Saclay, INRA, CNRS, AgroParisTech, GQE - Le Moulon, ²MIA, INRA, AgroParisTech, Université Paris-Saclay

Biography:

Hi everyone, I am Simon Rio from the Centre for Plant Biotechnology and Genomics (CBGP) in Madrid. Last year, I graduated from the University of Paris-Saclay in France where I did my PhD thesis under the supervision of Alain Charcosset. My thesis work focused on the impact of genetic structure and admixture in maize quantitative genetics studies, using both GWAS and genomic prediction. I am now working at the CBGP with Julio Isidro Sanchez on the optimization of multi-environment experimental designs, the application of GWAS and genomic prediction in oats and the study of the domestication of the Spanish grapevine.

Populations structured into genetic groups may display heterogeneous marker effects due to group-specific linkage disequilibrium, mutations and/or interactions between quantitative trait loci and the genetic background. These factors affect particularly the efficiency of genomic prediction of admixed individuals, whose genome is a mosaic of chromosome blocks with different group ancestry and that can combine favorable characteristics of each group. We developed two genomic prediction models adapted to the prediction of admixed individuals in presence of heterogeneous marker effects: Multi-group Admixed GBLUP Random Individual (MAGBLUP-RI) modeling the ancestry of alleles, and MAGBLUP Random Allele Effect (MAGBLUP-RAE) modeling group-specific distributions of allele effects. MAGBLUP-RI can estimate the segregation variance generated by admixture while MAGBLUP-RAE can disentangle the variability that is due to main allele effects from that due to group-specific deviation effects. Both models were evaluated for their precision in estimating variance components and for their genomic prediction accuracy using a maize panel including lines from the Dent and Flint groups, along with admixed individuals. Based on simulated traits, both models were accurate to estimate their respective variance components and proved their efficiency to improve genomic prediction accuracy compared to a standard GBLUP model. For real traits, a clear gain was observed at low marker densities whereas it became limited at high marker densities. The interest of including admixed individuals in multi-group training sets was confirmed using simulated traits, but was variable using real traits. Both MAGBLUP models and admixed individuals are of interest whenever there exist group-specific SNP allele effects.

Unravelling genomic regions associated with temperament in American Angus cattle

Dr Hinayah Rojas De Oliveira

Biography:

Hinayah obtained her B.Sc. (2014), M.Sc. (2015), and Ph.D. (2018) from the Universidade Federal de Vicosa, Brazil. During her Ph.D. she had the opportunity to conduct part of her research at the University of Guelph (Canada) and University of Georgia (USA). Hinayah has experience in genetic and genomic analysis of livestock animals in general, with focus on longitudinal traits. Currently she is a postdoctoral researcher at the University of Guelph (Canada) and Purdue University (USA), working under supervision of Drs. Flavio Schenkel, Luiz Brito and Stephen Miller.

Temperament has been identified as a key breeding goal for the beef cattle industry as it is directly related to productive performance, longevity, carcass and meat quality, animal welfare and handlers work safety. Various temperament indicator traits have been proposed over time. In the United States, the beef cattle industry has adopted the docility score (1-6 scale), which is a moderately heritable trait (~0.44). Understanding the underlying biology affecting the phenotypic expression of temperament by identifying genomic regions and metabolic pathways is of great interest. In this context, a single-step-GBLUP-based genome-wide association study was performed using 555,368 animals (143,945 had phenotypic records and 156,001 were genotyped and/or imputed to a 50K SNP panel). An animal multinomial threshold model was used to fit the docility scores (i.e., 1, 2, 3 and 4/5/6 to represent docile, restless, nervous and aggressive animals, respectively) into the liability scale. The top 10 SNPs with the largest effects were selected as relevant SNPs and used to perform the functional analyses of positional candidate genes. Important Quantitative Trait Loci were identified on chromosomes BTA5, BTA6, BTA10, BTA16, BTA24, BTA26, and BTA27. These candidate genes included VPS13C, CASC4, CTDSPL2, PRDX6, PSD3, ACO2, and TEF. The identification of important genes associated with docility contribute to biologically validate this trait used for selecting for desirable temperament and this information could be used to optimize genomic prediction of breeding values by giving higher weights to the important genomic regions.

Detecting ongoing selection and local adaptation in United States Bos taurus beef cattle

Mr Troy N. Rowan^{1,2}, Dr John M. Hickey², Dr Robert D. Schnabel¹, Dr Jared E. Decker¹

¹University Of Missouri, ²The Roslin Institute

Biography:

Troy Rowan is a Ph.D. student in the Division of Animal Sciences at the University of Missouri. Troy's research applies theory and tools from quantitative genetics, population genetics, and genomics to understand how selection changes the genome over short periods of time. His work leverages massive commercially-generated datasets from U.S. beef cattle to map ongoing selection on complex traits, as well as signatures of local adaptation. Beyond answering these basic science questions, Troy's work is focused on delivering real outcomes to producers and improving the efficiency and sustainability of beef production.

Cattle are under intense selection pressures, both artificial and natural, making them an interesting system to study how selection acts on short timescales across a range of environments. We leverage temporally-stratified genotypes to identify genomic changes that have occurred in cattle populations over the last ~50 years. Additionally, using geographic information and climate data, we identify unique signatures of selection associated with local adaptation.

Imputed 850K genotypes for over 54,000 Red Angus, Simmental, and Gelbvieh cattle, born from 1970 to 2016 were analyzed using Generation Proxy Selection Mapping. We fit birth year as the phenotype in a LMM to identify ongoing shifts in allele frequency greater than those caused by drift. In each dataset, we identified dozens of genomic regions under directional selection. These regions include large-effect loci, pleiotropic QTL, immune genes, and numerous novel targets of selection. Simulations are exploring the applications of this method to other systems.

To identify local adaptation genomic loci, we used discrete climate zones and continuous 30-year average measurements for temperature, precipitation, and elevation as phenotypes in univariate and multivariate LMMs. These analyses identified dozens of alleles strongly associated with local environments. Many were associated with genes identified in human adaptation studies. We also observe associations within GRIA4, known to mediate cold-adaptation in Arctic cattle. Loci with selection signals are retained as features to calculate region-specific genomic predictions. Using an evolutionary approach to study selection allows us to create deliverable results for use in genetic evaluations while providing important insights into adaptation biology.

Strategy to imputation of large genomic deletions and utilizing them in mapping and genomic prediction in cattle

Dr Goutam Sahana

Biography:

The work presented in the poster is part of the PhD study of Md Mesbah-Uddin. He did his PhD at Aarhus University, Denmark and INRAE, France. He is currently Postdoctoral Associate at Medical & Population Genetics, Cardiovascular Disease Initiatives (CVDi), Broad Institute of MIT and Harvard, Email: muddin@broadinstitute.org.

The poster will be presented by Goutam Sahana, Senior Researcher, Center for Quantitative Genetics and Genomics, Aarhus University, Email: goutam.sahana@qgg.au.dk.

Large genomic deletions are potential loss-of-function variants and can be lethal. Analyzing whole genome data of cattle, we report 8,480 large deletions (199bp to 773kb). Breakpoint analysis revealed that the majority of the deletions were most likely generated by micro homology-mediated end joining. Genotype likelihoods for deletions were computed using a Gaussian mixture model, assuming a linear relationship between observed read-depth with unobserved copy-number. Our strategy to impute deletion genotypes to the SNP array-typed population included extending the whole-genome sequence reference with deletion genotypes, followed by a two-step genotype refinement approach using Beagle4 and SHAPEIT2, and finally, joint imputation of SNPs, indels, and deletions using Minimac3. We achieved an imputation accuracy of $r^2 > 0.6$ at MAF as low as 0.7% for SNPs and indels, and 0.2% for deletions. Subsequently, we performed GWAS for eight fertility traits. We report significant associations ($P\text{-value} < 5 \times 10^{-8}$) of 30,384 SNPs, 178 indels and 3 deletions in Holstein, 17 SNPs in Jersey, and 23,481 SNPs, 189 indels and 13 deletions in Nordic Red cattle. Candidate genes within 500 kb of lead SNP included genes with annotations such as embryonic lethality, male and female infertility, oocyte degeneration, abnormal estrous cycle, and decreased ovulation rate in mouse, cattle, and zebrafish. We investigated the effect on genomic prediction of including one or several (genetic) variance components for imputed sequence variants and deletions in three cattle breeds. We used simulations to study the advantage of using across population information in genomic prediction, especially for population with small training data.

Towards a genomic evaluation of cheese-making traits including candidate SNP in Montbéliarde cows

Dr Marie-Pierre Sanchez

Biography:

Marie-Pierre Sanchez, engineer at INRA-GABI in France since 2000, is a quantitative geneticist by training. She first worked in pig genetics and joined the Bovine Genetics and Genomics team (G2B) in 2012. She finished a PhD on “Genetic analysis of milk protein composition and cheese-making properties predicted from MIR spectra” in 2019. She has a broad experience in genome-wide association studies. She has published 35 scientific articles (h-index = 14).

In the From’MIR project, milk cheese-making and composition traits were predicted from 6.6 million mid-infrared (MIR) spectra in 410,622 Montbéliarde cows (19,862 with genotypes). GWAS on imputed whole genome sequences revealed candidate SNP that were designed on the EuroG10K BeadChip. Here, we assess the reliability of Single-Step GBLUP breeding values (ssEBV) from test-day records of the first three lactations for cheese yields, coagulation traits, casein and calcium contents. The available data were split in two independent training and validation sets that respectively contained cows with the oldest and the most recent lactations. The training set, including 155,961 cows (12,850 with genotypes), was used to predict ssEBV of 2,125 validation genotyped cows. For each trait, the reliability of ssEBV was estimated considering: 1) SNP of the Illumina BovineSNP50 BeadChip with equal effect variances (50K), 2) 50K and candidate SNP with equal effect variances (50K+) and 3) 50K and candidate SNP with increased effect variances for 5 to 14 candidate SNP, depending on the trait (CAND). The 50K+ and CAND scenarios led to similar mean reliabilities (67%) and both outperformed the 50K scenario (65%). The CAND scenario gave the less biased ssEBV. These results led to the implementation of a genomic evaluation for cheese-making traits predicted from MIR spectra in Montbéliarde breed. This study was funded by the French Ministry of Agriculture, Agro-food and Forest (CASDAR), the French Dairy Interbranch Organization (CNIEL), the Regional Union of Protected Designation cheeses of Franche-Comté (URFAC) and the Regional Council of Bourgogne-Franche-Comté, under the project From’MIR.

Efficient variance components analysis across millions of genomes

Ali Pazokitoroudi¹, Yue Wu¹, Kathryn Burch², Kangcheng Hou², Aaron Zhou¹, Bogdan Pasaniuc^{3,4,5}, Sriram Sankararaman^{1,4,5}

¹Department of Computer Science, UCLA, ²Bioinformatics Interdepartmental Program, UCLA, ³Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, ⁴Department of Human Genetics, David Geffen School of Medicine, UCLA, ⁵Department of Computational Medicine, David Geffen School of Medicine, UCLA

Variance components analysis has emerged as a powerful tool in complex trait genetics, with applications ranging from heritability estimation to association mapping. While the application of these methods to large-scale genetic datasets has the potential to reveal important insights into genetic architecture, existing methods for fitting variance components do not scale well to these datasets. Here we present RHE-mc, a randomized algorithm for variance components analysis that is accurate and highly efficient. RHE-mc can simultaneously estimate multiple variance components as well as estimate variance components associated with continuous annotations. The resulting computational efficiency permits RHE-mc to jointly fit 300 variance components in less than an hour on a dataset of about 300,000 individuals and 500,000 SNPs, about two orders of magnitude faster than state-of-the-art methods. On a dataset of one million individuals and one million SNPs, RHE-mc can fit 100 variance components in about 12 hours.

Analyzing 22 diverse traits with genotypes from 300,000 individuals across about 8 million common and low frequency SNPs (minor allele frequency > 0.1%), we observe that the allelic effect size increases with decreasing MAF (minor allele frequency) and LD (linkage disequilibrium) across the analyzed traits consistent with the action of negative selection. Partitioning heritability across 28 functional annotations, we observe enrichment of heritability in FANTOM5 enhancers in asthma, eczema, thyroid and autoimmune disorders.

Finally, we discuss how these randomized estimators can be used to obtain scalable estimators of a number of models including multi-trait and non-additive models.

Maize intraspecific competition dynamics: the genomic architecture behind competitive ability

Miss Aimee Schulz

Biography:

Aimee Schulz is a second-year Plant Breeding and Genetics PhD student in the Buckler Lab at Cornell University. Combining her interests in plant breeding, ecology, and evolutionary genetics, Aimee is studying the genomic architecture of competitive ability in maize. Aimee graduated from Iowa State University with a BS in Genetics and Agronomy in 2019. She was an NSF REU student at the Donald Danforth Plant Science Center in the Topp Lab where she studied the genetics of maize root architecture across the genus Zea. During her undergrad, Aimee conducted research with Dr. Matthew Hufford on inbreeding depression in teosinte populations.

The presence of specific neighbors can have a direct phenotypic effect on an individual within a population. These differences in phenotype due to one's neighbors are called associative effects, and can be interpreted as a heritable environmental effect in quantitative genetics. Over time, maize has been selected for cooperative rather than competitive tendencies as a result of increased planting densities with identical genotypes. Uniformity in potentially competitive traits such as plant height has remained an important breeding goal; taller plants are able to shade out shorter plants and acquire a greater degree of resources, ultimately limiting the yield of neighboring plants. However, the question remains as to what extent other traits such as leaf angle and root architecture, and the genes underlying those traits, play a role in determining competitive or cooperative ability. We hypothesize that, depending on the genotype, plants will have varying degrees of competitive ability and impact on their neighbors' resulting level of fitness. In this study, we used phenotypic data from the maize Nested Association Mapping (NAM) population containing 26 families across 11 environments to map how the genetics of a neighbor impact the observed phenotype of an individual. We will further investigate the genetics underlying a plant's competitive ability through QTL mapping and genome wide association studies for SNPs impacting neighboring plants, utilizing additional populations. Mapping competitive ability can potentially benefit breeders in maize and other crops by identifying lines with negative or positive associative effects that could be leveraged to make greater selection gains.

Using EHR biomarker results to uncover pleiotropy between depression and the immune system

Ms Julia M. Sealock¹, Dr. Georgios Voloudakis², Dr. Arden Moscati², Dr. Heather Lee³, Dr. Lea K. Davis¹
¹Vanderbilt University, ²Icahn School of Medicine at Mount Sinai, ³Center for Genomic Medicine at Massachusetts General Hospital

Depression has previously been associated with an activated immune system, however, the relationship is not well understood. Combining genetics of depression with clinically measured immune markers stored in electronic health records (EHRs) could aid in understanding the relationship between depression and the immune system. Using Vanderbilt University Medical Center's EHR-linked biobank, BioVU, we extracted and cleaned 272 lab tests, including 70 immune markers for genetic analysis. Next, we calculated depression polygenic scores on 67,977 BioVU individuals of European genetic ancestry. Associations between polygenic scores and labs were calculated using a lab-wide association scan (LabWAS), controlling for sex and top 10 principal components of ancestry. The LabWAS of depression polygenic score showed associations with several labs, including three immune markers, white blood cell count (WBC) (OR=1.04, p-value=6.02e-11), urinary WBC (uWBC) (OR=1.04, p-value=8.41e-6), and absolute monocyte count (OR=1.03, p-value = 4.17e-5). In a conditional analysis controlling for depression or anxiety diagnosis, WBC remained associated with depression polygenic score (OR=1.04, p-value=4.22e-8). The association between depression and WBC was further investigated using local genetic correlation, which calculates genetic correlation within independent linkage disequilibrium blocks. Local genetic correlation revealed 2 independent regions (chr 6 and chr17) significantly correlated between the traits. Our results confirm previous reports of a link between depression and an elevated immune system, and provide evidence of pleiotropy. We plan to replicate our findings in other EHR systems and further investigate the direction of association and regions mediating the association between depression and WBC.

Investigating the effect of anti-hypertensive medication on psychiatric disorders: a mendelian randomisation study

Dr Sonia Shah

Biography:

Sonia was awarded her PhD at University College London on the discovery and application of genetic determinants of cardiovascular disease risk factors, with particular application in disease risk prediction and determining causality using Mendelian randomisation. Her interests lie in using genetic, epigenetic and genomic data to better understand the underlying architecture of complex disease traits. In 2018, she was awarded an NHMRC Early Career Researcher fellowship to investigate the relationship between cardiovascular and neurodegenerative disease using large-scale genetic and omic data.

Background

There is growing evidence for a bi-directional relationship between psychiatric disorders and cardiovascular diseases (CVD). Given the growing evidence for shared aetiology, drugs widely used for the treatment of CVD may cause, exacerbate, or relieve neuropsychiatric symptoms. Randomised control trials to investigate this are lacking.

Aim

Use Mendelian randomisation (MR) analysis to investigate the effect of anti-hypertensive drugs on psychiatric disorders.

Methods

We use blood and brain expression quantitative trait loci (eQTL) data to select genetic instruments for the exposure of interest (anti-hypertensive treatment). We conduct a two-sample summary-based MR analysis to investigate the likely effect of different classes of anti-hypertensive on schizophrenia, bipolar disorder and major depressive disorder.

Results

Lower expression of the ACE gene in both blood and brain tissue, analogous to use of ACE-inhibitors, was associated with lower blood pressure, but increased risk of schizophrenia.

Conclusions

Genetic analyses suggest that use of certain classes of anti-hypertensive drugs may increase the risk of developing psychiatric disorders or may be detrimental if used to treat hypertension in individuals with psychiatric disorders, and warrant further investigation.

Reciprocal causation mixture model for mendelian randomization analysis

Professor Pak Chung Sham

Biography:

Pak Sham's area of research is the genetics of complex diseases, particularly psychiatric disorders. He studied physiology and psychology at Cambridge University, and medicine at Oxford. After training in psychiatry, he obtained a fellowship from Wellcome Trust to specialize in genetic epidemiology at the Institute of Psychiatry London, including an attachment to Virginia Commonwealth University. In 2000 he was appointed Professor of Psychiatric and Statistical Genetics at King's College London, and in 2004 he moved to The University of Hong Kong where he was head of Psychiatry in 2007-11, and director of genomics in 2011-2019.

Mendelian randomization (MR) using GWAS summary data is a useful method for inferring causal relationships between risk factors and diseases. However, standard MR requires key assumptions which are often implausible due to the widespread pleiotropy of single nucleotide polymorphisms (SNPs) effects on multiple phenotypes. The selection of invalid instrumental variables (IVs) will violate MR assumptions and lead to spurious inferences concerning causation. Additionally, current MR methods require a separate analysis to examine the causal effect in the reverse direction. Here, we propose a novel strategy to estimate reciprocal causal relationships between two phenotypes. Our method uses GWAS summary statistics of all SNPs on the two phenotypes, and takes account of the correlations between SNPs based on reference linkage disequilibrium (LD) information. Each SNP can have an effect on neither phenotype (null), one of the phenotypes (trait-specific), or both phenotypes (pleiotropic), so that the effects across all SNP follow a bivariate normal mixture distribution. The model is fitted to GWAS summary statistics using composite likelihood - the reciprocal causal paths, together with nuisance parameters, are estimated by the EM algorithm. We tested our method by simulation under various scenarios, including strong pleiotropy. The results show that the method gives nearly unbiased estimates of the reciprocal causation paths, as well as correct type I error rates under the null hypothesis. Thus, compared with existing MR methods, our method can infer bidirectional causation without the need for IV selection, making full use of genetic information while explicitly modelling pleiotropy.

GxEsum: Genotype-by-environment model based on summary statistics

Miss Jisu Shin

Biography:

Jisu Shin is a master student and research assistant in UniSA Allied Health & Human Performance at University of South Australia. In 2018, she received her Bachelor in Animal Biosystem Science from Chungnam University in South Korea. She is mainly focusing on understanding the genetic architecture of complex diseases and developing statistical models.

Genetic variation in response to the environment is fundamental in biology and has been described as genotype-by-environment interaction (GxE), reaction norm or phenotypic plasticity. In the genomic era, there has been increasing interest in estimating GxE, using genome-wide SNPs, e.g. a whole-genome reaction norm model (RNM; Ni et al. 2019) that can estimate unbiased genome-wide GxE, correctly accounting for genotype-environment correlation. However, the existing approach is computationally demanding and infeasible to handle > 100,000 samples. Here we introduce GxEsum, a model for estimating GxE based on GWAS summary statistics, which can be applied to a large-scale biobank data (e.g. ~ 500,000 samples). In real data analyses using UK Biobank data after QC, the computing time and precision (i.e., power) of GxEsum based on 288,837 individuals are hundreds of times faster and ~ 3.5-fold higher than those of RNM based on 50,000 individuals. As the scale of available resources has been increased, GxEsum may be an efficient tool to estimate GxE that can be applied to large-scale data across multiple complex traits.

Genome-wide association study identifies novel variants associated with host resistance to bovine tuberculosis

Dr Masoud Shirali

Biography:

Dr Masoud Shirali is a senior quantitative geneticist and head of Genetics Unit in AFBI Hillsborough. He is an expert in designing and developing data-driven methods for understanding, prediction, and control of complex systems. He conducted his PhD in genetics and animal breeding and worked as a research fellow in University of Edinburgh for 7 years. He has researched in the areas of genomics, transcriptomics, metabolomics, and multi-omics. Masoud is the (co-)author of over 30 peer-reviewed scientific publications including publications in high impact factor journals such as Nature Genetics, Nature Neuroscience, Nature Communications, PLoS One, Heredity and Scientific reports.

This study aims to further enhance our knowledge of the genetic architecture underlying host resistance to Bovine tuberculosis (bTB) in dairy cattle. Determining genetic variants associated with bTB phenotypes may help with the design of future diagnostics / vaccines and efforts to breed more resistant cattle.

A population comprised 1,153 Holstein cows were selected from herds across Northern Ireland: 273 controls and 880 cases were selected from bTB affected herds. Samples were genotyped using the Illumina Bovine50 SNP chip.

A GREML analysis for family data (Zaitlen et al. 2013) using two genomic relationship matrixes (GRMs) by GCTA (Yang et al. 2011), was used to estimate variance components. To identify genomic regions associated with host resistance to bTB we undertook; 1. A genome-wide association study (GWAS) using a mixed model containing the two GRMs, by GCTA-MLMA (Yang et al. 2014); 2. A gene analysis using the single SNP GWAS summary statistics by MAGMA (de Leeuw et al. 2015); 3. The regional heritability mapping method (Nagamine et al. 2012). Significance level was set to the Bonferroni corrected p-value (0.05).

On the liability scale, the genomic heritability estimated by whole genome GRM and kinship GRM were 0.30 ± 0.06 and 0.24 ± 0.06 , respectively. The single SNP GWAS analysis resulted in detecting six independent loci. The gene analysis resulted in detecting two genes associated with host resistance to bTB. The RHM results demonstrated 13 significant regions.

Protein-Protein Interaction networks identify Hub proteins related to Nelore cattle tenderness

Dr Larissa F Simielli Fonseca

Biography:

Larissa Fernanda Simielli Fonseca is post-doctor researcher at São Paulo State University (UNESP), Brazil, focused in analysis of functional genomics data, mainly tropical cattle transcriptome and proteome studies. Her master and doctor degrees, also received at UNESP, were dedicated to understand genes expression profile in Nelore cattle. She has a longstanding interest in bovine genetic improvement, molecular genetic techniques approaches and functional genetic tools applied to improvement of quality and feed efficiency of Nelore beef cattle production.

Liquid Chromatography Mass Spectrometry (LC-MS/MS) was used to identify differentially expressed proteins in Nelore cattle muscle tissue. The analysis included 20 Longissimus thoracis muscle samples significantly divergent (t -test <0.05) for Warner-Bratzler Shear Force (WBSF), being 10 with tough (9.16 ± 0.53 kgf) and 10 with tender (3.31 ± 0.56 kgf) meat. Fold-changes obtained for the 656 identified proteins were used to generate Protein-Protein Interaction Networks, using Contextual Hub Analysis Tool (CHAT) and CentiScaPe app, implemented in Cytoscape software. Node degree and EigenVector were used to characterize the overall network topology, and IntAct EML-DBI database was used as reference to predict interactions. The ATPB, COX5B and COX5A bovine proteins were identified as hubs. These top three-proteins act in oxidative phosphorylation pathway in mitochondrial inner membrane, a major source of ATP in aerobic organisms. After slaughter, blood circulation is interrupted, ceasing oxygen and ATP supplies, causing anaerobic conversion of glycogen into glucose, producing lactate and reducing the intracellular pH. These biochemistry processes are important for meat tenderization, suggesting that the hub proteins are related to less meat tenderness, since they were more expressed in hard meat group and act in an aerobic environment. Acknowledgments: São Paulo Research Foundation (Grants #2016/23937-6, #2017/10630-2, #2018/20026-8 and #2019/16732-7) and Coordination for the Improvement of Higher Education Personnel - Brasil (CAPES - #001 and #88887.337783/2019-00).

Analysis of Genetic and Environmental Contributions to Lipids Reveals Differential Patterns of Disease Associations in a Large BioBank

Ms Kritika Singh^{1,2}, Ms Julia M Sealock^{1,2}, Dr. Dorret I. Boomsma³, Dr. Lea K. Davis^{1,2}

¹Vanderbilt Genetics Institute, Vanderbilt University Medical Center, ²Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, ³Dept Biological Psychology, Vrije Universiteit

Biological markers are measurable substances found in biospecimens that can indicate the severity or risk of a particular outcome, like disease or infection. Blood lipid measurements (i.e., LDL, HDL and Triglycerides) are important markers in determining an individual's risk of developing cardiovascular and metabolic diseases. The variation between people in these biomarkers is due to a combination of both genetics and environment. The common variant genetic contribution to each biomarker can be estimated using a polygenic score (PGS). The PGS reflects a measurable part of the common genetic contribution to lipid levels. Because genetics and environment together contribute to the variance in observed lipid values, removing variance due to genetics from the overall measured value can approximate the environmental contribution to lipids. The resulting residual variance, termed the 'lipid residual' reflects environmental variation and measurement error. We performed a phenome-wide association study (PheWAS) to examine the effect of PGS and the lipid residuals for LDL, HDL and Triglycerides on 1815 different medical diagnoses present in the electronic health records of 62,400 patients at Vanderbilt University Medical Center. For each lipid type, we observed that the lipid residual demonstrated a far greater number of disease-associations than the PGS indicating that the environmental contribution to each biomarker is a substantial driver of subsequent disease. After covarying for age, ancestry and batch effects, the environmental component of HDL ($\beta= 0.35, S.E.=0.03, p=9.90e-40$) and Triglycerides showed significant associations to tobacco use disorder ($\beta= 0.27, S.E.=0.02, p=3.74e-31$). We are currently investigating the associations of LDL residuals with statin effects.

Genomic Patterns of Structural Variation in the Sorghum Biomass Association Panel

Dr Kittikun Songsomboon¹, Dr Elizabeth Cooper¹

¹University Of North Carolina – Charlotte, USA

Biography:

Kittikun (Chris) Songsomboon is a postdoctoral researcher in Bioinformatics at the University of North Carolina at Charlotte. He holds a Ph.D. in Plant Breeding from Cornell University. His current research interest is to understand and potentially utilize genomic structural variations in sorghums. Besides, he also works with an iron deficiency condition of sorghums to determine the effects of iron on sorghum growth. His career goal is to integrate technology and approaches into his breeding program, of which he is starting as a newly hired faculty at Kasetsart University, THAILAND.

Sorghum (*Sorghum bicolor*) is the fifth most important cereal crop in the world and has been bred for multiple purposes, resulting in strikingly diverse phenotypes. Structural variants (SVs) in the genome are a key source of variation and could potentially underlie some of these agronomically important phenotypic differences. In this study, we identified genome-wide SVs in the Biomass Association Panel, a collection of 350 diverse sorghum genotypes collected from multiple countries and continents. Using Illumina-based, short-read whole genome resequencing data from every genotype, we found a total of 8,182 SVs after filtering (7,777 deletions, 338 duplications, and 67 inversions). The size of SVs ranged from 3 bp to 92,311 bp in length (median 931 bp). The global site frequency spectrum of the SVs fit a model of neutral evolution, suggesting that the vast majority of SVs are not under any type of selection. Most SVs were also shared among all geographic locations; only 4 SVs (all of which were in non-coding regions) were unique to a single country and only 7 were unique to a single race. Despite this, we observed a distinct clustering pattern that was not associated with current or historical geography, and there were 260 SVs that were unique to a single cluster representing a subset of lines collected from Ethiopia. Among these SVs were many mutations occurring in genes related to disease resistance and transport, which could be indicative of local adaptation. These SVs will be examined for their associations with agronomic traits for further breeding

SNP-based gender determination in *Bos indicus* and *Bos taurus* cattle and their crosses

Dr Eva M. Strucken

Biography:

Dr Strucken studied Agriculture and later Process and Quality Management, specializing in Animal Breeding, at the Humboldt-University of Berlin, Germany. She extended her studies into animal breeding and quantitative genetics during an exchange with the Wageningen University, Netherlands, and as a Marie-Curie fellow at the Roslin Institute, UK. She worked as a scientific researcher at the Humboldt-University before taking up a position as a research fellow at the University of New England, Australia, where she currently focuses on developing tools for breed proportion estimation and parentage assignment in African and Indian dairy cattle populations.

The most common measure of gender from SNP data is heterozygosity of markers on the X-chromosome, where males have zero heterozygosity other than genotyping errors. We assessed gender based on X-chromosome heterozygosity of 2,218 and 39,440 SNPs in two datasets of 1,058 and 1,136 pure indicine and taurine cattle and their crossbreds genotyped with the GGP50k and the 777k Illumina BovineHD Beadchip SNP assays, respectively. Male heterozygosity reach a maximum of 0.11 with the GGP and 0.04 with HD data. Female heterozygosity reach a minimum of 0.03 with 50k and 0.01 with 777k data, such that unambiguous gender determination was not possible. In human data, increased accuracy has been achieved when Y-chromosome markers were included, where females should have no recorded genotypes, except for genotyping errors. Y-chromosome genotypes on the GGP were unreliable but using the 808 Y-genotypes on the HD assay gave a clear separation of males versus females. Pure taurine cows had a mean and minimum heterozygosity of 0.32 (± 0.04) and 0.23 in the 50k, and pure indicine cows had 0.1 (± 0.01) and 0.09, and 0.19 (± 0.02) and 0.01 in the 50k and the 777k data, respectively. Ascertainment bias (higher heterozygosity in *Bos taurus* cattle) clearly plays a role in X-chromosome heterozygosity for the GGP50k SNPs but not obviously for the Illumina 777k SNPs. Inbreeding, accentuated by single chromosome inheritance through selected sires, also seemed to decrease X-chromosome heterozygosity. Autosomal inbreeding levels and indicine breed content can be used to improve gender determination in most cases.

Improving accuracy of genomic prediction by fitting epistasis

Dr Jinyan Teng

Biography:

Jinyan Teng is a Ph.D. student in Prof. Zhe Zhang's lab (animal breeding and genetics) at College of Animal Science, South China Agricultural University (SCAU). He received his Master Degree in 2020 from SCAU. His research is focused on the new methods and application of genomic prediction. Currently, their lab's project is to develop new models of genomic prediction by incorporating biological or multi-omics information. You can follow their research team on ResearchGate: <https://www.researchgate.net/lab/Zhe-Zhang-Lab-4>

Genomic prediction (GP) was widely used for predicting the genetic or phenotypic value of complex trait. Though numbers of genomic prediction studies based on additive genetic effects (A) archived a considerable accuracy, alongside the availability of big data, it is potential to include genetically interactive effects (D and I) in genomic prediction model. Since the biological interaction of gene sets from KEGG pathway was publicly accessible, incorporating the KEGG as a carrier of interaction (i.e. epistasis) into genomic prediction model may be a method potentially improving accuracy of genomic prediction. Conventional genome-enabled best linear unbiased prediction (GBLUP) and an pathway model (PGBLUP) that polygenes effect combined a separate random component matrix from KEGG in model were validated in a German Holstein population. The cattle population used in this study includes 2,000 genotyped individuals with 54K SNPs and six traits with highly reliable estimated breeding values. Model assessment was performed by 10 times 10-fold cross-validation. Predictive accuracy was defined as the Pearson's correlation coefficient between the predicted genetic values and phenotypic values. We found that the predictive accuracy of PGBLUP incorporating the best pathway outperformed GBLUP in six tested traits with the relative advantage of from 0.4% to 15%. This study concluded that KEGG pathway can as a carrier of interaction in genomic prediction model, and there is always an optimal pathway can greatly improve the predictive accuracy in the tested cattle population.

Improving polygenic risk prediction of complex diseases in non-European populations by reweighting SNPs by fixation index

Dr Shu Mei Teo

Biography:

Shu Mei completed her PhD with the National University of Singapore and Karolinska Institutet. Her thesis focused on statistical methods for the improved detection and analyses of structural variants.

In 2013, she moved to Melbourne as a post-doc. Her work on the longitudinal dynamics of the airway microbiome and its links with childhood asthma won her the 2019 Klosterfrau-Group Award for research excellence in the field of pediatric lung diseases. Shu Mei now works on polygenic risk scores (PRS), including their transferability to non-European populations, as well as using PRS to understand the molecular mechanisms through which genetics causes diseases.

Polygenic risk scores (PRS) combine the totality of genomewide disease-associated signals. They have the ability to identify individuals at high genetic predisposition to complex human diseases. However, current PRS are (unsurprisingly) biased toward prediction in Euro-centric populations, due to majority of large genomewide association studies (GWAS) conducted in those populations. The aim of this study is to improve the prediction of coronary artery disease (CAD) and Type-2-diabetes (T2D) in non-European populations. This is imperative in minimising disproportionate benefits from future use of PRS in clinical interventions.

The lack of transferability of PRS have been shown to be due primarily to differences in SNP allele frequency and linkage disequilibrium, hence we hypothesize that SNPs with little differentiation (i.e. low fixation index, F_{ST}) between the discovery (GWAS) and target populations should show less attenuation of trait-PRS effect sizes, as compared to SNPs with high differentiation. We investigate this by calculating PRS for CAD and T2D based on reweighting SNPs as a function of their between population F_{ST} in the British White (BW), South Asian (SA) and Black British (BB) populations in the UK Biobank.

Preliminary results show the best F_{ST} -reweighted PRS for BB resulted in a 7.9% and 6.9% relative improvement in odds or hazards ratio over the regular PRS, for T2D and CAD respectively. These scores offer ~1% improvement in the SA population, and no improvement in the BW population. We have shown that by leveraging on between population SNP differentiation, we can improve prediction in non-European populations.

Multi-phenotype GWAS of Chronic Kidney Disease in the Norfolk Island isolate

Ms Kim Ngan Tran

Biography:

Ngan Tran studied her master's degree in Chonnam National University, South Korea. There, she was introduced to high-throughput data analysis and got herself highly interested in Computational Biology. Before starting her PhD journey in 2019, she had chances to be involved in multiple projects in transcriptome analysis, genome assembly, and pharmacogenomics. She is working on quantitative genomics of complex traits in the Norfolk Island isolate. Specifically, her PhD project involves genomic analysis of chronic kidney disease in one of the most characterized and famous research cohorts in the world - the Norfolk Island Population.

Chronic kidney disease (CKD) is defined as the persistent impairment in kidney function. GWASs have revealed multiple genetic loci associated with CKD susceptibility. However, the complete genetic basis of CKD is not yet clear. Since CKD shares risk factors with cardiovascular diseases and diabetes, pleiotropic loci may play a role in the genetic basis underlying CKD but can go undetected when using single phenotype GWASs. The use of genetically isolated cohorts can offer additional advantages for gene mapping because of features such as founder effect and large multigeneration pedigrees. In this study, we performed multi-phenotype GWASs in the Norfolk Island isolate to identify new loci involved in traits related to CKD. In total, we included 380 related individuals of NI isolate with 29 continuous phenotypic measurements and 4.7 million genotyped and imputed SNPs. We performed factor analysis to extract phenotypic components representative for multiple traits, examined their narrow-sense heritability and finally used significantly heritable components as dependent variables in the downstream linear mixed association models. GWAS of principal component one, which was derived from the 3 CKD-primary traits, i.e. eGFR, serum creatinine, and serum urea, identified a loci mapping to KCNIP4 ($p_{\min}=1.67 \times 10^{-7}$) while there was no association peak for these traits when analyzed individually. Inclusion of other secondary CKD measurements also identified the KCNIP4 locus with greater statistical significance ($p_{\min}=1.59 \times 10^{-9}$). This study implicates KCNIP4 as a novel pleiotropic gene underlying CKD. Further studies are underway to assess functional relevance of this locus.

HSSGBLUP: a Single-Step SNP BLUP genomic evaluation software adapted to large livestock populations

Dr Thierry Tribout

Biography:

Dr. Thierry Tribout (male), engineer at INRAE since 1998, is a quantitative geneticist by training. He first worked in pig genetics, mainly in the implementation and optimization of the genetic evaluation of the French national pig breeds, and completed a PhD on the interest of genomic selection in pig breeding schemes. He joined the Bovine Genetics and Genomics team in GABI department in 2013. He mainly works on the genomic evaluation of cattle, and has expertise in genome-wide association studies.

In France, the current genomic evaluations based on multi-step methodology will soon be replaced by Single-Step Genomic BLUP evaluations in all dairy and beef cattle populations, using a genomic evaluation software developed by INRA.

The HSSGBLUP program is written in Fortran 90. The numbers of genotyped animals in the largest French populations exceeding hundreds of thousands of individuals, the software is based on a Single-Step biallelic marker effects BLUP formulation, to avoid the computational problems inherent to inverting large genomic relationship matrices as in classic breeding values Single-Step formulations. The program was conceived to perform genomic evaluations in very large populations while containing the amount of memory required, by using preconditioned conjugate gradient (PCG), sparse matrices approaches, an iteration on data strategy, and parallelizing the computations.

The imputed genotypes of the non-genotyped animals required in the model are calculated on-the-fly within iterations, without storage. These computations, as well as those of the covariate vector needed to fit the mean of unselected base population in the model to ensure the consistency of pedigree based and genomic relationships are based on the use of the Cholesky factor of the submatrix of the inverse of the pedigree relationship matrix relative to ungenotyped animals, which is efficiently obtained using MKL-Pardiso multiprocessing parallel solver.

The HSSGBLUP program can handle various models, such as models including direct and maternal genetic effects, weighted phenotypes, heterogeneous variances (multiplicative model), and adjustable genetic variances for markers to give appropriate weights to identified causal mutations or candidate variants.

Optimising genotype imputation to improve genomic selection for disease resistance in Atlantic salmon

Dr Smaragda Tsairidou

Biography:

I am Research Fellow in Aquaculture Genetics at the Roslin institute, University of Edinburgh (Professor Ross Houston group). My research combines quantitative genetics, and data-driven approaches and bioinformatics to develop cost-effective selective breeding strategies for resistance to sea lice and amoebic gill disease in Atlantic salmon. I am a Doctor of Veterinary Medicine. I received my MSc in Quantitative Genetics and my PhD in genetics and genomics at the University of Edinburgh, studying the genetics of bovine Tuberculosis. In my first post-doc in genetic-epidemiology, my research focused on infectivity and the use of selective breeding for disease control.

Genomic selection for polygenic traits, such as resistance to sea lice, can accelerate genetic gain and contribute to sustainable disease control in salmon aquaculture. However, it requires genome-wide genetic data on large populations, which can be prohibitively expensive. The aim of this study was to optimise the use of low-density genotypes and evaluate genotype imputation strategies for cost-effective genomic prediction.

Our study focused on sea lice count from a pedigreed Atlantic salmon population (Landcatch, UK), with a full-sib/paternal-half-sib structure. Post-smolt fish were challenged with *L. salmonis* and genotyped with a high density Atlantic salmon SNP chip. Variance components and breeding values were estimated using a GBLUP approach, and genotype imputation was tested by creating in silico low density SNP panels for offspring and high density panels for parents. To assess the impact of varying SNP panel density and composition, prediction accuracy was calculated through five-fold cross-validations and compared across SNP densities, with and without imputation.

The heritability for sea lice resistance was 0.19 (0.07). Reducing SNP density had little impact on prediction accuracy until 5,000 SNPs, below which the accuracy dropped. Imputation accuracy increased with increasing imputation panel density. Genomic prediction accuracy when offspring were genotyped for just 200 SNPs, and parents for 5,000 SNPs, was 0.53. This accuracy was similar to the full high density dataset and markedly higher than using low density true genotypes alone. These results suggest imputation from very low to medium density can be a cost-effective tool for genomic selection in Atlantic salmon breeding programmes.

Genome-wide association study of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain

Dr Yakov Tsepilov

Biography:

Dr. Yakov Tsepilov is the leading researcher and head of the lab at the Institute of Cytology and Genetics, Novosibirsk, Russia, and a part-time faculty member at the Department of Natural Sciences at the Novosibirsk State University. His main research interest is the quantitative genetics, human genetics of back pain and human lifespan and animal breeding.

Chronic musculoskeletal pain has a negative impact on all aspects of human life. Genetic studies of pain are complicated by the high complexity and heterogeneity of pain phenotypes. In this research, we aimed to reduce phenotype heterogeneity and reveal genes and pathways shared by chronic musculoskeletal pain at four locations: back, neck/shoulder, hip, and knee. Our study was based on the results of genome-wide association studies performed using UKBiobank data with a total sample size of 456,000 individuals. We applied principal component analysis based on the matrix of genetic covariances between the studied pain traits and constructed four genetically independent phenotypes (GIPs). The leading GIP (GIP1) explained the largest proportion of the genetic variance (78.4%). We identified and replicated five loci associated with GIP1; and one locus associated with GIP2. The genes confidently prioritized for the GIP1-associated loci were SLC39A8, ECM1, and FOXP2. For the remaining two GIP1-associated loci, we proposed several candidates but were unable to prioritize any of them convincingly. The most likely causal gene in the locus associated with GIP2 was GDF5. For GIP1, gene set/tissue/cell type enrichment analyses identified multiple terms related to the nervous system. Genetic correlation analysis revealed a genetic overlap between GIP1 and osteoarthritis as well as a set of anthropometric, sociodemographic and psychiatric/personality traits. We suggest that GIP1 represents a biopsychological component of chronic musculoskeletal pain, related to physiological and psychological aspects and possibly reflecting pain perception and processing. The research has been conducted using the UKBiobank (project #18219).

Genetic parameters for fertility and MIR predicted fertility in Australian dairy cattle

Dr Irene Van Den Berg

Biography:

Irene van den Berg is a research scientist at Agriculture Victoria in Melbourne, Australia. She has previously worked on the use of sequence data for genomic prediction, multi breed genomic prediction and QTL mapping, and is currently focussing on genetic parameters and genomic prediction of novel traits such as MIR predicted fertility and serum biomarkers.

Reversing the decline in dairy cow fertility has been important worldwide. However, genetic gain is hampered by low heritability estimates of fertility traits and many key fertility traits are only available on a small proportion of cows. Mid infra-red (MIR) spectroscopy data of milk samples are easier to collect, through routine milk-recording and might be useful to predict fertility. Our objective was to estimate genetic parameters and correlations for fertility and MIR predicted fertility in Australian dairy cattle and assess whether selection criteria for fertility could be expanded to include MIR traits and increase prediction accuracy.

Our dataset contained 9,040 Australian dairy cows with MIR predicted fertility, of which 4,878 cow records for overall pregnancy (preg), pregnancy after first insemination (pregAI1) and pregnancy after second insemination (pregAI2), and 3,259 cows were genotyped with the 50K SNP chip. We estimated heritabilities and genetic correlations using a pedigree relationship matrix (A), genomic relationship matrix (G) and H-matrix combining the pedigree and genomic relationship matrices (H).

Heritabilities for the fertility traits were not significantly different between the different models and ranged from 0.01 for pregAI2 with A and G to 0.04 for pregAI2 with G. The heritability of MIR predicted fertility was significantly higher than that of actual pregnancy traits, with estimates of 0.12, 0.07 and 0.10 for A, G and H, respectively. Provided genetic correlations between MIR predicted fertility and existing selection criteria for fertility are reasonable, then MIR could be a useful additional selection criterion in multi-trait selection indices for fertility.

Maternal Transmission Ratio Distortion in Iberian pigs from two different varieties

Dr Marta Vázquez-Gómez

Biography:

Marta Vázquez-Gómez is a postdoctoral researcher in the Animal Science Department at the Universitat Autònoma of Barcelona (Spain). She is currently working on the study of reproductive characteristics between different varieties of the Iberian pig. She obtained a Ph.D. degree from the Complutense University of Madrid (Spain) at the end of 2018, working on the postnatal effects of the birth-weight variability in this same breed. During her Ph.D., she participated in several international networks and carried out two research experiences at INRAe (France). Furthermore, she has been involved in studies using this pig breed as a model for metabolic disorders.

Transmission ratio distortion (TRD) is defined as the deviation from the expected Mendelian genotypic frequencies in the heterozygous parents. Although TRD can be a confounding factor in linkage tests, this remains mostly unknown in pigs and may relate to reproduction anomalies, remains mostly unknown in pigs, and without maternal information. We aimed to describe the TRD prevalence and its genomic distribution in Iberian sows. One hundred twenty-nine offspring females from 52 Entrepelado Iberian dams and 118 offspring females from 67 Entrepelado Iberian dams were genotyped with the Geneseek Genomic Profiler Porcine HD (Illumina, San Diego, USA). The offspring were sired by both purebred Retinto and Entrepelado Iberian boars, regardless of the dam variety. Only 21,803 SNPs remained with a family consistency rate higher than 75% and 20% or more within-variety heterozygous sows, after quality control. Maternal TRD was evaluated by a likelihood ratio test SNP-by-SNP, and results were corrected for multiple testing using the false discovery rate approach ($\alpha = 0.05$). Results provided 24 SNPs in Retinto variety and 68 SNPs in Entrepelado variety with significant maternal TRD ($q < 0.05$). Both varieties shared 10 common SNPs. Most of them increased the transmission of the minor allele (negative TRD values). Moreover, some genes contained more than one TRD locus in both Iberian pig varieties. These findings could provide useful data for future studies of gametogenesis and embryogenesis, and provide relevant information for the pig industry, particularly in traditional breeds.

Genetic determinism of rabbits' cecal microbiota

María Velasco

Biography:

María Velasco is a Ph.D. student at the Genetics and Animal Breeding department of IRTA in Spain. She has a Genetics degree from the Autonomous University of Barcelona; and an MSc diploma with maxima cum laude mention in Animal Breeding and Reproduction Biotechnology from Polytechnic University of Valencia and Autonomous University of Barcelona. Since 2015, she has been involved in the characterization of rabbit cecal microbiota. Her thesis objectives are to unravel the effect of different management factors on the microbiota, to study the genetic determinism of the rabbit cecal microbiota, and the relationship between microbiota and productive traits.

Rabbit cecum harbors a complex microbial ecosystem whose members constantly interact between them and with their host to ensure homeostatic balance maintenance. Our objective was to estimate heritability of rabbit cecal microbiota in order to assess whether interactions with the host also occur at genetic level. 16S rDNA amplicon MiSeq sequencing was conducted on cecal samples collected from 425 kits. These rabbits were bred in two experimental farms and fed with the same diet, supplemented or free of antibiotics, but under different intake levels (ad libitum/restricted). Bioinformatics analysis of the paired-end sequences was performed with QIIME software. Taxonomic assignment of the sequence variants inferred was based on Greengenes database gg_13_5_otus. We used Bayes Factor to test the existence of host's genetic determinism on CSS-normalized OTUs and genera relative abundances. The consideration of Bayes Factor as model choice criteria evidenced the existence of a certain host genetic determinism on 198 CSS-normalized OTUs and 12 genera relative abundances. Most of the analyzed microbial traits seem to have low medium heritabilities. However, 21 CSS-normalized OTUs and 2 genera have heritability estimates larger than 0.3.

Does rabbits' cecal microbial information improve prediction of individual growth, and cage feed intake and feed conversion ratio?

María Velasco

Biography:

María Velasco is a Ph.D. student at the Genetics and Animal Breeding department of IRTA in Spain. She has a Genetics degree from the Autonomous University of Barcelona; and an MSc diploma with maxima cum laude mention in Animal Breeding and Reproduction Biotechnology from Polytechnic University of Valencia and Autonomous University of Barcelona. Since 2015, she has been involved in the characterization of rabbit cecal microbiota. Her thesis objectives are to unravel the effect of different management factors on the microbiota, to study the genetic determinism of the rabbit cecal microbiota, and the relationship between microbiota and productive traits.

Microbial communities inhabiting gastrointestinal tract are expected to affect animal growth and feed efficiency given their well-known role in nutrient absorption and immunological processes. Our objective was to assess whether predictive ability of linear models for growth (G) and cage feed intake (FI) and feed conversion ratio (FCR) is improved by including microbial information; and to assess its effect on variance components estimates. Our dataset comprised 425 kits bred in two farms and fed with the same diet; supplemented or free of antibiotics. 16S rDNA-amplicon MiSeq sequencing assessment was conducted on cecal samples collected from kits at 66 days. Paired-end sequences were processed with QIIME software. We tested 2 different models for each trait including or not an individual random microbacterial effect (M) with a covariance structure proportional to the weighted UniFrac distance matrix; constructed from a normalized sequence variant table. For each trait we run 20 cross-validations, each splitting the data into training (80%) and validation (20%) sets, to compare predictive ability of both models (i.e. correlation between observed and predicted host trait records on validation sets). Fitting M in models did not improve predictive accuracies of G but for feed efficiency traits FI and FCR they increased from 0.79 to 0.81 and from 0.61 to 0.67, respectively.

Predicting a path to increased genetic gain in wheat using Artificial Intelligence

Dr Kai Voss-Fels

Biography:

Dr Kai Voss-Fels is a Senior Research Fellow at The University of Queensland in Brisbane focusing on plant breeding and quantitative genetics.

Kai is interested in developing and implementing new genomics-assisted breeding approaches to improve yield, quality and resistances in major crops, such as wheat, barley, sugarcane, rapeseed and chickpea, with a particular focus on the integration of quantitative genetics, genomics and computational approaches. This also includes the study of GxE interaction and non-additive gene action, which represent a source of non-linearity observed in crop breeding, and thereby a key challenge for increasing genetic gain in crops.

The rate of genetic gain in wheat must be doubled over the next 2-3 decades to secure global food supply. Production trends worldwide suggest that current breeding strategies need to be improved in order to achieve this goal. One of the most critical steps in a wheat breeding program is to select parents for targeted crossing but strategies for decision support are not readily available. Conventionally, parents are selected based on their performance per se or their breeding value which limits the inference that can be made about the probability for maximising the number of favourable chromosome segments in the offspring of a given cross. This is particularly challenging when a combination of quantitative traits (yield) and mono- or oligogenic traits (e.g. disease resistance, quality) are considered simultaneously. This problem is well suited to evolutionary computation approaches, where algorithms inspired by biological evolution, such as reproduction, mutation, recombination, and selection are employed to find solutions to complex problems. In this ongoing study, we use evolutionary computing in a large commercial data set comprising more than 34,000 genotyped breeding lines which have been tested over multiple years and locations across Australia. The algorithms predict optimal crosses that most efficiently stack complementary alleles for quality, disease resistance and yield. Generated offspring from those crosses is advanced through “speed breeding” which allows a rapid turnaround of several generations and breeding cycles per year. Finally, these approaches will be compared for efficiency with alternative breeding strategies, such as standard genomic selection and phenotypic selection.

Influence of rare variants on heritability estimates using whole genome sequence data

Mr Pierrick Wainschein¹, Professor Peter M. Visscher¹, Professor Jian Yang¹

¹*University Of Queensland, Brisbane, Australia*

Heritability, the proportion of phenotypic variance explained by genetic factors, can be estimated from pedigree data, but such estimates are uninformative with respect to the underlying genetic architecture. While approximately one-third to two-thirds of heritability is captured by common SNPs from known causal loci, it is not known whether the remaining heritability is due to the imperfect tagging of causal variants by common SNPs or other reasons. Here we show that pedigree heritability for height and body mass index appears to be fully recovered from whole-genome sequence data on 30,340 unrelated individuals of European ancestry. Partitioning variants by minor allele frequencies (MAF) and linkage-disequilibrium with variants nearby, we showed that much of the “missing heritability” is due to rare variants in particular the ones in low linkage-disequilibrium with its neighbouring variants. Moreover spurious associations arising from population stratification in rare variants or from local non-genetic effects not accounted for by using standard correction methods could lead to biased estimates. To address these issues we demonstrate that although the GREML method is sensitive to population-stratification, a stringent selection of samples based on the rare variants can correct it and leveraged the spatial information in the UKBiobank Exome data set and to that non-genetic effects on heritability estimate are negligible. Our results imply that the missing heritability of complex traits and disease is accounted for by rare variants, in particular those in regions of low LD, and that these estimates are robust in respect with various potential bias associated with rare variants.

Evaluating the accuracy of imputed whole-genome sequence data in admixed dairy cattle

Dr Yu Wang

Biography:

Yu Wang is a postdoctoral scientist in quantitative genetics and genomics at Livestock Improvement Cooperation (LIC), New Zealand. She holds a Ph.D. in animal breeding and genetics from the University of Hohenheim, Germany. Before she joined LIC, she worked as a post-doc researcher at Massey University on the “Better Breeding Values” project under Genomics Aotearoa, which involved further development and implementation of a pipeline that comprises whole genome sequence imputation, sequence level genome-wide association studies, followed by single-step prediction using sequence imputed variants. Her main research interests are quantitative genetics, genomic selection, genetic diversity and conservation genetics.

The imputation accuracy was evaluated when high-density SNP genotypes (~644K) from an admixed population of 165,364 New Zealand dairy cattle were imputed to whole-genome sequence (WGS). WGS data were available for 336 Holstein-Friesian (H), 174 Jersey (J) and 535 H×J crossbred animals, among which 603 were sequenced with read depth >10x. The raw WGS data were aligned to the ARS-UCD1.2 bovine reference genome and variant calling was performed using GATK. VQSR filtering and standard quality control processes were conducted, resulting in ~20 million variants across the genome. Either high-depth sequenced animals or all sequenced animals were used as the reference population. The following software were singly or jointly used for phasing: Beagle 4.1, LinkPhase 3 and Beagle 5.0. Imputation was then performed from the phased data using Beagle 5.0. The quality of the imputation on chromosome 5 was evaluated by comparing the average dosage R², or based on 248 imputed animals sequenced for validation. Our study demonstrated that using Beagle 5.0 for phasing and imputation achieved high accuracy (average dosage R²=0.905) via using all sequenced animals and the reliable variants determined by high-depth sequenced animals. The sequence data from 248 validation animals exhibited an error rate of 1.11%, and correlation between imputed and called variants of 98.8%. Beagle 4.1 pre-phasing using genotype likelihood as input brought marginal benefit, however, it may be beneficial when low-depth sequenced animals were included in the reference. The imputed dataset will be used for future genome-wide association studies for casual variant detection and genomic selection.

Quantification of the accuracy of polygenic scores in ancestry divergent populations

Ms Ying Wang

Biography:

Ying Wang is a current PhD candidate in the Program in Complex Trait Genomics, IMB, UQ, Australia. She is under the supervision of Drs. Loic Yengo and Peter M. Visscher. She has been working on, within and across populations complex traits and diseases prediction using summary statistics from large-scale GWAS.

Polygenic scores (PGS) have been widely used to predict complex traits and risk of diseases using variants identified from genome-wide association studies (GWASs). To date, most GWASs have been conducted in populations of European ancestry, which limits the use of GWAS-derived PGS in non-European populations. Here, we develop a new theory to predict the relative accuracy (RA, the prediction accuracy relative to Europeans) of PGS across ancestry. We used simulations and real data from the UK Biobank to evaluate our results. We found across various simulation scenarios that the RA of PGS based on trait-associated SNPs can be predicted unbiasedly from modelling linkage disequilibrium (LD), minor allele frequencies (MAF), cross-population correlations of SNP effect sizes and heritability. Altogether, we find that LD and MAF differences between ancestries explain alone up to ~70% of the loss of prediction accuracy in African ancestry for traits like height and body mass index. Our results suggest that common genetic variation controlling complex traits and diseases is mostly shared across continents.

Integrating genetic and environmental information to improve phenotype prediction for body mass index

Mr Huanwei Wang

Biography:

I'm Huanwei Wang, a PhD student in the University of Queensland under the supervision of Prof. Jian Yang and Dr. Allan McRae. I'm now working at Westlake University in Hangzhou, China, as a visiting student due to the COVID19 travel restriction to Australia. My research direction is quantitative genetics and I use state-of-art statistical methods to explore the big genotype-phenotype biobank data. More specifically, three PhD research projects are about genotype-by-environmental interaction, epistasis and genetic prediction.

Predicting human complex traits or diseases is key for personalized prevention, intervention and treatment. The accuracy of genetic predictors has been increasing in the last decade with the increase of training sample sizes and the improvement of statistical methods. However, the use of environmental information in the prediction has been less well studied, although increasing numbers of environmental measurements are available in the large biobank data such as the UK Biobank (UKB). Here we aim to integrate both genetic and environmental information to improve prediction performance using body mass index (BMI) as a model trait. We generated a genetic risk score (GRS) with weights estimated by the SBayesR method (ref 1) and an environmental risk score (ERS) with weights estimated by GSMR method (ref 2). These scores were combined to generate a genetic and environmental risk score (GERS). Using a 10-fold cross-validation for 348,501 unrelated individuals with 163 environmental factors and 1,317,930 HapMap 3 SNPs in the UKB, our preliminary results show the GERS can on average increase the prediction accuracy (R^2) for BMI from 0.136 to 0.195 in comparison with GRS across 10 replicates. Our results indicate the value of integrating both genetic and environmental information in predicting BMI, and our method is applicable to the phenotype prediction of other complex traits or diseases.

1. L. R. Lloyd-Jones et al., bioRxiv. 522961 (2019)
2. Z. Zhu et al., Nature Communications. 9, 224 (2018)

Leveraging genetically correlated traits improves the detection of susceptibility loci for endometrial cancer

Dr Xuemin (Patrick) Wang

Biography:

I received a PhD degree in genetics and plant breeding from The University of Queensland in May 2020. Fascinated by quantitative genetics and the availability of data in human genetics, in January 2020 I made a significant career change, moving from plant genetics to cancer genetics. Since then, I have been working on endometrial cancer as a postdoc. I aim to conduct high-impact research in the understanding of cancer etiology and outcome, and to contribute to the development of drugs by research to cure cancer.

Widespread pleiotropy among traits can be exploited for the detection of genomic loci associated with complex traits. We recently conducted the largest genome-wide association study (GWAS) for endometrial cancer (12,906 cases), identifying 16 genetic loci associated with disease risk. However, these risk loci together explained nearly a quarter of the portion of the familial relative risk attributable to common, readily-imputable variants. Another crucial way to the prevention and the reduction of mortalities of endometrial cancer is to identify and understand its risk factors. Many risk factors of endometrial cancer have been identified observational studies and Mendelian Randomisation analyses. In this study, we leveraged the availability of GWAS summary-level data of diverse traits from the UKB Biobank and other consortia to identify risk factors of endometrial cancer.

Use of WGS and novel genomic selection strategies to improve selection of age at puberty in tropically adapted beef heifers

Mrs Christie Warburton

Biography:

Christie Warburton is currently a PhD candidate at the Queensland Alliance for Agriculture and Food Innovation, University of Queensland. Her thesis focus is on improving the prediction accuracy of multi-breed genomic selection in tropically adapted cattle. Growing up in the Northern Australian beef industry, she is passionate about the industry and hopes her research will provide valuable outcomes to breeders.

Genomic prediction for age at puberty (AP) has currently been limited in tropically adapted beef heifers due to the low accuracy of prediction. The aim of this research was to investigate novel methods of pre-selecting whole genome sequence (WGS) single nucleotide polymorphisms (SNP) and alternative analysis methodologies to determine if prediction accuracy for AP can be improved.

Genotypes and phenotypes were obtained from two research herds, the Northern Breeding Project research herd from the Beef Cooperative Research Centre for Beef Genetic Technologies (Beef CRC), and the Queensland Smart Futures (SMF) dataset. Genotypes were imputed to 23 million whole genome sequence variants and eight strategies were used to pre-select SNP from genome wide association study (GWAS) results. Pre-selected SNP were included in three analysis models, GBLUP with a single genomic relationship matrix (GRM), GBLUP multi-genomic relationship matrix (MGRM) and BayesR. Each analysis was conducted within three commercially available marker panels, 6K, 50K and 800K.

Prediction accuracies for AP were highest in BayesR (~0.5) analyses, and the addition of pre-selected WGS SNP had little effect on the BayesR estimates. The inclusion of WGS SNP pre-selected using a meta, conditional or joint (COJO) analysis by chromosome fitted in a MGRM model, had the highest prediction accuracies in the GBLUP analyses, across all marker panels. The prediction accuracy for the 6K marker panel using this methodology (0.42) was improved to the same prediction accuracy as the 800K control analysis (0.42). While more phenotypes are required, these results are promising and warrant further investigation.

Genotype by environment interactions and macro-environmental sensitivity in Australian Sheep

Mr Dominic Waters

Biography:

Dominic Waters is a PhD student at the University of New England. He completed his Honours degree in 2020, where he investigated genotype by environment interactions and genetic control of robustness in Australian sheep. His current PhD project aims to use genomic information to better understand the mechanisms of genotype by environment interactions in both livestock and plants.

The objective of this study was to explore genotype by environment interactions (GxE) and macro-environmental sensitivity in a multi-breed sheep population. Post-weaning weight (PWWT), post-weaning growth rate (PWGR), post-weaning scanned fat depth (PCF), carcass weight (CWT) and intra-muscular fat (IMF) were investigated. The number of sheep analysed ranged from 28,860 for PWWT to 14,969 for IMF. Environment was defined by the best linear unbiased estimation (BLUE) of contemporary group effects in an animal model, using PWGR as the response variable. Each trait was analysed using a trivariate mixed effects models which considered performance in low, average and high growth environments as separate traits. Reaction norms were also used for each trait to estimate macro-environmental sensitivity as well as provide an alternative method for estimation of GxE.

In the trivariate approach, genetic correlations were significantly different from unity for all traits across all environments except IMF. Correlations ranged from 0.62 to 0.72 for GR, PCF and CWT, 0.77 to 0.85 for PWWT and 0.90 to 0.99 for IMF. Scaling effects were only significant for PCF.

Overall, genetic correlations and scaling effects estimated using reaction norms were comparable to those estimated using multi-trait analysis. Genetic variation in slope was significant for all traits except IMF. Quadratic polynomials were significant for PWWT and GR only, indicating that sensitivity for these traits is dependent on the level of the growth environment.

This study highlights the potential for breeding sheep that are more robust to different environments.

Shared genetic etiology between obsessive-compulsive disorder, obsessive-compulsive symptoms in the population, and insulin signaling

Ms Joanna Widomska

Biography:

Joanna is a PhD Candidate at the Donders Center for Brain, Cognition and Behaviour, Nijmegen, The Netherlands. Her research focuses on genetic underpinnings underlying neuropsychiatric disorders, mainly Obsessive-Compulsive Disorder and Tourette Syndrome. She analyses and integrates various genomics data to provide insights into molecular aetiology of these comorbid disorders.

Obsessive-compulsive symptoms (OCS) in the population have been linked to obsessive-compulsive disorder (OCD) in genetic and epidemiological studies. Insulin signaling has been implicated in OCD. We extend previous work by assessing genetic overlap between OCD, population-based OCS, and central nervous system (CNS) and peripheral insulin signaling. We conducted genome-wide association studies (GWASs) in the population-based Philadelphia Neurodevelopmental Cohort (PNC, 650 children and adolescents) of the total OCS score and six OCS factors from an exploratory factor analysis of 22 questions. Subsequently, we performed polygenic risk score (PRS)-based analysis to assess shared genetic etiologies between clinical OCD (GWAS data from the Psychiatric Genomics Consortium), the total OCS score and OCS factors. We then performed gene-set analyses with a set of OCD-linked genes centered around CNS insulin-regulated synaptic function and PRS-based analyses for five peripheral insulin signaling-related traits. For validation purposes, we explored data from the independent Spit for Science population cohort (5047 children and adolescents). In the PNC, we found a shared genetic etiology between OCD and 'guilty taboo thoughts'. In the Spit for Science cohort, we additionally observed genetic sharing between OCD and 'contamination/cleaning', and 'symmetry/counting/ordering'. The CNS insulin-linked gene-set associated with 'symmetry/counting/ordering' in the PNC. Further, we identified genetic sharing between peripheral insulin signaling-related traits: type 2 diabetes with 'aggressive taboo thoughts', and levels of fasting insulin and 2 h glucose with OCD. In conclusion, OCD, OCS in the population and insulin-related traits share genetic risk factors, indicating a common etiological mechanism underlying somatic and psychiatric disorders.

The evolution of the genetic architecture of traits under genomic selection

Dr Yvonne Wientjes

Biography:

I obtained my PhD degree from Wageningen University and Research (WUR), where I investigated the potential to combine populations for Genomic Selection using both simulations and theoretical derivations. During my postdoc, I investigated how to use Genomic Selection to improve crossbred performance by selecting purebred animals. Currently, I'm working as a researcher at WUR and I'm investigating the long-term effects of genomic selection on the genetic architecture of traits. The main aim is to answer the question whether genomic selection can be applied for many generations without depleting genetic variance. This research is funded by a prestigious personal grant.

With non-additive gene action, the average effect of a causal allele depends on its allele frequency and the allele frequencies at interacting loci. Since drift and selection change allele frequencies, they also change the average effects of alleles. Genomic selection has likely accelerated this process, because it is more accurate and focusses more on genes of large effect. We aimed to quantify the changes in genetic architecture and genetic variation under genomic vs. traditional selection. We simulated a population for 50 generations with selection for a single trait, with varying levels of additive, dominance and epistatic effects. Results show that genomic selection always outcompeted other selection strategies for short-term gain. For long-term gain, mass selection was close to genomic selection, and even outcompeted it when epistasis was present. Loss in additive genetic variance was comparable across modes of gene action, and around 75% for genomic selection, slightly less for pedigree BLUP selection, and only 50% for mass selection. Over 50 generations of genomic selection, the correlation of average effects with average effects in generation 0 was 1 with additive gene action, 0.85 when dominance was also present, and 0.5 when also epistatic effects were present. This correlation decreased only slightly faster with genomic than with pedigree BLUP selection, but considerably faster than with mass selection. We also investigated the genomic changes due to different selection methods. We conclude that the presence of non-additive gene action and method of selection have a large impact on the long-term effects of selection.

The Genetic Basis of Thermal Tolerance in a Multi-Parental Population of Fruit Flies

Ms Patricka Williams-Simon, Ronel Ghidey¹, Enoch Ng'oma¹, Troy Zars¹, Elizabeth King¹

¹University of Missouri

Biography:

My name is Patricka Williams-Simon. I am a Black female who is also a sixth-year Ph.D candidate at the University of Missouri in Dr. Elizabeth King Lab.

*Generally, I am interested in the mechanistic genetic basis that control complex traits. For my research, I aim to dissect the genetic basis of learning, memory, and thermal tolerance in a Multi-Parental Population of fruit flies, *D. melanogaster*. In the near future, I plan on continuing my training as a postdoc where I gain the skills to become an independent scientist. My major career goal is to become a scientific researcher at an RO1 institution, where I can lead and a diverse group of people to think critically about both research and the broader impacts of doing science in an inclusive environment.*

Thermal tolerance is a complex trait that is a fundamental survival skill in many species. For example, everyday tasks such as foraging, finding a mate, and escaping predation, are highly dependent on how well an organism can tolerate extreme temperatures. Understanding the natural variants of the genes that control this trait is of high importance if we want to better comprehend how this trait evolves in natural populations. Here, I take a quantitative genetics approach to dissect the genetic basis of thermal tolerance using *Drosophila* Synthetic Population Resource (DSPR). This multi-parental population consists of approximately 1,800 Recombinant Inbred Lines (RILs), which allows for high-resolution genome wide scans, and the identification of loci contributing to naturally occurring genetic variation. Using a highly sensitive apparatus known as the “heat box” we presented the flies with a constant 41°C for a total of 9 ½ mins. and recorded the time it takes for an individual to become incapacitated as our measure of thermal tolerance. We measured 741 RILs and found that thermal tolerance ranges from, 24.8 sec – 361 sec, which is ~14-fold difference between the lowest to the highest tolerant RILs. We then performed a genome scan and identified several loci (QTL) affecting thermal tolerance, including one large effect locus. We performed fine mapping, including RNA-seq to identify differentially expressed candidate genes within these QTL that affect thermal tolerance. Future work will aim to validate the function of these genes

Predicting key GWAS outcomes under a point-normal polygenic model

Ms Tian Wu

Biography:

Tian Wu is a PhD candidate from the Department of Psychiatry, the University of Hong Kong. Her research interests include the polygenic risk score construction and accuracy improvement.

Genome-wide association studies (GWAS) have successfully identified thousands of common single nucleotide polymorphisms (SNPs) associated with complex disorders and related traits. It has been observed that for each complex phenotype, associated SNPs become detectable as the sample size reaches a certain minimum, above which the number of significantly associated SNPs increases with increasing sample size in an approximately linear fashion. However, a detailed modelling of the entire relationship between the number of significant SNPs and sample size has not been performed. Here, we consider a point-normal polygenic model, to predict the number of independent significant SNPs given GWAS sample size, key parameters of genetic architecture of the phenotype (e.g., SNP-heritability, polygenicity and disease prevalence) and study design (e.g., case-control ratio of the GWAS, or for meta-analysis, the case-control ratio of each sub-study). Under the same assumptions, we also predicted several important outcome indices of the GWAS, including the overall phenotypic variance explained by significant SNPs, and the predictive accuracy of polygenic scores that weight SNPs by the effect size estimates from GWAS, after shrinkage by various methods. Calculations were performed from analytically derived formulae, and were validated by simulations. We compared the predictions of our method to the observed results of GWAS on height and BMI, and found that they were in agreement. Thus, our method could be a useful tool for the design of GWAS and for predicting the future behaviour of GWAS as sample sizes increase further.

Promoter-anchored chromatin interactions predicted from genetic analysis of epigenomic data

Dr Yang Wu

Biography:

Dr. Yang Wu is postdoc research fellow at the University of Queensland. Yang's research focuses on understanding the genetic regulatory mechanisms underpinning complex trait variation in humans. Yang obtained his PhD in statistical genetics from the University of Queensland and was recognized as the Australian research "Rising stars" by The Australian newspaper 2020.

Promoter-anchored chromatin interactions (PAIs) play a pivotal role in transcriptional regulation. Current high-throughput technologies for detecting PAIs, such as promoter capture Hi-C, are not scalable to large cohorts. Here, we present an analytical approach that uses summary-level data from cohort-based DNA methylation (DNAm) quantitative trait locus (mQTL) studies to predict PAIs. Using mQTL data from human peripheral blood (n=1,980), we predicted 34,797 PAIs which showed strong overlap with the chromatin contacts identified by previous experimental assays. The promoter-interacting DNAm sites were enriched in enhancers or near expression QTLs. Genes whose promoters were involved in PAIs were more actively expressed, and gene pairs with promoter-promoter interactions were enriched for co-expression. Integration of the predicted PAIs with GWAS data highlighted interactions among 601 DNAm sites associated with 15 complex traits. This study demonstrates the use of mQTL data to predict PAIs and provides insights into the role of PAIs in complex trait variation.

A causality perspective of genomic breed composition for composite animals

Dr Xiaolin Wu

Biography:

I joined the Council of Dairy cattle Breeding in 2020 as the product development manager, holding an adjunct associate professor position at the University of Wisconsin – Madison (UWM). I previously worked as a computational geneticist at UWM, Principal Scientist at the Bayer CropScience, and director of Biostatistics and Bioinformatics at the Neogen GeneSeek. I published over 140 papers and six books and book chapters. My research experience covered molecular genetics, quantitative genetics, and bioinformatics in the past twenty years. Topics of interest in recent years include and are not limited to genomic prediction, genomic breed composition, and parentage analysis.

Genomic breed composition (GBC) of an individual animal refers to the partition of its genome according to the inheritance from its ancestors or ancestral breeds. Various statistical approaches have been proposed to estimate GBC in animals, but the interpretations of estimates have varied with these methods. In the present study, we proposed a causality interpretation of GBC based on path analysis. We applied this method to estimating GBC in two composite breeds of beef cattle, namely Brangus and Beefmaster. The path analysis decomposed the relationships between the ancestors and the composite animals into direct and indirect path effects, and GBC was measured by the relative ratio of the coefficients of direct (D-GBC) and combined (C-GBC) determination from each ancestral breed to the progeny, respectively. Estimated GBC varied only slightly between different genotyping platforms and between the three SNP panels (1K, 5K, and 10K). In the Brangus cattle, because the two ancestral breeds had a very distant relationship, the estimated D-GBC and C-GBC were comparable to each other in the path analysis, and they corresponded roughly to the estimated GBC from the linear regression and the admixture model. In the Beefmaster, however, the strong relationship in allelic frequencies between Hereford and Shorthorn imposed a challenge for the linear regression and the admixture model to estimated GBC reliably. Instead, D-GBC by the path analysis included only direct ancestral effects, and it was robust to bias due to high genomic correlations between reference (ancestral) breeds.

421

The poor man's multi-trait solver

Dr Alencar Xavier

Biography:

My work on statistical genetics is focused on genomic-assisted breeding with emphasis on theoretical and computational aspects of data-driven plant breeding, such as modeling, prediction and selection using various sources of information. My research regards the development and implementation of new quantitative methods using mixed models, Bayesian methods and machine learning, along with high-performance computing. I work at Corteva Biostatistics and as an adjunct faculty at Purdue University.

Selections are performed on many correlated traits. However, the computation of multivariate models is prohibitive under scenarios with numerous observations, markers and traits. Pre-genomic methodologies were developed in the 80's to deal with large pedigree sets, but those have not been evaluated with genomic data. The two operations involved in the computation of multivariate genetic model are: estimation of (1) regression coefficients and (2) covariance components. This study presents an iterative solver that couples a multivariate Gauss-Seidel algorithm for estimation of marker effects with simultaneous estimation of covariance components via Tilde-Hat method. Accuracy, bias, and computation time were compared to AI-REML using simulations.

Genome-wide analyses of behavioral traits biased by misreports and longitudinal changes

Mr Angli Xue

Biography:

Angli Xue is a PhD candidate at Program in Complex Trait Genomics (PCTG) group at University of Queensland (UQ), Australia. Before joining UQ, he received the bachelor in 2015 from Zhejiang University (ZJU), China. His research focuses on dissecting the genetic architecture of human complex diseases, performing meta-analysis and integrating multi-omics data to identify associated gene loci and gene regulating mechanisms. He is also interested in investigating how modifiable risk factors affect human health, such as tobacco smoking and alcohol consumption.

Genome-wide association studies (GWAS) have discovered numerous genetic variants associated with human behavioral traits. However, it is often under-recognized that misreports and longitudinal changes (MLC) in behavioral traits are non-random, causing severe biases in GWAS and follow-up analyses. Here, we demonstrated in the UK Biobank (n=455,607) that individuals with higher disease burden were more likely to reduce or misreport their alcohol consumption (AC) level, and that spurious AC GWAS signals caused by MLC were enriched in metabolic/cardiovascular traits and disorders. Almost all the previously reported negative estimates of genetic correlation between AC and common diseases became positive/non-significant after correcting for MLC. For example, the genetic correlation estimate between AC and hypertensive disease was -0.050 (s.e. = 0.023), but turned to 0.071 (s.e. = 0.024) after correcting for MLC. Also, the previously reported U-shaped association between AC and common diseases was likely due to the enrichment of underreported and/or reduced AC in the reference group (i.e. usually modest drinking group). Furthermore, the estimate of genetic correlation between AC and alcohol use disorder increased from 0.565 (s.e. = 0.046) to 0.721 (s.e. = 0.051) after correcting for MLC. We also observed biases due to disease ascertainment for AC in other data sets, and for other behavioral traits (e.g., smoking, physical activities) in the UK Biobank. Our findings provide a plausible explanation of the long-standing controversy about the effects of AC on health outcomes and caution for future analyses of self-reported behavioral traits in biobank data.

Improved genomic prediction of clonal performance in sugarcane by exploiting dominance and epistasis effects

Ms Seema Yadav

Biography:

Seema Yadav is a PhD candidate at Queensland Alliance of Agriculture and Food Innovation (QAAFI), the University of Queensland. She is based at the Centre for Animal Science and working with Prof Ben Hayes and Dr. Kai Voss-Fels. Her PhD project is focused on implementing genomic selection to accelerate genetic gains in Australian sugarcane breeding programs. Before joining the UQ, she was working as an international consultant with Quantitative Genetics cluster at International rice research institute, Philippines. She has double master degrees in Mathematics and Statistics.

The rate of genetic gain is slow in sugarcane compared to other crops. One reason might be a significant contribution of non-additive effects in the expression of complex traits. The availability of dense marker information provides the opportunity to exploit non-additive effects in genomic selection (GS) for improved clonal predictions which can lead to accurate identification of varieties for further testing and subsequent release to farmers. In this study, we assessed a series of models extending the GBLUP approach that accounted for additive, dominance and additive-additive epistatic effects with or without heterozygosity for cane per hectare (TCH), commercial cane sugar (CCS), and Fibre content of 3000 genotyped clones from Australia breeding programs. Alternatively, the reproducible kernel Hilbert space (RKHS) model was used to capture the epistasis effects. We considered two forward prediction scenarios to investigate the effect of training size and year by genotype interaction on prediction accuracy. Our results highlight the importance of non-additive variance for complex traits in sugarcane, especially additive by additive variation. Average heterozygosity had a significant effect on TCH, suggesting that directional dominance is important for this trait. Prediction accuracies improved by at least 17% for TCH when non-additive and heterozygosity effects were included in the model, whereas only 1% improvement was found for CCS. For Fibre, an additive model gave the best predictions. The extended-GBLUP outperformed RKHS, likely due to the addition of heterozygosity effect. Genomics-based sugarcane breeding is likely going to benefit from exploiting non-additive effects, especially in designing crossing schemes. Keywords: Dominance, Epistasis, GBLUP, Genomic Prediction, Heterozygosity, RKHS, Sugarcane

Identifying potential risk factors for endometriosis

Ms Fei Yang

Biography:

Fei Yang is a third-year PhD candidate in Prof Grant Montgomery's lab within Genetics and Genomics Division at The Institute for Molecular Bioscience, The University of Queensland. Her research focuses on identifying the genetic risk factors for endometriosis and investigating genetic influence on transcript regulation in endometrium including transcript isoform usage and alternative splicing. She earned a master degree in clinic veterinary medicine in 2018 and a bachelor degree in veterinary medicine in 2015 from Jiangxi Agricultural University in China.

Endometriosis is a complex disease for which multiple risk factors have been reported as associated in epidemiological studies, including reproductive phenotypes, lifestyle related factors and cancer. Evidence of genetic associations between endometriosis and potential risk factors can help understand endometriosis pathogenesis, identify shared target genes and pathways. The aim of this study was to estimate the genetic correlation between endometriosis and potential risk factors and identify causal relationships and shared risk loci.

Existing epidemiological evidence suggests a possible relationship between endometriosis and melanoma. Studies report that a history of endometriosis is associated with increased risk of melanoma and both diseases occur more frequently in younger aged women. Using GWAS summary data from a recent endometriosis meta-analysis (n=198,927) and melanoma meta-analysis (n=31,353) we performed LDSC and found no evidence of a genetic correlation. We did however identify a significant causal relationship between the two diseases using Generalized Summary-data-based Mendelian Randomization ($p=2.1e-05$). Further investigation into potential confounding factors affecting this causal relationship like sex effects, or other potential mechanisms like hormone regulation is required.

Endometrium is considered to be the source of cells initiating endometrial lesions. Correlation of endometrial eQTLs with other tissues revealed that genetic effects on gene expression in endometrium were highly correlated with reproductive and digestive tissues, providing evidence that genetic regulation of gene expression is shared between biologically similar tissues and cell types. Analysis of reported risk factors related to these correlated tissues including age of menarche and menstrual cycle length is ongoing.

Genetic analysis of blood cell type proportions in neurological diseases

Dr Yuanhao Yang

Biography:

Yuanhao Yang is a postdoctoral research fellow in the Cognitive Health Genomics, at the Mater Research Institute (MRI). Before starting research at MRI, Yuanhao completed his PhD in statistical genetics and epidemiology at Queensland University of Technology in 2016, and then worked as a postdoctoral research fellow in the Institute for Molecular Bioscience at the University of Queensland until Sep 2018. His research focuses on quantitative genetics and epidemiology of complex traits.

Associations between blood cell type proportions (CTPs) and neurological diseases such as Parkinson's disease (PD) have been reported in recent DNA methylation analyses, but whether these associations have a genetic basis is unclear. Here, we report genetic correlations and the results of Mendelian randomisation analyses between six neurological diseases (PD, Alzheimer's disease [AD], ALS, multiple sclerosis [MS], migraine and stroke) and 36 human blood cell traits, using genome-wide summary statistics. We found limited evidence for genetic correlations between neurological diseases and blood cell traits, with the exception of migraine and platelet count. In contrast, there was strong evidence for causal effects of AD, MS and migraine on multiple blood cell traits, after removal of pleiotropic SNPs. We also observed an interesting bidirectional association between MS and lymphocyte count in which MS was associated with increased lymphocyte count, but high lymphocyte count was associated with decreased susceptibility to MS. Analyses considering blood cell traits as exposures for neurological diseases yielded fewer and weaker associations, with two exceptions: increased basophil percentage was protective for MS and higher red blood cell count was a risk factor for ALS. PD was notable for exhibiting no Bonferroni-significant associations with blood cell traits, despite a priori evidence for an association with blood CTPs. Our study provides novel insights into the genetic relationships between neurological diseases and blood cell traits, and suggest that for some neurological diseases, correction for differences in blood CTPs in DNA methylation analyses may dilute true disease signals.

Restricted diet drives autism-gut microbiota associations, which are weak compared to age, dietary and stool traits

Miss Chloe X Yap

Biography:

Chloe Yap is an MD-PhD-GCBusLead (Medical Leadership Program) student and Fulbright Future Scholar at MRI-UQ and IMB, University of Queensland. Her thesis is focused on systems genomics of autism, with the hope of contributing to biomarker discovery to improve the process of diagnosis. She is particularly interested in understanding the interplay between mental health and physical health through the lens of “big data” omics approaches, and to ultimately contribute to the translation of precision medicine initiatives as a clinician-scientist.

Background: There has been increasing excitement around the potential contribution of the gut microbiota to autism spectrum disorder (ASD) – a poorly-understood and heterogeneous neurodevelopmental condition characterised by social difficulties and repetitive and restrictive behaviours. However, the lack of well-powered studies and inconsistent correction for confounders means that the very foundations of this link are tenuous. **Methods:** Here, we present results from the largest metagenomics study of the autism stool microbiota to date (n=246, including 99 children on the spectrum), using samples from the Australian Autism Biobank and the Queensland Twin Adolescent Brain study. **Results:** We find negligible evidence for association of ASD diagnostic status with the gut microbiota. Instead, we find significant relationships with dietary diversity and propose a model of causality whereby ASD-associated restricted interests are associated with restricted diet, less-diverse microbial taxonomic diversity, and looser stool consistency. In contrast to ASD diagnostic status, we demonstrate that this metagenomics dataset is well-powered to detect associations with other traits, accounting for over 95% of variation in age, up to 75% for dietary traits, and 55% for stool consistency. **Conclusions:** Overall, our results suggest that changes to the microbiota in ASD may reflect consequences of behaviour and dietary preferences, and we caution against undue emphasis on the microbiota having an upstream role in ASD.

Enrichment of inbreeding depression across genomic annotations of the human genome

Dr Loic Yengo¹

¹University Of Queensland

Biography:

Loic Yengo is a statistical geneticist and team leader within the Program of Complex Traits Genomics (PCTG). Loic joined the University of Queensland in 2016 with a research background in applied mathematics, statistics, and molecular epidemiology. His current research focuses on understanding the genetic and phenotypic consequences of non-random mating (inbreeding and assortative mating) in human populations, and on developing analytical methods for genome-wide association studies (GWAS).

Offspring of related individuals often exhibit significant reduction in fitness-related traits, also termed as inbreeding depression (ID). ID is observed in many species, including humans, though its genetic and molecular basis remains elusive. Here, we develop a new method to quantify enrichment of ID within specific genomic annotations of the genome. We analysed the phenomes and genomes of ~350,000 unrelated participants of the UK Biobank, and found significant evidence across multiple traits that variants located within conserved regions in primates (~18x; $p=10^{-4}$), conserved regions in mammals (~16x; $p=1.3 \times 10^{-4}$), DNase-I hypersensitive sites (~5x; $p=1.7 \times 10^{-6}$) and specific chromatin marks (~2x, $p=5.8 \times 10^{-5}$) disproportionately contribute to ID. Importantly, we find that ~30% of the enrichment of ID is explained by low linkage disequilibrium (LD) within these genomic regions. Overall, our study provides new empirical evidence that the ID is due to widespread partially-recessive deleterious alleles within low LD regions of the human genome.

Whole-exome sequencing analyses identify risk genomic loci for Nasopharyngeal Carcinomas

Dr Yanni Zeng

Biography:

Dr Yanni Zeng is an associate professor from Zhongshan Medical School, Sun-yat Sen University(SYSU), Guangzhou, China. Yanni studied Biology in East China Normal University in Shanghai and KIZ, CAS in Kunming(2006-2013). This was followed by a PhD in psychiatric genetics(2013-2016), supervised by Prof. Andrew Mcintosh from the division of psychiatry and Prof. Chris Haley from IGMM, University of Edinburgh(UK). In 2016-2017 she works with Prof. Chris Haley's group as a postdoc exploring determination of DNA methylation and its impact on complex traits. In 2018 she joined SYSU. Her current research interest focuses in genetic, epigenetic and environmental determination of complex disease/traits.

Background:

Nasopharyngeal carcinoma (NPC) featured a high prevalence in southern China, southeast Asia and Northern Africa, yet our understanding of its genetic etiology is confined to common variants. Only 16.9% of genetic risk of NPC is attributable to the common variants, indicating a urgent need to study rare variants

Methods:

Exome-wide sequencing was conducted for blood samples of 2465 NPC patients and 2304 healthy controls recruited from south China and southeast Asia. This collection also includes 354 NPC patients with a family history. A SNP-set (Sequence) kernel association test (SKAT) was applied to localize aggregate genetic effects from germline rare variants in coding genes on NPC.

Results:

Seven genes were identified for consistently elevated frequency of rare mutations in NPC patients across samples as compared with healthy controls. A subset of these genes additionally showed a higher frequency of mutation in subgroups of patients with family history. The exon-based SKAT analysis combined with single variants association tests localized major contributing rare variants for the association observed at gene level. A proportion of these contributing rare variants showed distinguished minor allele frequency in Asia population as compared with European population. Functional pathways were also identified for increased frequency of risk alleles from both common and rare variants in NPC patients.

Conclusions:

This study identified important germline rare mutations, coding genes and molecular pathways contributing to the genetic etiology of NPC. We provided potential biomarkers for personalized NPC genetic risk evaluation and potential explanations of the regional differences of NPC prevalence.

Bayesian analysis of GWAS summary data reveals differential signatures of natural selection across human complex traits and functional genomic categories

Dr Jian Zeng

Biography:

Dr Jian Zeng is a statistical geneticist and NHMRC Emerging Leadership Fellow at the Institute for Molecular Bioscience (IMB) at the University of Queensland (UQ). He received his PhD in animal breeding and genetics at Iowa State University and joined the Program in Complex Trait Genomics (PCTG) at UQ in 2016. Bringing the skills used in genomic prediction for livestock breeding, his research has been focusing on developing methodology and software tools for understanding the genetic architecture of complex traits and for polygenic risk prediction of common diseases in humans. He was awarded a NHMRC Investigator Grant (EL1) in 2019.

Understanding how natural selection has shaped genetic architecture of complex traits is of importance in medical and evolutionary genetics. Bayesian methods have been developed using individual-level GWAS data to estimate multiple genetic architecture parameters including selection signature. Here, we present a method (SBayesS) that only requires GWAS summary statistics. We analysed data for 155 complex traits (n=27k-547k) and projected the estimates onto those obtained from evolutionary simulations. We estimate that, on average across traits, about 1% of human genome sequence are mutational targets with a mean selection coefficient of ~ 0.001 . Our results suggest relatively strong selection on genetic variants for disease and cognitive traits and relatively small number of mutational targets for disease and reproductive traits. SBayesS analyses incorporating functional annotations reveal that selection signatures vary across genomic regions, with both the number of associated variants and the magnitude of effect sizes enriched in coding and conserved regions.

Empirical variance component regression for sequence-function relationships

Dr Juannan Zhou¹, Dr Sze Wong¹, Dr Wei-Chia Chen¹, Dr Justin B. Kinney¹, Dr Adrian Krainer¹, Dr David M. McCandlish¹

¹*Cold Spring Harbor Laboratory, Ithaca, USA*

Contemporary high-throughput mutagenesis experiments have provided an unprecedented view of the complex patterns of genetic interaction that occur between multiple mutations within a single protein or regulatory element. By simultaneously measuring the effects of thousands of combinations of mutations, these experiments have revealed that the genotype-phenotype relationship typically contains genetic interactions not only between pairs of sites, but also higher-order interactions between larger numbers of sites. Here we provide a method of analyzing data from these experiments using Gaussian process regression with an empirical Bayes prior. The key insight is that many previously proposed methods can be recast as members of a family of Gaussian process regression models with hyperparameters corresponding to the expected fraction of variance due to each order of genetic variance. By analyzing the distance correlation function of the observed data, we can extract point estimates of these variance components. Based on these point estimates, we then construct a prior over all possible sequence function relationships, where the prior is concentrated on models with a similar correlation structure to that observed in the data. We apply this method to analyze high-throughput measurements for the splicing efficiency of human 5' pre-mRNA splice sites. Conducting low-throughput validation experiments to assess the quality of our model's predictions, we show we can accurately predict the splicing efficiency of previously unmeasured 5' splice site genotypes.

Role of Testosterone in the Sex-specific Genetic Architecture of Complex traits

Dr Zhihong Zhu

Biography:

Zhihong Zhu was awarded his PhD in Statistical Genetics from Zhejiang University, China, 2012, with primary research in the plant genomics. He then switched focus to study human genetics. He moved to Brisbane, Australia, 2013, being employed as a Postdoctoral Fellow in the Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, University of Queensland, before moving to the Program in Complex Trait Genomics, Institute for Molecular Bioscience, University of Queensland in 2016. He moved to Aarhus University in 2020. His primary research interests are development and application of statistical methodology to understand genetic mechanism of human traits and diseases.

Sex-differences are found in most human complex traits. The genetic basis for these differences is largely understudied, with most genome-wide association studies treating sex as a nuisance parameter. Using an unrelated subset of the UKBiobank, we investigated the genetic contribution to sex-differences in 68 phenotypes using sex-specific GWAS on up to 166,736 males and 193,828 females. Estimates of heritability using LD Score regression did not differ between the sexes, except for both systolic and diastolic blood pressures having lower heritability in males due to medication use. The genetic correlation between males and females was significantly smaller than one for 28 (35%) traits. Most genetic correlations remained large (>0.84), with the exception of testosterone levels that had a distinct genetic architecture across the sexes ($r_g = 0.03$, s.e. 0.05). Significant autosomal SNP-by-sex interactions were identified in 20 traits, with 78 in testosterone and 40 in other traits. Additionally, 11 SNP-by-sex interactions were identified on the X chromosome (6 for testosterone). The genetic correlation between testosterone and 26 other traits was significantly different across the sexes. BMI had a negative genetic correlation with testosterone in males (-0.28 , $SE=0.03$), but a positive genetic correlation in females (0.17 , $SE=0.03$). These results demonstrate the role of testosterone in human sex-differences extends beyond simple mean differences in the sexes.



ICQG·6

Abstract book sponsored by



**ONE FAMILY.
ONE PURPOSE.**