7th International Conference of Quantitative Genetics

www.icqg2024.ista.ac.at

Institute of Science and Technology Austria

TABLE OF CONTENTS

1 ICGQ7

ICQG7 will bring together researchers with a focus on theory and methodological development. We aim to represent the full range of applications of quantitative genetics – from plants, crops and trees to livestock to humans including common disease, to wild populations and laboratory model species.

The conference provides a forum to highlight novel, principled statistical approaches which may be relevant to the problems faced across a range of applications. In the genomics era the integration of quantitative genetics theory across species applications is converging, and new quantitative trait phenotypes such as single cell gene expression are being studied. This represents an exciting time for understanding and translating the contribution of genetic variation of quantitative traits.

We aim to focus on the presentation and discussion of state-of-the-art results, theoretical developments and new methodologies, and we will prioritise unpublished research. There will be time for discussion after each talk and to foster informal interactions among scientists of all career stages. We will seek a conference program includes a diverse range of speakers and discussion leaders from institutions and organizations worldwide.

The conference is five days long and held in a single location, with the hope of creating lasting collaborations. In addition to invited talks, speakers will be selected from submitted abstracts, and the conference has designated time for sessions from individuals of all career stages.

2 Program

Monday, July 22, 2024

	Audimax	Arcaded Courtyard	Big HS	HS 33
08:20-10:00	Opening remarks **Bogdan Pasaniuc** Polygenic risk scores for precision medicine: promises and challenges **Chiara Sabatti** Searching for causal variants in polygenic traits			
10:00-10:30		Coffee break		
10:30-12:00	**Thomas Hansen** The structure of evolutionary quantitative genetics **Jeff Endelman** Directional Dominance in Polyploids: Trait Analysis and Mate Selection			
14:00-16:00		**Session 1**		

16:00-16:20	**Kerry Geiler-Samerotte** The Genotype-Phenotype-Phenotype-Phenotype Map		**Torsten Pook** Strategies to improve selection compared to selection based on estimated breeding values	**Frank Albert** Genetic variation in protein degradation
16:20-16:40	**Laura Luebbert** Efficient and accurate detection of viral sequences at single-cell resolution reveals novel viruses perturbing host gene expression		**Zhiwu Zhang** Emerging Marker Assisted Selection and Genomic Selection	**Jose Aquicira Hernandez** Scalable single-cell models for robust cell-state-dependent eQTL mapping
16:40-17:00	**Joshua Popp** Dynamic genetic regulation of gene expression in heterogeneous differentiating cultures		**Tobias Niehoff** Exploiting progeny variances for selection decisions improves genetic gain and variance in genomic breeding programs	**Daniel Kaptijn** Genetic regulation of single-cell personal gene correlations (co-eQTLs) is highly enriched for GWAS variants
17:00-17:20	**Natalia Ruzickova** Interpretable genomic predictions via effect propagation in gene regulatory networks		**Marcio Resende** Integrating single kernel Phenomic Selection with Genomic Selection: Applications in Corn Breeding	**Guillaume Ramstein** Prediction of variant effects by foundation AI models: in vivo validation at nucleotide and haplotype resolution in plant populations
17:30-19:00		Welcome reception		

Tuesday, July 23, 2024

	Audimax	**Arcaded Courtyard**	**Big HS**	**HS 33**
08:20-10:00	**Piter Bijma** Dark Genes: How			

	transmission of infections boosts heritable variation and response to selection **Jason Wolf** Genetic analysis of intrafamilial interactions			
10:00-10:30		Coffee break		
10:30-12:00	**Susan Johnston** The causes and consequences of sex differences in recombination rates **Zach Lippman** Synergy among cryptic variants in a plant regulatory network drive nonlinear phenotypic effects			
14:00-16:00		**Session 2**		
16:00-16:20	**Peter Keightley** The impact of spontaneous mutation accumulation on quantitative variation in a mammalian species		**Jian Zeng** Genome-wide fine-mapping improves identification of causal variants	**Xiaoning Zhu** Deciphering the genetic mechanisms of complex traits in chicken AIL populations using multi-omics data
16:20-16:40	**Gregor Gorjanc** Quantitative genetic modelling of diverse populations using ancestral recombination graphs		**Yakov Tsepilov** GentroPy package for ancestry specific systematic fine-mapping of GWAS data, colocalization and drug targets prediction	**Hao Tong** Leveraging interactome and transcriptome to enhance genomic prediction in plant breeding
16:40-17:00	**Andrea Doeschl-Wilson** Estimating and		**Yixuan He** Multi-trait and multi-ancestry polygenic risk	**Xiandong Ding** An efficient analysis method for integrating

	dissecting host genetic variation underlying infectious disease transmission – methodology and empirical evidence		score approach improves genetic discovery and risk prediction of respiratory diseases	multiple omics data based on deep learning
17:00-17:20	**Owen Powell** Improving the prediction of non-additive effects with hierarchical genomic prediction models		**Lin Qing** Multi-ancestry genome-wide association study meta-analysis deciphers the genetic architecture of male fertility in pig	**Julia Sidorenko** Reconciling linkage and association studies of complex traits
17:30-18:15	**Naomi Wray** Quantitative Genetics of Psychiatric Disorders			

Wednesday, July 24, 2024

	Audimax	**Arcaded Courtyard**	**Big HS**	**HS 33**
08:20-10:00	**Jacqueline Sztepanacz** Estimating genetic variation and selection in high-dimensional data **Guy Sella** A population genetic interpretation of genome-wide association studies in humans			
10:00-10:30		Coffee break		
10:30-12:45	**Jack Dekkers** Implementation of a Mechanistic Growth Model for Pigs into Bayesian Methods for Genomic Prediction and GWAS **Christine Baes** Quantitative Genetic Solutions for Optimizing Livestock Sustainability: Innovations, Genomic Applications, and Future Directions **Gustavo de los Campos** Improving cross-ancestry PGS Prediction through			

	Transfer Learning using Informative Penalized Regressions and Bayesian Mixture Models			
from 17:50	Vienna Panorama Sightseeing Tour Meeting point: Vienna State Opera			

Thursday, July 25, 2024

	Audimax	**Arcaded Courtyard**	**Big HS**	**HS 33**
08:20-10:00	**Julien Ayroles** Quantitative Genetic Solutions for Optimizing Livestock Sustainability: Innovations, Genomic Applications, and Future Directions **Michael Goddard** Identifying causal variants for histone modification			
10:00-10:30		Coffee break		
10:30-12:00	**Po-Ru Loh** Influences of genomic structural variation on human complex traits **Joelle Mbatchou** Using large language models for rare variant association testing in large-scale biobanks			
14:00-16:00		**Session 3**		
16:00-16:20	**Al Depope** Light-speed whole		**Leke Victor Aiyesa** A new unrestricted	**Anna Hewett** Inbreeding

	genome association testing and prediction via Approximate Message Passing		assessment toward utilizing individual plant phenotypes and genotypes for breeding	depression throughout the growth period of wild Swiss barn owls
16:20-16:40	**Xia Shen** Modelling the genetic architecture of complex traits via stratified high-definition likelihood		**Michelle Stitzer** Transposable element abundance subtly contributes to lower fitness in maize	**Richard Bernstein** Effective population size in honeybees from pedigree and SNP data
16:40-17:00	**Matias Schrauf** Altered Prior Mean of Allelic Effects: An Approach for Adequately Considering Gene Edited Variants within Genomic Predictions		**Neda Rahnamae** Can hybridization allow the emergence of a Super-Genotype in Arabis floodplain species?	**Elizabeth Mittell** The effects of a missing fraction on selection in adult size traits in a wild population
17:00-17:20	**Dom Waters** Reduced rank factor analytic models for capturing genotype by environment interactions in livestock		**Yvonne Wientjes** Changes in allele frequency and GWAS results across years in two pig populations under selection	**Kelly Swarts** Isolating adaptive variation from natural forest trees
17:30-18:15	**Ed Buckler** From Climate Change to AI: Improving Agriculture by Learning from Global Biological Diversity			
from 18:15		Evening reception		

Friday, July 26, 2024

	Audimax	**Arcaded Courtyard**	**Big HS**	**HS 33**
08:20-10:00	**Loic Yengo** Convergence of heritability			

	estimates from orthogonal experimental designs **Augustine Kong** Participation bias in genetic studies and estimate adjustments			
10:00-10:30		Coffee break		
10:30-12:00	**Lingzhao Fang** The Farm Animal Genotype-Tissue Expression (FarmGTEx) Project for Advancing Agriculture and Biomedicine **Amelie Baud** The Hologenome 2.0			
14:00-16:00		**Session 4**		
16:00-16:20	**Lars Rönnegård** Warning: Selection for decreased variability in milk yield may lead to asocial cows!		**Natalia Leite** Marker Effect P-Value for Large Genotype Populations with the Algorithm for Proven and Young	**Tom Druet** Unravelling the genetic architecture of height and muscular development traits in Belgian Blue cattle and using it for genomic prediction
16:20-16:40	**Christie Warburton** IBS versus IBD – new insights from whole genome sequence data		**Anthony Long** X-QTL mapping using multi-parent synthetic populations is powerful and efficient conditional on experimental design	**Martin Johnsson** The structure of potentially functional genetic variation in cattle

16:40-17:00	**Ilse Krätschmer** Direct, indirect and epigenetic effects in families		**Teresa McGee** Increasing power in association mapping with genetic replicates by recognizing variance heterogeneity and exploring implications of near zero-variance	**Emre Karaman** Incorporating prior biological information into genomic predictions: An example from mastitis in Danish Jersey and Nordic Red cattle
17:00-17:20	**Thomas Ellis** The Effect of Population Structure Correction on GWAS Before and After Random Mating		**Nick Machnik** Causal inference for multiple risk factors and diseases from genomics data	**Naveen Kadri** Detection of QTL for global recombination rate in Fleckvieh cattle
17:30-18:15	Closing remarks and early career and talk prizes			

3 Invited Speakers

3.1 Ayroles, Julien

TRANSCRIPTIONAL DYNAMICS UNDER SELECTION: UNRAVELING POLYGENIC ADAPTATION AND STRESS RESPONSES

3.2 Baes, Christine

QUANTITATIVE GENETIC SOLUTIONS FOR OPTIMIZING LIVESTOCK SUSTAINABILITY: INNOVATIONS, GENOMIC APPLICATIONS, AND FUTURE DIRECTIONS

3.3 Baud, Amelie

3.4 THE HOLOGENOME 2.0

Genetics is traditionally understood as the study of how an individual's phenotype is affected by its own genotype. However, in recent years it has become clear that the genotypes of the individual's social partners (relatives and peers) and the genetic makeup of the individual's microbiota are also important. Furthermore, commensal microbes can be exchanged between socially interacting conspecifics. Therefore, we believe the individual should be studied not in isolation but, instead, together with its social partners and microbiota, which we refer to as the "Hologenome 2.0".
I will discuss the expected evolutionary consequences of the Hologenome 2.0 and present the empirical strategies we are developing to study it, using genetically diverse (outbred) and deeply phenotyped laboratory rodents as study system. We have already uncovered widespread and, in some cases, strong genetic effects arising from social partners (cage mates), and pinpointed specific genes involved. We are now developing an approach to gain insights into the traits of conspecifics that are most influential, leveraging the phenome-wide data available in our rodent cohorts to do so. We have also studied host/microbiome interactions in thousands of outbred rats, and have started developing methods to quantify and characterise microbial transmission between cage mates.

3.5 Bijma, Piter

DARK GENES: HOW TRANSMISSION OF INFECTIONS BOOSTS HERITABLE VARIATION AND RESPONSE TO SELECTION
Genetics is traditionally understood as the study of how an individual’s phenotype is affected by its own genotype. However, in recent years it has become clear that the genotypes of the individual’s social partners (relatives and peers) and the genetic makeup of the individual’s microbiota are also important. Furthermore, commensal microbes can be exchanged between socially interacting conspecifics. Therefore, we believe the individual should be studied not in isolation but, instead, together with its social partners and microbiota, which we refer to as the "Hologenome 2.0".

I will discuss the expected evolutionary consequences of the Hologenome 2.0 and present the empirical strategies we are developing to study it, using genetically diverse (outbred) and deeply phenotyped laboratory rodents as study system. We have already uncovered widespread and, in some cases, strong genetic effects arising from social partners (cage mates), and pinpointed specific genes involved. We are now developing an

approach to gain insights into the traits of conspecifics that are most influential, leveraging the phenome-wide data available in our rodent cohorts to do so. We have also studied host/microbiome interactions in thousands of outbred rats, and have started developing methods to quantify and characterise microbial transmission between cage mates.

3.6 Buckler, Ed

PARTICIPATION BIAS IN GENETIC STUDIES AND ESTIMATE ADJUSTMENTS

3.7 de los Campos, Gustavo

IMPROVING CROSS-ANCESTRY PGS PREDICTION THROUGH TRANSFER LEARNING USING INFORMATIVE PENALIZED REGRESSIONS AND BAYESIAN MIXTURE MODELS

In the last two decades, thousands of Genome-Wide Association Studies (GWAS) have been published. Increasingly, the findings reported by these studies inform the development of Polygenic Scores (PGS) that can be used to predict phenotypes and disease risk. The Polygenic Scores Catalog includes thousands of PGS. However, the overwhelming majority of the PGS were derived using data from Europeans and have poor predictive performance when used to predict phenotypes of individuals of non-European ancestry. Transfer Learning (TL) is a technique by which knowledge gained in one data set is used to improve the model's performance in another. We propose two novel TL methods to build PGS. The first one is a Penalized Regression that penalizes deviations from prior estimates (e.g., European-derived effects)–we present algorithms for L2, L1, and Elastic Net penalties. The second one is a Bayesian mixture model that uses external estimates as prior means–this approach offers the possibility of transferring knowledge from multiple sources of prior information. After introducing the above-described methods, I will present extensive benchmarks of these methods against other methods used to TL in PGS prediction using data from the UK-Biobank, All of Us, and the HCHS/SOL cohort and discuss the advantages and disadvantages of each approach.

3.8 Dekkers, Jack

IMPLEMENTATION OF A MECHANISTIC GROWTH MODEL FOR PIGS INTO BAYESIAN METHODS FOR GENOMIC PREDICTION AND GWAS

Mechanistic growth models have been developed to optimize formulation of diets for livestock. These models assume observed performances (growth rate, feed intake, body composition) are a nonlinear function of unobserved latent traits and environmental effects. We developed a Bayesian Hierarchical model to integrate longitudinal growth phenotypes (daily body weight and feed intake) and composition traits (back fat and lipid and protein mass) into genomic prediction for pigs using a mechanistic growth model. Breeding values for the latent variables are modeled using a multi-trait random regression marker effects model (GBLUP), while the latent variables are linked to the observed phenotypes using the mechanistic growth model. Through the mechanistic model, genomic predictions for properly formulated latent traits allow prediction of phenotypes at unobserved ages (e.g. mature weight) and have the potential to predict performance under different environments, as has been demonstrated for maize, and to predict non-additive genetic effects. Several computational strategies were implemented to reduce the computational demands of the model, including orthogonalization of the latent variables and of SNP genotypes, which greatly reduce the number of covariates fitted when the number of SNPs is greater than the number of genotyped individuals. To enable GWAS, estimated effects from these equivalent reduced models can be back-transformed to obtain effect estimates of the original SNPs. Although the latter estimates are invariant to the transformation used, this does not hold for the estimated variance contributed by a genomic window, which is the preferred criterion to detect QTL under linkage disequilibrium. However, breeding values for a genomic window, which are used to compute the

variance contributed by the window, can be separated into a component that is invariant to the transformation and that depends on the data, while the remainder does not depend on the phenotypic data. Using this concept, an MCMC sampler was developed to estimate the ior distribution of window variance for GWAS. This sampler can be applied to any Bayesian GWAS model. It was observed that, as the size of the window becomes smaller, the proportion of the window variance that depends on the data decreases, suggesting a strategy to optimize window sizes for GWAS. Methods were applied to data from two purebred pig lines for genomic prediction and GWAS. Further work is needed to evaluate the ability of the model to predict GxE and non-additive effects, including crossbred performance based on purebred phenotypes. Funded by USDA-NIFA 2020-67015-31031.

3.9 Endelmann, Jeff

DIRECTIONAL DOMINANCE IN POLYPLOIDS: TRAIT ANALYSIS AND MATE SELECTION

Directional dominance models provide a minimally realistic representation of quantitative traits, based on their ability to capture essential features, such as inbreeding depression/heterosis. To begin, I will review the recent theoretical development of a breeding value parameterization for directional dominance with digenic interactions, for any ploidy. Estimates of inbreeding depression and dominance variance from a tetraploid potato breeding program will be presented. A practical application of this theory is the optimal design of mating plans based on maximizing the expected mean of the F1 generation. This genetic merit objective, as well as its decomposition into mid-parent value and mid-parent heterosis, can be visualized as conic sections (e.g., hyperbolas and ellipses) in the (X,Y) plane for parental allele frequencies X and Y. For long-term genetic gain, maximizing genetic merit must be balanced against conservation of genetic diversity. In animal breeding, this has been achieved by limiting the inbreeding rate during mate selection. The polyploid generalization of this approach has been derived and is now available in the software COMA (Convex Optimization of Mate Allocation). Results from stochastic simulation of a breeding program will be presented to benchmark COMA against other established methods, such as optimum contribution selection and weighted genomic-estimated breeding values.

3.10 Fang, Lingzhao

THE FARM ANIMAL GENOTYPE-TISSUE EXPRESSION (FARMGTEX) PROJECT FOR ADVANCING AGRICULTURE AND BIOMEDICINE

3.11 Geiler-Samerotte, Kerry

THE GENOTYPE-PHENOTYPE-PHENOTYPE-PHENOTYPE MAP

3.12 Goddard, Michael

IDENTIFYING CAUSAL VARIANTS FOR HISTONE MODIFICATION

Many genomic variants have been associated with complex or quantitative traits (QTL) but linkage disequilibrium (LD) among variants makes it difficult to identify the causal variant. If causal variants throughout the genome shared a similar surrounding DNA sequence, this could be used to identify the causal variants. This is possible if you can delineate the segment of the genome that contains the causal variant.

QTL are enriched in functional parts of the genome as defined by features such as histone modification assayed by Chipseq and DNase hypersensitive regions (DHS). Beer et al() found gapped kmers that are enriched under DHS regions and, using them, predicted polymorphisms that would affect the height of the DHS peak.

We used data on heterozygous sites that affect the height of a Chipseq peak for the histone modification (H3K4me3). These sites show allele specific binding (ASB) in that one allele generates a higher Chipseq peak than the other. For each ASB site we identified other ASB sites throughout the genome that had a similar

sequence and used these sites to predict the which allele would be higher expressed at the target site. The predicted direction of ASB agreed with the observed direction 70% of the time which is significantly more than the 50% expected by chance. Thus sites where the direction of ASB can be predicted must be enriched for causal variants because the same sequence generates the same ASB throughout the genome which cannot be due to LD. Our prediction using ASB data was more accurate than that based on gapped kmers (gkm SVM). We next tested whether these putative causal variants for ASB also caused allele specific expression at nearby genes. We found that they were enriched for ASE but no more so than other heterozygous sites under ASB peaks. This may be due to LD between all sites under the same Chipseq peak or it may indicate that many sites that alter histone modification do not alter gene expression.

3.13 Hansen, Thomas

THE STRUCTURE OF EVOLUTIONARY QUANTITATIVE GENETICS

3.14 Johnston, Susan

THE CAUSES AND CONSEQUENCES OF SEX DIFFERENCES IN RECOMBINATION RATES

The rate of meiotic recombination often shows large differences between the sexes. It can be strongly female-biased (humans), strongly male-biased (macaques/sheep) or somewhere in between. Understanding why this happens is key to understanding the evolution of recombination rates, yet the causes and consequences of this phenomenon remain unknown. This talk will focus on our most recent work in house sparrows (Passer domesticus), with broader context from work in mammals and fish. We use genomic data in large pedigrees to characterise individual recombination rates and landscapes to: (a) investigate the heritability and genomic basis of variation in recombination rates; (b) identify genomic correlates with fine-scale sex-differences in recombination landscapes; and (c) use genomic prediction approaches to understand the relationship between recombination and fitness within each sex. Our work provides a foundation for broader understanding of the vast diversity of recombination rates in eukaryotic genomes.

3.15 Kong, Augustine

PARTICIPATION BIAS IN GENETIC STUDIES AND ESTIMATE ADJUSTMENTS

The rate of meiotic recombination often shows large differences between the sexes. It can be strongly female-biased (humans), strongly male-biased (macaques/sheep) or somewhere in between. Understanding why this happens is key to understanding the evolution of recombination rates, yet the causes and consequences of this phenomenon remain unknown. This talk will focus on our most recent work in house sparrows (Passer domesticus), with broader context from work in mammals and fish. We use genomic data in large pedigrees to characterise individual recombination rates and landscapes to: (a) investigate the heritability and genomic basis of variation in recombination rates; (b) identify genomic correlates with fine-scale sex-differences in recombination landscapes; and (c) use genomic prediction approaches to understand the relationship between recombination and fitness within each sex. Our work provides a foundation for broader understanding of the vast diversity of recombination rates in eukaryotic genomes.

3.16 Lippman, Zach

SYNERGY AMONG CRYPTIC VARIANTS IN A PLANT REGULATORY NETWORK DRIVE NONLINEAR PHENOTYPIC EFFECTS

3.17 Po-Ru, Loh

INFLUENCES OF GENOMIC STRUCTURAL VARIATION ON HUMAN COMPLEX TRAITS

Genetic association analyses of large genotype-phenotype data sets have identified many thousands of SNPs and indels associated with human complex trait variation. However, few genetic association studies have analyzed genomic structural variants: i.e., polymorphisms modifying >50 base pairs of DNA sequence. Because of their large size, structural variants collectively contribute more base pairs of variation within an individual's genome than SNPs and indels. However, structural variants have been difficult to identify and genotype from the SNP-array and short-read sequencing data generated by biobanks to date.

We have recently undertaken several efforts to better ascertain and genotype SVs from biobank sequencing data and to explore their influences on human complex traits. I will describe two of these efforts, starting with an analysis of protein-altering copy-number variants (CNVs) in UK Biobank that leveraged haplotype-informed methods to sensitively detect protein-altering CNVs from whole-exome sequencing data. Subsequent association and fine-mapping analyses identified many likely-causal CNV-trait associations, including a low-frequency partial deletion of *RGL3* exon 6 that appeared to confer one of the strongest protective effects of gene LoF on hypertension risk (OR = 0.86 [0.82–0.90]). Additionally, protein-coding variation in rapidly-evolving gene families within segmental duplications—previously invisible to most analysis methods—appeared to generate some of the human genome's largest contributions to variation in type 2 diabetes risk, chronotype, and blood cell traits.

In a second analysis of UK Biobank WGS data, we found that a sequence of SVA retrotransposon insertions in an early intron of the *ASIP* (agouti signaling protein) gene has likely shaped human pigmentation multiple times. We identified a recent 3.3kb SVA retrotransposon insertion polymorphism that appears to underlie one of the strongest common genetic influences on pigmentation and skin cancer risk within European populations. *ASIP* expression in human skin displayed the same association pattern as lighter pigmentation and skin-cancer risk did, with the same insertion allele exhibiting 2.2-fold (1.9–2.6) increased expression of *ASIP*. This effect had an unusual apparent mechanism: an earlier, non-polymorphic, human-specific SVA retrotransposon 3.9kb upstream appeared to have caused *ASIP* to exhibit nonproductive splicing, which the new (polymorphic) SVA insertion largely eliminated. These results suggest that a sequence of retrotransposon insertions contributed to a species-wide increase, then a local and variable decrease, of human pigmentation.

3.18 Mbatchou, Joelle

USING LARGE LANGUAGE MODELS FOR RARE VARIANT ASSOCIATION TESTING IN LARGESCALE BIOBANKS

3.19 Pasaniuc, Bogdan

POLYGENIC RISK SCORES FOR PRECISION MEDICINE: PROMISES AND CHALLENGES

SEARCHING FOR CAUSAL VARIANTS IN POLYGENIC TRAITS

Polygenic traits are influenced by a large number of genetic variants, each with modest effect, making it particularly challenging to identify precisely those DNA polymorphism with a causal effect. Yet, to increase our understanding of the biology behind these traits and to equitably and robustly deploy this knowledge, it is important to attempt this identification.

We will illustrate how an approach based on knockoffs can substantially improve detection power, while maintaining control of the false discovery rate. The method handles linkage disequilibrium, population structure, and can be used in combination with deep learning to study gene-environment interactions.

3.20 Sabatti, Chiara

SEARCHING FOR CAUSAL VARIANTS IN POLYGENIC TRAITS

Polygenic traits are influenced by a large number of genetic variants, each with modest effect, making it particularly challenging to identify precisely those DNA polymorphism with a causal effect. Yet, to increase our understanding of the biology behind these traits and to equitably and robustly deploy this knowledge, it is important to attempt this identification.

We will illustrate how an approach based on knockoffs can substantially improve detection power, while maintaining control of the false discovery rate. The method handles linkage disequilibrium, population structure, and can be used in combination with deep learning to study gene-environment interactions.

3.21 Sella, Guy

A POPULATION GENETIC INTERPRETATION OF GENOME-WIDE ASSOCIATION STUDIES IN HUMANS

I will briefly describe a population genetic model for the genetic architecture of complex, quantitative traits and illustrate its usefulness in interpreting findings from genome-wide association studies (GWASs) in humans. Notably, GWASs in humans have revealed that the genetic architectures of complex traits vary widely in terms of the numbers, effect sizes, and allele frequencies of significant hits. Fitting the model to GWAS hits for 95 highly polygenic quantitative traits from the UK Biobank reveals that differences in their architecture arise mainly from two evolutionary parameters: the mutational target size and heritability per site, which vary by orders of magnitude among traits. When the effect sizes of variants are measured in units that account for the differences in these parameters, the architecture of all 95 traits becomes strikingly similar. Additionally, time permitting, I will show how the same model predicts when genes and other functional genomic elements that are important to trait biology stand out in GWAS.

3.22 Sztepanacz, Jacqueline

ESTIMATING GENETIC VARIATION AND SELECTION IN HIGH-DIMENSIONAL DATA

Genetic correlations caused by the pleiotropic effects of alleles on multiple traits may limit the evolvability of populations, leading to slow or constrained evolutionary responses. Conversely, pleiotropy can facilitate adaptation by concentrating genetic variation in directions of phenotype space under selection. However, quantifying the effects of pleiotropy on evolutionary response requires multivariate studies of genetic variation and selection, both of which present biological and statistical challenges. In this talk, I share some new methods to estimate genetic variation and selection in high-dimensional data, and show how it can help us understand the critical role of pleiotropy in determining the distribution of genetic variation within populations and evolutionary responses to selection.

3.23 Wolf, Jason

GENETIC ANALYSIS OF INTRAFAMILIAL INTERACTIONS

Genetic correlations caused by the pleiotropic effects of alleles on multiple traits may limit the evolvability of populations, leading to slow or constrained evolutionary responses. Conversely, pleiotropy can facilitate adaptation by concentrating genetic variation in directions of phenotype space under selection. However, quantifying the effects of pleiotropy on evolutionary response requires multivariate studies of genetic variation and selection, both of which present biological and statistical challenges. In this talk, I share some new methods to estimate genetic variation and selection in high-dimensional data, and show how it can help us understand the critical role of pleiotropy in determining the distribution of genetic variation within populations and evolutionary responses to selection.

3.24 Wray, Naomi

QUANTITATIVE GENETICS OF PSYCHIATRIC DISORDERS

For a long time, the theoretical and analytical approaches to quantitative traits ran parallel in livestock and human genetics, the former tracing back to Fisher (mixed models) and the latter to Wright (path coefficients; adopted by twin researchers). For binary traits, similar differences existed, liability threshold model in livestock and epidemiological models in humans, despite early attempts by a handful of researchers, notably the late Charlie Smith, to change that. These past differences were mainly driven by applications – estimation and prediction in livestock and causality inference and disease mapping in humans. The genome era has changed all that, by cross-fertilisation (and re-invention) of both analytical methods and applications. There remain differences that are caused by different population structures and these can affect the choice of analysis method and the interpretation of results. I will provide an overview of research synergies between human and livestock genetics and will invite discussion about future synergies.

3.25 Yengo, Loic

CONVERGENCE OF HERITABILITY ESTIMATES FROM ORTHOGONAL EXPERIMENTAL DESIGNS

4 Oral Presentations

A NEW UNRESTRICTED ASSESSMENT TOWARD UTILIZING INDIVIDUAL PLANT PHENOTYPES AND GENOTYPES FOR BREEDING

Aiyesa, Leke Victor[1]; Link, Wolfgang[2]; Scholten, Stefan[3]; Beissinger, Timothy M.[4]

[1]Plant Breeding Division, Department of Crop Sciences, Faculty of Agriculture, University of Goettingen, Carl-Sprengel-Weg 1, 37075 Goettingen, Germany.; [2]Crop Plant Genetics Division, Department of Crop Sciences, Faculty of Agriculture, University of Goettingen, Von-Siebold-Str 8, 37075, Goettingen, Germany.; [3]X, the moonshot factory, Mountain View, California, United States.; [4]Centre for Breeding Research (CiBreed), University of Goettingen, Albrecht-Thaer Weg 3, 37075, Goettingen, Germany.

Advances in phenotyping and genotyping enable the evaluation of a large number of individual plants (IPs) in field conditions. In the near future, the cost of genotyping and phenotyping will no longer be a limiting factor in plant breeding; instead, logistical challenges such as developing and maintaining germplasm families will be. To explore this, we assembled a panel of 1000 IPs from 50 European maize landraces (~20 IPs/population) across 9 countries. Field experiments for 2 years provided data on 15 agronomic traits, analyzed alongside 120,261 SNP markers and 18,671 haplotypes to assess the genomic potential of IPs. Genetic diversity analysis revealed that 52% of the genetic variance was partitioned among IPs within populations, uncovering a serial

founder effect during maize migration to northeastern Europe from the tropical south. We identified 447 SNPs and 87 haplotype blocks strongly associated with local adaptation and agronomic traits in European maize. Notably, 127 candidate genes including Tb1, ZCN7, and ZmMADS69, were identified, which regulate flowering properties in maize. Prediction accuracies using 6 statistical and 3 machine learning methods varied from 0.25 for ear length to 0.68 for silking date and 0.54 for grain weight. Although comparable to certain prediction results from developed panels of inbreds, they nonetheless exhibit higher error variance. This study illustrates the genetic gains achievable through field-grown IPs and provides insights into implementing IP-based experiments and analysis. The curated dataset offers a valuable IP resource for breeding research and improvement for European maize and others.

GENETIC VARIATION IN PROTEIN DEGRADATION

Collins, Mahlon; Avery, Randi; W Albert, Frank

University of Minnesota, Department of Genetics, Cell Biology, & Development, Minneapolis, MN 55422, USA

A key question in complex trait genetics is how genetic variation affects cellular activities that in turn shape organismal traits. Protein degradation is a key cellular trait with critically important roles in gene expression regulation and proteostasis that are essential for cell function. We study natural genetic variation in protein degradation in a cross between two genetically different isolates of the yeast Saccharomyces cerevisiae. We focus on the ubiquitin-proteasome system (UPS), the main protein degradation machinery in eukaryotic cells. Cellular reporters (termed "tandem fluorescent timers") of UPS activity allow us to conduct bulk segregant analysis in millions of recombinant, living, single cells. Using fluorescence-activated cell sorting, we collect pools of thousands of cells with high or low UPS activity. Whole-genome sequencing of these pools reveals loci that influence the UPS with high statistical power. We recently showed that UPS activity is a genetically complex trait that is shaped by numerous loci throughout the genome, often with pathway-specific effects on the molecular branches of the UPS (Collins et al., 2022 & 2023). Genome engineering at five loci revealed multiple causal DNA variants at each of several genes with core UPS functions, ranging from ubiquitin ligases to a proteasome component. The causal variants included coding and noncoding variants and ranged from rare mutations to alleles that are common in the global yeast population. Other loci mapped to pleiotropic genes known to affect many complex traits. Here, we present key extensions of this work. First, UPS activity is highly responsive to the cellular environment. We therefore studied how genetic variation in UPS activity is modified by gene-by-environment interactions (GxE). Genetic mapping of six pathway-specific UPS reporters in eight environments revealed 419 loci across 48 pathway/environment combinations. GxE was enriched at pleiotropic loci that likely shape the UPS via indirect

mechanisms, suggesting that GxE is more likely to occur where there is a larger number of molecular steps between the causal variant and the trait. Second, we have mapped variation in the degradation and abundance of 49 individual proteins. Protein degradation was shaped by a median of seven loci per protein for a total of 382 loci, 98% of which acted in trans. Trans-acting protein degradation loci contained core UPS genes and pleiotropic regulators, in addition to new loci with unknown molecular basis. At least 20% of loci that influenced a protein's abundance also influenced the same protein's degradation, suggesting degradation as a causal mechanism for protein-specific effects on gene expression. Surprisingly, at about half of these locus pairs, increased degradation was linked to higher (rather than lower) abundance of the protein, suggesting that in these cases protein degradation may act to partially buffer protein abundance against variation in the levels of the genes' mRNAs. Together, this work establishes protein degradation as a new model for the genetics of cellular traits.

SCALABLE SINGLE-CELL MODELS FOR ROBUST CELL-STATE-DEPENDENT EQTL MAPPING

Disease risk variants identified through genome-wide association studies (GWAS) are enriched in non-coding genomic regions, suggesting a potential regulatory role. However, regulatory variants identified through expression Quantitative Trait Locus (eQTL) mapping exhibit only a modest overlap with GWAS variants. This may be because traditional eQTL models aggregate RNA profiles from many cells and fail to capture the distinct regulatory programs of causal cell states. We recently proposed a single-cell-resolution eQTL modeling approach (Nathan A et al, 2022) that offers a high-resolution approach to linking genetic variation to gene expression in cell types (such as T cells) or states (e.g. cytotoxicity and proliferation). However, accurately modeling single-cell expression represents a challenge for single-cell eQTL mapping. In some instances, individual genes may exhibit distinct count distribution patterns in single-cell data due to data sparsity and heterogeneous contributions of various cell states. This violates parametric assumptions and may lead to p-value inflation and a higher false-positive rate. Furthermore, current eQTL methods fail to scale to large datasets of millions of cells. Therefore, they are limited to testing only a few variants across a whole locus already defined from pseudobulk analysis.. To solve these problems, we propose a non-parametric linear mixed-effects modeling approach to map cell-state-dependent single-cell eQTLs. Key features of this model include: 1) adaptive bootstrapping to derive statistically calibrated p-values without any assumption about the gene expression distribution, 2) pseudocell aggregation and covariate regression prior to eQTL mapping to reduce computation time while retaining statistical power. We applied this method to 500,089 memory T cells from 259 Peruvian individuals and identified cell-state-dependent eQTLs for autoimmune disease-risk genes such as IL2RA and SH2B3 which had no evidence of eQTL using pseudobulk analysis. Some eQTLs from single-cell resolution analysis were not detected with bulk analysis and had regulatory effects specific to rare states. These results suggest that testing eQTL interactions with cell states can identify cell-state-

dependent regulatory variants that would be obscured in traditional eQTL mapping.

EFFECTIVE POPULATION SIZE IN HONEYBEES FOM PREDIGREE AND SNP DATA

Bernstein, Richard; Du, Manuel; Hoppe, Andreas

Institute for Bee Research Hohen Neuendorf

Effective population size is a useful way to put the inbreeding developments of a population into a nutshell. There is a wide array of tools to estimate effective population size from classic formulas derived for idealized populations, over the increase of inbreeding in the population, to linkage-based methods. Genomic methods are especially appealing for populations where no pedigree records are available, and linkage-based methods even offer estimations when only genotypes of a single generation are available. Honeybees pose challenges to all these methods due to their haplodiploid mating biology, and polyandry. In the most common forms of controlled mating, virgin queens are brought to geographically isolated areas such as islands or valleys, where a sister group of drone producing queens (DPQ) is stationed. DPQs are often not phenotyped, and not used to produce daughter queens, which is the role of breeding queens (BQ). On a mating station, the virgin BQs mate during flight with 10 to 20 drones. Consequently, unknown paternity must be accounted for. We used forward in time simulations to generate a base population, and develop it under random selection with different schemes of mating. The number of DPQs on a mating station, the number of mating stations, and the number of daughter queens varied across the mating schemes. The simulations provided bi-allelic marker data for queens and drones. Each scheme was simulated over 40 years, repeated 50 times and the results were averaged. The simulation software keeps track of the individual sire drones of daughter queens, and provide accurate pedigree relationships. We used this as a gold standard to evaluate the quality of other methods. Since drones cannot be tracked in practice, a specialized algorithm was used to calculate inbreeding from realistic pedigrees in years 0, 9, 19 and 39, in order to estimate the effective population size. Genomic inbreeding can be calculated from queen genotypes as in diploid species. For linkage-based methods, the extremely high recombination rate of honeybees (20 cM/Mb) must be considered. Preliminary results suggest that estimates from realistic pedigree information are very reliable. Alternatively, the genomic inbreeding of queens yields similar results, when the allele frequencies of the markers in the base population are known. Linkage based methods often underestimated the effective population size, but produced significantly different results, when the underlying populations followed very different mating schemes. We applied the methods to real data sets from Central Europe, where pedigree records date back by 70 years, and a high-density SNP chip was used to genotype more than 4000 honeybees within the last decade.

LIGHT-SPEED WHOLE GENOME ASSOCIATION TESTING AND PREDICTION VIA APPROXIMATE MESSAGE PASSING

The recent release of whole-genome sequence (WGS) data for all UK Biobank participants facilitates investigating the impact of rare variants on complex traits. The common statistical approach of single-marker, or single-gene burden score regression, gives marginal associations that do not account for linkage disequilibrium, and in large sample size high density WGS data, even weak associations that are physically distant from causal variants will be discovered as significant. This limits our understanding of the genetic basis of human traits. Here, we present a new algorithmic paradigm, gVAMP, that (i) directly fine-maps WGS variants and gene burden scores, conditional on all other measured DNA variations genome-wide, and (ii) provides optimal polygenic risk score prediction. This is done in one-shot, in lightning speed, allowing the analysis of datasets considered impossible before. On DNA Nexus, it takes just over one day to fine-map many thousands of human height-associated WGS variants and create a polygenic risk score with 48% prediction accuracy. We find 60 genes where rare coding mutations significantly influence phenotype, 76 X-chromosome associations, and thousands of autosomal associations localised to the single-locus level for five additional traits. Additionally, gVAMP outperforms summary statistic polygenic score methods, and outperforms REGENIE for standard association testing in a fraction of the compute time across 13 traits in imputed sequence data. In summary, we showcase how to fine-map variants genome-wide in large sequencing datasets and provide open-source software.

AN EFFICIENT ANALYSIS METHOD FOR INTEGRATING MULTIPLE OMICS DATA BASED ON DEEP LEARNING

Liu, Huatao; Wang, Chuduan; Ding, Xiangdong

China Agricultural University

Fat deposition is closely related to pig production efficiency, pork quality, and reproductive traits. At the same time, pigs are an ideal model animal for studying human obesity. Therefore, exploring the key genes that affect pig fat deposition has always been a research hotspot. Fat deposition, as a complex trait, is often regulated by multiple omics levels, and the integration and analysis of multiple sets of data is crucial for deciphering its regulatory mechanism. In order to effectively utilize the supplementary information contained in multi omics data, it is necessary to develop models that can represent different data layers and integrate them into a single framework. As an emerging big data analysis method, machine learning can effectively fit complex data and accurately identify samples and genes. The high learning ability and flexibility of deep neural networks, the fundamental structure of deep learning, make them more advantageous in analyzing high-dimensional large-scale datasets. Therefore, we propose a new multi omics integration analysis method based on deep neural networks, which takes genotype coding values and gene expression data as inputs, fits different omics data through two different coding sub networks, and then connects a sub neural network to integrate multi omics data, establishing the connection between omics data and target phenotype, and ultimately identifying important omics features related to phenotype through feature substitution. We selected three main organs involved in fat deposition, namely

pig adipose tissue, liver, and muscle, for multi omics sequencing, and integrated publicly available data from similar samples to construct a multi omics dataset for pig fat deposition. Based on the our dataset, the model was trained and tested, achieving a prediction accuracy of over 0.9, while identifying several key genes related to fat deposition. We also validated the model using human medical datasets and mouse datasets, and the results showed that the method outperformed traditional regression models, LASSO, and eight general machine learning models in multiple datasets. The model constructed in this study not only achieved higher prediction accuracy, but also more efficiently and accurately identified phenotype related omics features. This study provides new ideas for the integration and analysis of multi omics in biological data, and lays the foundation for analyzing the complex regulatory mechanisms of fat deposition.

Estimating and dissecting host genetic variation underlying infectious disease transmission – methodology and empirical evidence

Doeschl-Wilson, A.[1]; Prentice, J.C.[2]; Pooley, C.M.[3]; Pong Wong, R.[4]; Marion, G.[5]; Houston, R.[1]; Robledo, Diego[2]; Cabaleiro, S.[3]; Villanueva, B.[4]

[1]The Roslin Institute, University of Edinburgh, Easter Bush, EH25 9RG, UK;; [2]BioSS, The King's Buildings, Edinburgh, EH9 3FD, UK;; [3]Benchmark Genetics, Edinburgh Technopole, Edinburgh, EH26 0GB, UK;; [4]Centro Tecnológico del Cluster de la Aquicultura (CETGA), A Coruña, Spain.,; [5]INIA-CSIC, Ctra. de La Coruña, km 7.5, 28040, Madrid, Spain;

Individuals differ widely in their contribution to the spread of infection within and across populations, which may be partly genetically determined. In particular, three key epidemiological host traits under potential genetic control affect infectious disease spread: susceptibility (propensity to acquire infection), infectivity (propensity to transmit infection to others) and recoverability (propensity to recover / die). Interventions aiming to reduce pathogen spread may target improvement in any one of these traits, but the necessary statistical methods for obtaining genetic risk estimates are lacking. Here, we introduce the Bayesian methodology and a new inference software SIRE (Susceptibility – Infectivity – Recoverability – Estimator) for estimating genetic co-variances and individual risks for all three host epidemiological traits from temp epidemic data, assuming pedigree or genomic relationships are known. Extensive validation of the methods with diverse types of data from a wide range of simulated epidemics differing in population and contact group size and structure provided valuable insights of data requirements to achieve good prediction accuracies for all three host traits. We then applied this methodology to data from a large-scale disease transmission experiment in turbot, where we recorded disease and survival status of 1800 genotyped fish distributed across 72 contact groups, to obtain the first empirical genetic parameter estimates for host infectivity, in addition to susceptibility, as well as infection induced mortality (inverse of recoverability). Confirming expectations from evolutionary theory, we found substantial genetic variation in host infectivity, in addition to host susceptibility. In contrast, genetic variation in infection induced mortality was close to zero. Despite high

uncertainties, estimated genetic correlations between the traits were unfavorable, highlighting the importance for considering all three host traits underlying disease transmission and survival in populations. Genome wide association studies suggest that all three traits are mostly under polygenic control, although some putative SNPs of moderate effect were detected for all three traits. These are located in different genomic regions, indicating that traits may be under different genetic regulation and evolve differently. Finally, using genetic-epidemiological prediction models, we demonstrate how the ability to estimate genetic effects for host susceptibility, infectivity and recoverability could be utilized for more effective disease control.

UNRAVELLING THE GENETIC ARCHITECTURE OF HEIGHT AND MUSCULAR DEVELOPMENT TRAITS IN BELGIAN BLUE CATTLE AND USING IT FOR GENOMIC PREDICTION

As a result of intensive selection, Belgian Blue beef cattle have extreme muscular development. They present a double-muscling phenotype resulting from the fixation of an 11 bp deletion in the myostatin gene causing a premature stop codon. However, after fixation of this mutation, muscularity has been further increased and this selection has been accompanied with a reduction in stature. Here, we used a cohort of more than 15,000 cows imputed at the sequence level and phenotyped for muscular development traits and height to study the genetic architecture of these traits using GWAS and heritability partitioning approaches. Among the 15 significantly associated variants, we found an enrichment of common coding variants with large effects. A first set of these coding variants were found in genes also associated with size or growth disorders in other mammals (e.g. LCORL, PAPPA2, ADAM12, EZH2), while five others were breed specific recessive deleterious variants, including variants causing genetic defects such as crooked tail syndrome or stunted growth, that conferred a heterozygous advantage (e.g. increased muscularity). Evidence of regulatory effects was also found for variants in CCND2 and ARCM12, which interestingly increases the activity of EZH2. Taken together, these variants with a large effect accounted for only a small proportion of the genetic variance. We then used GREML and a Bayesian grouped mixture of regressions model called GMRM to partition heritability according to different genomic compartments including, coding sequence, regions upstream or downstream of genes (+/- 1kb), open chromatin, intronic and intergenic regions. The percentage of genetic variance associated with each category, called %SNP heritability, was highly variable across traits and methods. Nevertheless, we estimated that on average, variants in open chromatin regions had a higher contribution to the genetic variance (> 45%), while variants in coding regions had the strongest individual effects (> 25-fold enrichment on average). Conversely, variants in intergenic or intronic regions showed lower levels of enrichment (0.2 and 0.6-fold on average, respectively). We therefore investigated whether incorporating annotation information into genomic selection, for example by giving more weight to coding or regulatory variants, could improve its accuracy. First, genomic predictions were made with the GBLUP and GMRM models using the full sequence data. In this case, estimating specific parameters for each annotation group (e.g. effect variances,

priors for mixture proportions) only marginally improved GS accuracy compared to predictions without annotation. Next, we reduced the number of variants by selecting mainly coding and putative regulatory variants or with an LD pruning approach. In both cases, the accuracy levels were comparable to those obtained with the full sequence data. Overall, our results show that a few large effect coding variants are often associated with complex traits in livestock species, although regulatory variants are likely to make the largest contribution to genetic variation. However, further investigation is needed to efficiently exploit this information in genomic selection.

THE EFFECT OF POPULATION STRUCTURE CORRECTION ON GWAS BEFORE AND AFTER RANDOM MATING

Ellis, Thomas James[1]; Gunis, Joanna[1]; Clauw, Pieter[1]; Rabanal, Fernando[2]; Nordborg, Magnus[1]; Thomas, Ellis[3]; Gunis, Joanna[3]; Clauw, Pieter[3]; Rabanal, Fernando[3]; Nordborg, Magnus[3]

[1]Gregor Mendel Institute, Austrian Academy of Sciences, Dr.-Bohr-Gasse 3, 1030 Vienna; [2]Department of Molecular Biology, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany.; [3]Agri-Food and Biosciences Institute - AFBI

Although considerable progress has been made in developing statistical methods for modeling population structure confounding in GWAS, these methods are not perfect, and power will inevitably be reduced if genome-wide correlations between loci are too strong. The traditional alternative is using a controlled cross between two genotypes in which unlinked loci are effectively randomized with respect to each other. This has the disadvantage that resolution is much poorer than GWAS, and allele frequencies are distorted. To get around this problem, we generated a mapping population that kept the natural allele frequencies, but decreased the amount of linkage disequilibrium. We did this by carrying out well over 1000 crosses between parents chosen at random from natural lines of Arabidopsis thaliana, and propagated each cross to the F9 generation. We are using genotype data to compare the genetic structure of the parents and F9s. We compare the performance of GWAS in these two cohorts to assess the effect of the population structure correction on the identification of genotype-phenotype associations for three traits related to growth and reproduction.

GENTROPY PACKAGE FOR ANCESTRY SPECIFIC SYSTEMATIC FINE-MAPPING OF GWAS DATA, COLOCALIZATION AND DRUG TARGETS PREDICTION

GE, Xiangyu Jack[1]; Considine, Daniel[1]; López Santiago, Irene[1]; Tsukanov, Kirill[2]; Buniello, Annalisa[2]; McDonagh, Ellie M.[2]; Suveges, Daniel[3]; Ochoa, David[3]; Tsepilov, Yakov A.[3]; GE, Xiangyu Jack[1]; Considine, Daniel[1]; López Santiago, Irene[1]; Tsukanov, Kirill[2]; Buniello, Annalisa[2]; McDonagh, Ellie M.[2]; Suveges, Daniel[3]; Ochoa, David[3]; Tsepilov, Yakov A.[3]

[1]Open Targets, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK 2.; [2]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton,

Cambridgeshire CB10 1SA, UK; [3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

One of the challenges in the genetics field was that there was no high-throughput, scalable, openly available solution for the systematic in-silico follow up analysis of GWAS for drug target prioritization. Previously, we developed Open Targets Genetics (https://genetics.opentargets.org/), a comprehensive tool for predicting causal genes in established human GWAS loci using a machine learning framework. Though effective, the original analysis pipeline faced computational challenges, including scalability to ever-growing datasets, and was limited to GWAS of European ancestry. Here we present GentroPy, a toolkit optimized for scalable and reproducible genomic analysis, that encompasses the Open Targets Genetics pipelines for locus-to-gene prediction. This Python package enables the ingestion of the ever-expanding volume of multi-omic datasets and the ability to incorporate non-European ancestries. The package employs PySpark for the ETL pipeline, with default utilization of Google Cloud Platform for computational tasks, although local operations are possible. The package includes the first optimized version of the genetic data ingestion and harmonization pipelines, clumping, PICS and SuSiE-inf fine mapping, eCAVIAR and COLOC colocalization, outliers detection method CARMA, summary statistics imputation, and locus-to-gene score computation pipelines. It uses GnomAD v2.1 as the LD reference for seven major population ancestries. The package was applied to all GWAS Catalog studies, FinnGen R10 studies, GWAS catalog curated associations, eQTL catalogue, and UKBB-PPP proteomics data. It resulted in more than 600,000 credible sets related to more than 15,000 different traits. It was combined with more than 2 million different molQTL credible sets in target prediction machine learning framework. From raw data to locus-to-gene prediction scores, the entire analysis now takes less than 24 hours to complete. The publicaly available package, accessible at https://github.com/opentargets/gentropy, facilitates the interpretation and analysis of GWAS and functional genomic studies for target identification.

MULTI-TRAIT AND MULTI-ANCESTRY POLYGENIC RISK SCORE APPROACH IMPROVES GENETIC DISCOVERY AND RISK PREDICTION OF RESPIRATORY DISEASES

He, Yixuan[1]; Ho Jee, Yon[2]; Cho, Michael[3]; Moll, Matthew[4]; Wang, Ying[5]; Tsuo, Kristin[6]; Martin, Alicia[6]

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; [4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; [5]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; [6]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA

Introduction: Respiratory diseases such as chronic obstructive pulmonary disease (COPD) and lung cancer are leading causes of morbidity and mortality globally. While these diseases share many comorbidities as well as genetic, clinical, and lifestyle risk factors, most studies to date have modeled genetic risk in single traits in primarily European ancestry groups without modeling the high genetic correlation between these related traits or linkage disequilibrium (LD) and allele frequency patterns between populations. In this study, we substantially increase genomic discovery for lung function in East Asians, then develop and validate PRS-xtra (cross TRait and Ancestry), a multi-trait and multi-ancestry polygenic risk score (PRS) approach, for predicting respiratory diseases. Methods: We implemented a stepwise approach to model multi-trait and multi-ancestry features for 8 strongly correlated traits–COPD, asthma, lung cancer, FEV1, FVC, FEV1/FVC, smoking, and cigarettes/day–in African (AFR), Admixed American (AMR), East Asian (EAS), and European (EUR) ancestry groups. We first conducted the largest meta-analysis for lung function in East Asians from the Korean Cancer Prevention Study - II and Taiwanese Biobank. To model genetic correlation between traits, we then leveraged the largest and most diverse GWAS from the Global Biobank Meta-analysis Initiative, Pan UKBiobank, and GWAS & Sequencing Consortium of Alcohol and Nicotine use in a multi-trait analysis of GWAS (MTAG). We implemented PRS-CSx to model LD within and between populations to derive 39 PRSs in 289K individuals from the All of Us research program (AoU). We used regularization to penalize individual PRSs to synthesize PRS-xtra and evaluated PRS-xtra in a held out test set of 124K individuals from AoU. Results: The meta-analysis of lung function in East Asians (N=132K) identified 44, 73, and 31 independent loci for FEV1, FVC, and FEV1/FVC, respectively, of which 25 are novel. All traits were significantly correlated with each other, with the strongest correlations being between asthma and COPD (rg=0.705) and lung cancer and cigarettes/day (rg=0.614). Modeling genetic correlations across traits using MTAG, the number of independent loci increased for all traits, with COPD in the EAS cohort having the largest increase from 4 to 26. PRS-xtra and PRS were significantly correlated with each other ($p<0.0001$), but PRS-xtra demonstrated substantial improvements in predictive performance, especially in non-EUR populations. For example, PRS-xtra significantly improved predictive accuracy of asthma and COPD in AMR compared to PRS, with AUCs increasing from 0.509 (0.493-0.524) to 0.630 (0.616-0.645) and 0.513 (0.486-0.541) to 0.628 (0.603-0.653), respectively. PRS-xtra is also able to predict COPD exacerbations significantly better than PRS in the validation cohort, with AUC increasing from 0.572 (0.558-0.586) to 0.600 (0.587-0.614). Conclusion: We conducted the most powerful multi-trait and multi-ancestry genetic analysis of respiratory diseases and auxiliary traits to date. We propose PRS-xtra as a method to model genetic correlation across traits and LD differences between ancestry groups that demonstrate significantly better disease prediction, thereby advancing more equitable and generalizable prediction models.

INBREEDING DEPRESSION THROUGHOUT THE GROWTH PERIOD OF WILD SWISS BARN OWLS

Hewett, Anna; Lavanchy, Eléonore; Cumer, Tristan; Topaloudis, Alexandros; Simon, Celine; Roulin, Anne-Lyse; Roulin, Alexandre; Goudet, Jérôme

Department of Ecology and Evolution; University of Lausanne (UNIL)

Inbreeding depression can be extremely detrimental for wild populations, the lowered fitness of inbred individuals can reduce population size and in the worst cases lead to extinction. However, studying inbreeding depression in wild populations is difficult because fitness and genomic/pedigree data is usually needed, which can be hard to obtain. Here, we used a long-term dataset of >3000 wild barn owls gathered across Switzerland in combination with individual genomic inbreeding coefficients (Funi and FROH) estimated from high and low coverage whole genome sequencing – where genotypes of low coverage individuals were imputed from high coverage individuals. Using repeated measurements of wing length, bill length and mass over the growth period of an individual (and beyond) we ran non-linear hierarchical mixed models to assess the impact of inbreeding on three growth rate parameters: initiation of growth, growth rate and the final trait value (asymptote). We show that inbreeding depression is present not only in the final trait value, but also during the period of growth in some traits. Additionally, we show that the inbreeding coefficient used (Funi or FROH) can affect the inferences which may be due to the genetic architecture of inbreeding depression. Finally, we also show that the heritability of traits changes over the course of life, with environmental factors contributing more to phenotypic variation in early life.

THE STRUCTURE OF POTENTIALLY FUNCTIONAL GENETIC VARIATION IN CATTLE

Adepoju, Dolapo; Aivazidou, Stella; Lanigan, Simon; Klingström, Tomas; M. Johansson, Anna; Rius-Vilarrasa, Elisenda; Johnsson, Martin

Department of Animal Biosciences, Swedish University of Agricultural Sciences, Uppsala, Sweden. Växa Sverige, Stockholm, Sweden.

Population genetic features such as population history and historical selection influence the distribution of genetic variants and their association, and thus the performance of genomic prediction. In our ongoing research on the genome dynamics of livestock breeding, we explore the effects of features of the genome on breeding in farm animals. We used linkage disequilibrium-based methods to estimate the recent population size history of cattle, including publicly available data from Swedish local breeds (Swedish Red cattle, Swedish Mountain cattle, Fjällnära cattle, Ringamåla cattle, Bohus Polled, Väne cattle, and Swedish Red Polled) and major dairy breeds (Holstein and Jersey). While there is substantial variation in these estimates, they qualitatively agree that recent population size has been relatively large (on the order of thousands) followed by a rapid decline at around the onset of systematic breeding. We used publicly available sequence data to estimate variant intolerance of cattle genes (using Residual Variant Intolerance Scores), highlighting genes that have more or less common potentially functional variants than expected from total variation. This analysis identified immune genes among the most variant tolerant, whereas the least tolerant genes were enriched for basal molecular processes such as

transcriptional regulation. This is consistent with results from human genetics, whereas the correlations of scores for orthologous genes between species was low. We used bioinformatically predicted functional variants (using the MutPred family of tools) as a proxy for causative variants, and estimated the linkage disequilibrium landscape around them. The results suggested that these variants tend to be rare and have low linkage disequilibrium with surrounding variants. Taken together, these results appear to clash with the mainstream view that genomic prediction accuracy in animal breeding is driven by clusters of correlated segments in long-range linkage disequilibrium. Potentially, these are rare functional variants, that are not being captured by genomic prediction, and are in the process of being lost.

DETECTION OF QTL FOR GLOBAL RECOMBINATION RATE IN FLECKVIEH CATTLE

Kadri, Naveen; Pausch, Hubert

ETH, Zurich

Detection of QTL for global recombination rate in Fleckvieh cattle Naveen Kadri and Hubert Pausch Animal Genomics, ETH Zurich, 8092 Zurich, Switzerland Recombination between homologous chromosomes during meiosis plays a key role in gametogenesis and thus fertility. It also contributes to increasing genetic diversity by creating new allelic combinations. By examining phased genotypes for more than 38,000 genome wide SNPs in 429,265 Fleckvieh cattle using LINKPHASE3, we identified 10,181,122 crossovers in 392,753 male gametes and 2,362,152 cross overs in 102,266 female gametes. Consistent with previous reports in Bovidae, the recombination rate was higher in males than females (25.92 v/s 23.10), mainly due to an increased crossing over in the telomeric regions of autosomes in males. The number of crossovers in the autosomes (global recombination rate; GRR) was moderately heritable in both sexes (0.15 ± 0.008, 0.08 ± 0.005 in males and females respectively) and exhibited moderate genetic correlation (0.45 ± 0.04). Genome wide association testing between 28 million imputed sequence variants and mean GRR in 6,161 sires and 56,603 dams identified 13 significant ($p < 5e-8$) QTL, five (on BTA6, 10, 19 and 23) associated with GRR in both sexes and eight (on BTA1, 3, 5, 8, 9, 24, 25, and 26) associated only in females. While eight of the 13 QTL colocalized with known GRR-associated genes PRDM9 (BTA1), MSH4 (BTA3), RNF212 (BTA6), SHOC1 (BTA8), RNF212B (BTA10), MLH3 (BTA10), MSH5 (BTA23) and CEP55 (BTA23), five QTL on BTA3, 15, 19, 24, and 25 were novel. We identified SYCP1, CCDC73, CAMTA2, KATNAL2, and TRAAP as compelling positional and functional candidate genes for the five novel QTL. Our study provides additional evidence for the male biased recombination towards the telomere in cattle. We identify five novel QTL including a large effect QTL on BTA19 explaining 10% the genetic variance in both sexes. Further, the positions of more than 12.5 million cross overs identified in Fleckvieh cattle - the largest catalogue of crossover positions in a single cattle breed - is a valuable resource for further studies on genetics of crossover placement.

GENETIC REGULATION OF SINGLE-CELL PERSONAL GENE CORRELATIONS (CO-EQTLS) IS HIGHLY ENRICHED FOR GWAS VARIANTS

Kaptijn, Dan[1]; Losert, Corinna[2]; Korshevniuk, Maryna[1]; Oelen, Roy[2]; Consortium, sc-eQTLGen[1]; Westra, Harm-Jan[2]; van der Wijst, Monique[1]; Bonder, Marc Jan[2]; Heinig, Matthias[1]; Franke, Lude[2]

[1]Helmholtz Munich, Institute of Computational Biology, Oberschleißheim, Germany;; [2]University Medical Center Groningen, Genetics Department, Groningen, Netherlands

Genes are known to operate in regulatory networks, and recently, we have shown that genetic variants can affect these in a cell-type specific manner through so-called co-expression QTLs (co-eQTLs). These co-eQTLs may help in interpreting the upstream regulators of genetic variants that have been associated with disease. However, identifying co-eQTLs requires large sample sizes which limited our previous study. Here we leverage the single cell datasets consolidated by the sc-eQTLGen consortium to extend those analyses. Within the sc-eQTLGen consortium we conducted the largest cell type eQTL mapping to date in PBMC (n-donors 2,000). The first results of the sc-eQTLGen consortium identified in total 14,749 eQTLs (FDR 10%) for the six major PBMC cell types corresponding to 7,310 eGenes and 2,810 unique cell type eQTLs. Building on these results we mapped co-eQTLs on 4 in-house datasets, processed with the sc-eQTLGen consortium pipeline, which includes 1,188 PBMC samples. We identified co-eQTLs across the six major PBMC cell-types. In the largest dataset (oneK1K) we observed 1,045 co-eQTLs for CD4+ T cells involving 487 unique genes and 157 unique SNPs. Associated genes were enriched in coherent functions. Furthermore, co-eQTL SNPs were approximately twice as often associated with complex traits and diseases (from the GWAS catalog) as eQTL SNPs not showing a co-eQTL effect (p=0.001). The other datasets showed high agreement of the identified co-eQTLs (Rb>0.67). Our results indicate that there are wide-spread personal and cell-type specific differences in genetic regulation. Application of our pipeline to all datasets of the sc-eQTLGen Consortium will enable further validation and extension of identified co-eQTLs laying the foundation for building cell-type specific gene regulatory networks. The GWAS enrichment results suggest that co-eQTLs can be useful for the interpretation of disease mechanisms.

INCORPORATING PRIOR BIOLOGICAL INFORMATION INTO GENOMIC PREDICTIONS: AN EXAMPLE FROM MASTITIS IN DANISH JERSEY AND NORDIC RED CATTLE

Karaman, Emre[1]; Cai, Zexi[2]; Chegini, Arash[3]; Lidauer, Martin[4]; C.M. Moreira, Gabriel[1]; K. Chitneedi, Praveen[2]; Janss, Luc[3]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus C, Denmark; [2]Natural Resources Institute Finland (Luke), Jokioinen, Finland; [3]Unit of Animal Genomics, GIGA Institute, University of Liège, Liège, Belgium; [4]4Research Institute for Farm Animal Biology (FBN), Dummerstorf, Germany 5https://www.bovreg.eu/project/consortium/

The assumption of the same normal distribution for all genetic markers (genome-wide single-nucleotide polymorphisms, SNPs) in a standard genomic prediction analysis may be violated for some of the complex traits, where certain regions may contribute more to the trait's phenotypic variation among individuals. A widely adapted approach for prioritizing functionally relevant genomic regions in genomic predictions has been to update standard SNP panels (e.g. the Illumina's Bovine 50K SNP panel) that are used in official genomic evaluations of cattle, with additional SNP markers. Adding new SNPs to such panels for each new discovery, however, may not be feasible. In this study, a strategy for incorporating prior biological information into genomic predictions was investigated, which does not require extending SNP panels, but requires linking the SNPs in the standard 50K panel to those discoveries. As a proof of concept, this strategy was applied to genomic prediction of mastitis in Danish Jersey (JER) and Nordic Red dairy cattle (RDC). In total, 9,939 JER and 34,394 RDC cows with de-regressed proofs and genotypes were available. Analyses were carried out separately for each breed. Cows born in 2017 or before were assigned to reference populations (JER: 8,737, RDC: 31,101), while cows born in and after 2018 were assigned to validation populations (JER: 1,202, RDC: 3,293). Prior information was available from genomics (QTL), transcriptomics (eQTL), metabolomics (mQTL) and chromatin accessibility (ATAC-seq). The standard genomic best linear unbiased prediction (GBLUP) method using 50K SNP panel without prior information was used for benchmarking a Bayesian whole-genome regression method (Bayesian log-variance method, BayesLV), with which the prior biological information was used. This novel approach improved reliabilities for JER but not for RDC. Using all information available (QTL, eQTL, mQTL, ATAC-seq), the reliabilities were improved as much as about 14% for JER (0.187 vs 0.164), whereas for RDC, the reliabilities were slightly reduced (0.140 vs 0.145). When beneficial, standard analysis pipelines in official genomic evaluations can thus be extended to include "weights" for SNPs, with the weights obtained from methods considering prior biological information, still using the same standard 50K SNP panel.

THE IMPACT OF SPONTANEOUS MUTATION ACCUMULATION ON QUANTITATIVE VARIATION IN A MAMMALIAN SPECIES

Keightley, Peter

Spontaneous mutations are the source of genetic variation for quantitative traits, and fuel evolutionary adaptation and the response to artificial selection. The new variation arising from mutation for a quantitative trait is usually expressed as the mutational heritability, h^2_M, the increase in heritability from one generation of spontaneous mutation. Spontaneous mutations also introduce heritable changes that tend to reduce the mean fitness of populations. In natural populations, this erosion of fitness is countered by selection, which keeps deleterious mutations at low frequencies and ultimately removes most of them. The classical way of studying the impact of spontaneous mutations is via mutation accumulation (MA) experiments, where lines of small effective population size are bred for many generations in conditions where natural selection is largely removed. MA experiments in microbes, invertebrates and

plants have generally demonstrated that fitness decays as a result of mutation accumulation. Based on between-MA line divergence, estimates of h^2_M for quantitative traits are typically around 10^-3, i.e. the heritability increases by about one-thousandth each generation. However, the phenotypic consequences of mutation accumulation in vertebrates are largely unknown, because no replicated MA experiment has previously been carried out. This gap in our knowledge is relevant for human populations, where societal changes have reduced the strength of natural selection, potentially allowing deleterious mutations to accumulate. Furthermore, there is scant information on the magnitude of h^2_M for quantitative traits in vertebrates, he only information coming from small scale selection and inbreeding experiments in mice, which suggest h^2_M could be as high as 10^-2. In this study, we carried out the first MA experiment in a mammal using a genetically characterised inbred mouse strain with multiple MA lines maintained by full-sib mating. The experiment also includes a contemporary comparison of the MA lines with a cryopreserved control that had undergone minimal mutational accumulation, a feature that allows us to attempt to distinguish genetic from environmental change. We studied the impact of spontaneous mutation accumulation on the mean and genetic variation for several quantitative and life history traits and gene expression in 55 MA lines maintained for 21 generations by brother-sister mating. We also sequenced individuals from the MA lines using Illumina and PacBio technologies. Results are presented on rates of single nucleotide mutations, mutation spectra, and the new variation arising from insertion-deletion mutations and transposable element movements. We integrate these molecular estimates of the mutation rate with estimates of the new heritability for quantitative traits (including gene expression) and the rate of erosion of fitness per generation from spontaneous mutation accumulation. We use these results to predict the amount of new variation for quantitative traits in humans and other vertebrates, and predict the rate of fitness loss from spontaneous mutation accumulation in humans, and determine whether this should be of concern in the foreseeable future.

DIRECT, INDIRECT AND EPIGENETIC EFFECTS IN FAMILIES

Krätschmer, Ilse; Mahmoudi, Mahdi; Hofmeister, Robin J.; Delaneau, Olivier; Mägi, Reedik; Robinson, Matthew R.

ISTA

An individual's phenotype reflects a complex interplay of the direct effects of their DNA, epigenetic modifications of their DNA induced by their parents, and indirect effects of their parents' DNA. We first show how these direct, epigenetic imprinting, maternal and paternal genetic effects are confounded in theory and simulations for child-mother-father trios. Modelling sibling differences also fails to separate direct from imprinting and sibling genetic effects. We then present a Bayesian framework that is able to estimate the independent contribution of each of these genetic components to phenotypes by taking into account all the components and their correlations. Using trios in the Estonian Biobank, we find a significant partitioning of variance into direct, parental and imprinting variation in adult height, BMI, cardiovascular disease and high blood pressure. Separating

the indirect nurturing from genetic and epigenetic effects is key to understanding genotype-phenotype associations in the human population.

Marker Effect P-Value for Large Genotype Populations with the Algorithm for Proven and Young

Leite, Natalia[1]; Bermann, Matias[1]; Tsuruta, Shogo[1]; Misztal, Ignacy[2]; Lourenco, Daniela[2]

[1]University of Georgia, USA,; [2]Topigs Norsvin, the Netherlands

The single-step method (ssGBLUP) allows for simultaneous evaluations of populations composed of genotyped and non-genotyped animals. As a GBLUP-based method, the output of ssGBLUP is breeding values. However, interest might rely on the estimation of SNP effects and on how genome segments are associated with the phenotype of interest. When that is the case, SNP effects can be obtained from a linear transformation of breeding values, and P-values can be used as a measure of estimation certainty. P-values are derived from the prediction error variance of SNP effects, which relies on the prediction error (co)variance matrix of breeding values obtained from the inverse of the left-hand side of the mixed model equations (LHS). However, inverting the LHS becomes unfeasible with large, genotyped populations. Therefore, alternative strategies to overcome this problem and obtain p-values with increasing genotyped populations might be evaluated. In this study, we aimed to estimate marker effect p-values when the algorithm for proven and young (APY) was used to approximate the predictor error (co)variance for animals in a large, genotyped population. We estimate P-values of SNP marker effects for PWG in an Angus cattle population. The dataset was composed of around 844K phenotyped animals, 450K genotypes, and 1.8M pedigree records. Analyses were split into two sets. In the first set, a reduced genotype data of 50K was used, so p-values obtained from a direct-inverse of LHS using the regular G-1 (Exact_G) or APY G-1 (Exact_GAPY) were compared with APY G-1 and an approximation of the prediction error variance that did not require the inversion of the LHS (Approx_GAPY). In the second set of analyses, P-values with Approx_GAPY were estimated with the full genomic set of 450K (Approx_GAPY_450K), and computational requirements were recorded. All analyses were performed in three replicates. In the first set of analyses, genome-wide association with Exact_G uncovered two significant regions in chromosomes 7 and 20. Along all replicates, the same regions were also identified with Exact_GAPY and Approx_GAPY, indicating that similar resolutions are obtained with exact and approximated p-values. Results from Approx_GAPY_450K showed that, with a complete genome set, besides the two regions in chromosomes 7 and 20, two new regions in chromosomes 6 and 14 were identified. On average, the entire procedure for obtaining P-values with 450K genotyped animals had an elapsed wall clock time of 24h with a maximum memory usage of 87.6 GB. Altogether, our results suggest that with APY and the approximation of the prediction error variance, current computational boundaries for obtaining marker effect P-values for large genotyped populations should be lifted.

MULTI-ANCESTRY GENOME-WIDE ASSOCIATION STUDY META-ANALYSIS DECIPHERS THE GENETIC ARCHITECTURE OF MALE FERTILITY IN PIG

Lin, Qing[1]; Zhong, Zhanming[2]; Chen, Hong[3]; Cai, Xiaotdian[4]; Xu, Zhiting[5]; Zhang, Xiaoke[1]; Chen, Xinyou[2]; Ningchao, Ningchao[3]; Li, Jiaqi[4]; Speed, Doug[5]; Wang, Qishan[1]; Zhao, Yunxiang[2]; Fang, Lingzhao[3]; Zhang, Zhe[4]

[1]College of Animal Science, South China Agricultural University;; [2]Center for Quantitative Genetics and Genomics, Aarhus University;; [3]College of Animal Sciences, Zhejiang University;; [4]College of Animal Science and Veterinary Medicine, Shandong Agricultural University;; [5]Animal Science and Technology college

[Introduction] Male fertility is one of the most important reproductive phenotypes in human and livestock. Breeding boars was an essential biomedical model to investigate the genetic basis of male fertility because of the accessibility of genotypes and phenotypes from artificial inseminations. Genome-wide association studies (GWASs) yielded large-scale phenotypes-associated variants, while most of them located in the non-coding regions. The functional information resource, such as FAANG and FarmGTEx project, provides an insight to interpret the regulatory mechanism of complex traits in farm animals. However, the interpretation of male fertility was not clear. [M&M] Here, to exploit the genetic architecture of male fertility, we collected multi-ancestry populations (including 14,210 pigs with 1,156,029 records) to conduct genome-wide association study (GWAS) and meta-analysis for 6 semen traits (including semen volume, sperm abnormality, sperm concentration, sperm motility, number of sperms and number of motile sperms). We then utilized the functional genomics data, including chromatin states, to annotate the significant signals. In addition, we integrated the PigGTEx resource with GWAS meta-analysis to identify the potential candidate genes. We further incorporated the functional genomics and the integrative results to prioritize the candidate genes associated with male fertility. Finally, we exploited the correlation between male fertility and other complex phenotypes in human. [Results] We performed 90 individual GWASs for 15 populations with 6 semen traits. And then, we implemented GWAS meta-analysis. The GWAS meta-analysis for male fertility yielded 105 significant loci, corresponding 141 independent association signals. For the functional annotation, we found that the trait-associated variants enriched in the regulatory elements including the promoters and transcribed genes, which indicated that the regulatory elements played an important role in male fertility. In addition, we found that the significant associated loci of male fertility were correlated with the neural and early developmental tissues, such as blastocyst and frontal cortex. Furthermore, the integrative analysis with PigGTEx resource found that some of candidate genes was associated with male fertility, such as NAXE and ACSM5. [Conclusion] These results decipher the genetic basis of male fertility in pigs, and provides an essential link between human and pig in male fertility.

X-QTL MAPPING USING MULTI-PARENT SYNTHETIC POPULATIONS IS POWERFUL AND EFFICIENT CONDITIONAL ON EXPERIMENTAL DESIGN.

Long, Anthony D.[1]; Macdonald, Stuart J[2]

[1]Ecology and Evolutionary Biology, 32Steinhaus Hall, University of California, Irvine, CA 92697, USA; [2]Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045, USA

We are exploring a bulked segregant/X-QTL approach for dissecting complex traits in Drosophila melanogaster. Prior work constructed the Drosophila Synthetic Population Resource (DSPR; www.flyrils.org) via diallel crosses of two sets of eight highly inbred Drosophila strains, followed by 50 generations of random mating to break up haplotype blocks, and the generation of >1600 RILs via brother-sister mating. In this work 663 DSPR RILs from the eight-way "A-half" of the DSPR were mixed to re-create the initial outbred synthetic population with un-recombined founder block sizes averaging 3cM in length (i.e., a 20-fold genetic map expansion). X-QTL experiments sample several replicate pools of 500 or more control versus selected flies from this synthetic population, with the selected proportion ranging from 2-10%, and each resultant pooled-DNA sample Illumina sequenced to 40X-100X. Key differences between our approach and traditional X-QTL analysis are that our segregant pools are both highly recombined and derived from an eight-way cross as opposed to being the F2 of a cross between two parental haplotypes. In order to powerfully identify QTL it is essential to use raw SNP frequencies to impute founder haplotype frequencies, and then statistically test for localized shifts in underlying imputed haplotype frequencies. In contrast, genome scans using directly ascertained SNPs are far less powerful, as the errors on directly ascertained frequencies are larger than those on imputed frequencies. Although perhaps counter-intuitive, despite per pool sequence coverage often being on the order of one tenth the number of individual alleles contributing to a pool, the large pool size is essential and is exploited via the haplotype-imputation procedure. Especially for traits that can "self-select", extremely large replicated X-QTL experiments are achievable. Provided those experiments employ a large number of individuals per pool, an X-QTL approach can be considerably more powerful and allow for greater mapping resolution than the very best RIL-based mapping studies. We have successfully used the X-QTL method to dissect three complex traits (caffeine-, malthion-, and zinc-oxide-resistance), and several other experiments are currently in progress. We will present analytical and experimental results illustrating the utility of X-QTL mapping in flies.

EFFICIENT AND ACCURATE DETECTION OF VIRAL SEQUENCES AT SINGLE-CELL RESOLUTION REVEALS NOVEL VIRUSES PERTURBING HOST GENE EXPRESSION

Luebbert, Laura[1]; K. Sullivan, Delaney[2]; Carilli, Maria[3]; Eldjárn Hjörleifsson, Kristján[1]; Viloria Winnett, Alexander[2]; Chari, Tara[3]; Pachter, Lior[3]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California;; [2]UCLA-Caltech Medical Scientist Training Program, David Geffen School of Medicine, University of California, Los Angeles;;

[3]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California

More than 300,000 mammalian virus species are estimated to cause infectious disease in humans. They inhabit a wide range of human tissues, including the lungs, blood, and brain, and often remain undetected. Efficient and accurate detection of viral infection is vital to understand its impact on human health, and to make accurate predictions to limit negative effects, including the prevention of future epidemics. The increasing use of high-throughput sequencing methods in research, agriculture, and healthcare provides an opportunity for the cost-effective surveillance of viral diversity and investigation of virus-disease correlation. However, there are no existing workflows for accurate real-time detection of viral infection from sequencing data. We introduce a method that accurately and rapidly detects viral sequences in bulk and single-cell transcriptomics data, enabling the detection of ongoing infection by RNA viruses covering up to 10^12 virus species.

CAUSAL INFERENCE FOR MULTIPLE RISK FACTORS AND DISEASES FROM GENOMICS DATA

Machnik, Nick; Mahmoudi, Mahdi; Kratschmer, Ilse; Bauer, Markus J.; Robinson, Matthew R.

Institute Of Science and Technology Austria, Am Campus 1, Klosterneuburg, 3400, Austria, Boehringer Ingelheim RCV GmbH & Co KG, Dr. Boehringer-Gasse 5-11, Vienna, 1120, Austria.

Causal learning in genomics relies on Mendelian Randomization (MR) as the standard approach. A prerequisite for MR are valid instrumental variables, which are traditionally obtained from GWAS results. However, as GWAS estimates are made without controlling for LD and a wide-range of other confounding factors, and lack trait-specificity due to linkage, pleiotropy and population-level factors such as assortative mating, they are unlikely to ever represent a valid instrument variable. Selection of invalid instruments, combined with the high false positive rates of the most common MR methods, has created the current explosion of MR research papers. Here, we demonstrate that by following a graphical inference approach, one can learn a graphical model of genetic variables and medical traits that can greatly improve MR results by: i) explicitly testing the instrumental-variable assumptions and favouring trait-specific genetic variants, and ii) very accurately distinguishing direct from indirect effects between traits. We show that in high-LD datasets, our CI-GWAS model outperforms mvSuSiE fine-mapping performance in both power and precision and more accurately fine-maps trait-specific variants. Utilising these variants in multi-variable MR approaches gives increased precision and lower error in finding causal effects and their sizes, as compared to using GWAS selected variants. We apply CI-GWAS to 17 traits and 8.2M variants recorded for 500k individuals in the UK-Biobank and find 544 likely trait-specific causal variants and a limited array of plausible causal relationships between the traits.

INCREASING POWER IN ASSOCIATION MAPPING WITH GENETIC REPLICATES BY RECOGNIZING VARIANCE HETEROGENEITY AND EXPLORING IMPLICATIONS OF NEAR ZERO-VARIANCE

McGee, Teresa[1]; Ashner, Marissa[2]; Corty, Robert[1]; Xie, Jialiu[2]; Valdar, Will[3]

[1]Department of Genetics, University of North Carolina at Chapel Hill;; [2]Department of Biostatistics, University of North Carolina at Chapel Hill ;; [3]Department of Genetics, University of North Carolina at Chapel Hill

Modern quantitative trait locus (QTL) mapping in populations with genetic replicates uses a linear mixed model to test for SNP-phenotype association. Existing procedures assume equal variance, i.e. that the phenotype of each strain (or individual) is known with equal precision; however, this assumption does not always hold. We propose a method, weighted Inbred Strain Association Mapping (wISAM), which accounts for heteroscedastic residual variance in the study population using a strain variance-weighted regression technique. We adapt an empirical Bayes variance shrinkage method to stably estimate these weights. We compare wISAM with existing methods and demonstrate that it can provide additional statistical power for QTL mapping in the presence of heteroscedasticity. We also apply wISAM to a study on the genetic basis of drug metabolizing enzyme abundance in the Collaborative Cross mouse panel. We demonstrate how the naive use of wISAM can introduce problems for the analysis of protein abundance data, specifically when there is a spike in the phenotypic distribution resulting in all individuals of a strain having the same phenotypic value (e.g., zero-inflated phenotypes). We explore why such zero-variance phenotypes can cause problems in QTL mapping, despite traditional shrinkage methods and explore practical solutions, including two step QTL analyses and our empirical bayes variance shrinkage approach. ¬

THE EFFECTS OF A MISSING FRACTION ON SELECTION IN ADULT SIZE TRAITS IN A WILD POPULATION

Mittell, Elizabeth; Pemberton, Josephine; Kruuk, Loeske; Morrissey, Michael

University of Edinburgh and University of St Andrews

In evolutionary quantitative genetics, the missing fraction problem can cause bias in estimates of selection of traits expressed later in life. The missing fraction problem occurs when there is (1) viability selection early in life and (2) correlation between traits important for early-life survival and traits expressed later in life. Although these conditions may well be common in wild populations, the problem has received little empirical attention – possibly because it can appear intractable, given that it is impossible to measure phenotypes that were never expressed. However, it is not impossible to correctly measure lifetime selection, or correctly predict evolutionary trajectories, of later-life traits in the presence of the missing fraction. Here we present two methods that can be used to address the missing fraction and apply them to empirical data from a wild population of Soay sheep on the islands of St Kilda, Scotland. First, we show that there are genetic signatures of prior viability selection in genetic correlations between first year survival and several adult size traits. Second, we explore the causal drivers of these genetic signatures using phenotypic episodes of selection

analyses. Our analyses indicate effects of early-life selection on adult size traits, and hence that the missing fraction may substantially affect estimates of selection and predictions of evolutionary responses within this wild population.

EXPLOITING PROGENY VARIANCES FOR SELECTION DECISIONS IMPROVES GENETIC GAIN AND VARIANCE IN GENOMIC BREEDING PROGRAMS

M. Niehoff, Tobias A.; Napel, Janten; L. Calus, Mario P.

Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg 1, 6700AH Wageningen, the Netherlands

Selecting animals based on breeding values aims at maximizing the average performance of the next generation. If the aim is to maximize the average performance of the grand offspring generation, selection criteria that weigh the breeding value with the gametic Mendelian Sampling variance (MSV) of animals could be used as proposed by multiple authors. We aimed to design criteria that not only consider the gametic MSV of animals but also the MSV of their offspring, grand offspring, great-grand offspring and so forth. This extends the planning horizon at the time of selection by one or more generations. In order to devise selection criteria that look further ahead, we first developed analytical equations that enable to predict the genetic variance among descendants of certain animals based on allele effects, recombination frequencies and phased genotype states. Since allele effect sizes are estimated in real breeding programs and are influenced by errors, we tested our selection criteria in a generic simulated swine breeding program with daily gain as an example trait. We used criteria that look 1, 2 or 3 generations ahead next to ordinary truncation selection based on estimated breeding values. All criteria were tested with three different training population sizes to mimic differences in genomic prediction accuracy. The highest benefits were observed in scenarios in which the training population was largest, i.e., the genomic prediction accuracy was highest. In addition, the benefit of considering MSV of future generations was higher with longer planning horizons of our criteria. For example, compared to selection based on breeding values, planning 1, 2 or 3 generations ahead with the highest prediction accuracy resulted in 0.8%, 1.6% and 2% higher genetic levels and up to 6%, 12% and 20% higher genetic variances, respectively. At the same time, more beneficial alleles were retained in the population relative to the same number in the breeding value selection scenario. When looking 3 generations ahead, the inbreeding rate over 20 generations was only slightly higher with 1.39% than the inbreeding rate observed for the breeding value selection scenario (1.35%). In all scenarios exploiting MSV, the population average of breeding values in early generations was lower than with selection on estimated breeding values. However, the commercial genetic level, i.e., the level of boars used to breed finisher pigs, was at least as high as when selecting based on breeding values and never lower in all generations, even in scenarios with the lowest prediction accuracies. All MSV considering criteria also retained significantly more genetic variance regardless of the prediction accuracy. In conclusion, the benefit of considering MSV in selection decisions may be small and dependent on the genomic prediction accuracy but we did not observe any apparent negative side-

effects. For implementation, breeding programs require a genetic map, estimated marker effects and phased genotypes which are readily available for commercial livestock species.

STRATEGIES TO IMPROVE SELECTION COMPARED TO SELECTION BASED ON ESTIMATED BREEDING VALUES

Pook, Torsten[1]; Hassanpour, Azadeh[2]; Niehoff, Tobias[1]; Calus, Mario[2]

[1]Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg 1, 6700AH Wageningen, the Netherlands #; [2]University of Goettingen, Department of Animal Sciences, Animal Breeding and Genetics Group, Albrecht-Thaer-Weg 3, 37075, Goettingen, Germany

Selecting animals based on estimated breeding values (EBVs) is shown to maximize short-term response to selection when using random mating. However, when the aim is not only about population-wide genetic gain but about obtaining a small number of high-performance individuals (breeding nucleus) or multiple generations are considered, an optimal strategy is not so clear-cut, as the maintenance of genetic diversity may become an important factor. In this context, various strategies from optimum contribution selection to mate allocation to breeding for rare alleles have been proposed to reduce inbreeding, increase genetic diversity and selectively mate animals. Our aim was to first evaluate several strategies to maintain genetic diversity next to each other, and then combine them into a joined optimal strategy. To assess the efficiency of different strategies against each other, stochastic simulations using the software MoBPS were conducted. Applying a weighting factor on the estimated allele effects based on the frequency of the beneficial allele (1/p3) resulted in an increase of the long-term genetic gain of 0.27 genetic standard deviations (gSD) after 50 generations (+1.1%) while reducing inbreeding rates by 22% compared to selection according to traditional EBVs. In the short term, there were small losses of up to 0.05 gSD. Putting 15% of the index weight on the average kinship to top animals led to no meaningful short-term losses (-0.02 gSD) while reducing inbreeding rates by 20% and yielding 0.80 gSD more long-term genetic gain (+3.1%). Lastly, we considered a novel mate allocation strategy where we first calculate the expected inbreeding of all potential matings and subsequently sample the matings to perform from those potential matings with expected inbreeding below a certain threshold. Using the 50% quantile of all expected inbreeding values as a threshold yielded 0.33 gSD more long-term genetic gain while reducing inbreeding rates by 25% and causing practically no short-term losses (-0.02 gSD). To optimally combine different strategies, an evolutionary algorithm was used to estimate optimal weights for each considered strategy. As a breeding objective, the genetic gain after each generation was considered equally up to 50 generations to emphasize both short and long-term gains. The finally obtained optima suggests a selection index with 71% weight on traditional EBVs and 29% of the weight split between various of the previously suggested diversity characteristics. Among others, results suggest putting 6% on the average kinship to top individuals and 12% on estimated allele effects weighted by allele frequencies. Additionally, matings between animals with high expected

inbreeding were avoided (threshold: 55% quantile). To capitalize on the additional retrained diversity, the selection intensity was increased by 24%. This resulted in short-term genetic gains of up to 6.0% and long-term gains of 1.08 gSD (+4.1%) while still reducing inbreeding rates by 6%. This indicates that although fewer animals were selected these still carry more variation. Results show that a combination of breeding strategies for the management of genetic diversity will be more efficient than a single one. Using a combination of strategies not only helps with long-term genetic progress but also allows for higher short-term genetic gain by allowing for increased selection intensities without higher inbreeding rates.

DYNAMIC GENETIC REGULATION OF GENE EXPRESSION IN HETEROGENEOUS DIFFERENTIATING CULTURES

Popp, Joshua M[1]; Rhodes, Katherine[2]; Jangi, Radhika[3]; Li, Mingyuan[4]; Barr, Kenneth[5]; Tayeb, Karl[6]; Battle, Alexis[7]; Gilad, Yoav[5]; Popp, Joshua M[6]; Rhodes, Katherine[7]; Jangi, Radhika[2]; Li, Mingyuan[3]; Barr, Kenneth[4]; Tayeb, Karl[2]; Battle, Alexis[3]; Gilad, Yoav[4]

[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore MD (JMP, AB);; [2]Department of Medicine, University of Chicago, Chicago IL (KR, KB);; [3]Department of Biology, Johns Hopkins University, Baltimore MD (RJ, ML);; [4]Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago IL (KT);; [5]Department of Computer Science, Johns Hopkins University, Baltimore MD (AB);; [6]Department of Genetic Medicine, Johns Hopkins University, Baltimore MD (AB);; [7]Department of Human Genetics, University of Chicago, Chicago IL (YG)

Identifying the molecular effects of human genetic variation across cellular contexts is crucial for understanding the mechanisms underlying complex traits, yet existing datasets have focused predominantly on healthy adult tissues. One key limitation is that many contexts, especially those arising during dynamic processes such as development, are inaccessible in typical human samples. In vitro cell cultures have offered access to dynamic regulatory effects, but are typically limited to one differentiating cell-type or dynamic context per experiment. In this study, we introduce heterogeneous differentiating cultures (HDCs), a class of in vitro models that can be used to explore diverse cellular contexts efficiently. Here, we focus on unguided HDCs, which are based on embryoid body systems using an extended culturing time to consistently generate dozens of cell-types derived from all three developmental germ layers. We collected single-cell RNA-sequencing data from over 900,000 human cells generated from 53 iPS lines from unrelated Yoruba individuals from Ibadan, Nigeria. We then employed a series of latent variable modeling techniques to characterize how the impacts of genetic variation on gene expression vary with respect to diverse cell-types, pseudo-temp stages, and gene programs, many of which have not been characterized at the population level in humans. We called expression quantitative trait loci (eQTLs) in 29 distinct cell-types, which revealed novel regulatory effects for hundreds of genes. Notably, these novel eQTLs were

enriched among genes crucial for developmental pathways, such as those implicated in cardiovascular and central nervous system formation. Despite collecting data at a single time-point, asynchronous differentiation enabled us to additionally reconstruct dynamic trajectories along the three germ layers and identify dynamic eQTLs in each. And finally, we applied topic modeling to study regulatory variability imposed by cellular programs such as the cell cycle and ciliary activity which are not compartmentalized by cell-type or temp stage, and demonstrate how these programs additionally modulate the impacts of genetic variation. We identified dozens of genes with novel interaction eQTLs (ieGenes) overlapping at least one GWAS locus with no previously reported regulatory function. Among these findings were a schizophrenia-associated locus influencing the neurodevelopmental transcription factor BCL11B and a variant linked to appendicular lean mass affecting the expression of COL1A2 during muscle cell differentiation. Our study uses latent variable modeling to examine the dynamic landscape of genetic regulation of gene expression in a uniquely flexible in vitro model system, setting the stage for accelerated discovery of disease mechanisms and a more comprehensive characterization of the molecular architecture of complex traits.

IMPROVING THE PREDICTION OF NON-ADDITIVE EFFECTS WITH HIERARCHICAL GENOMIC PREDICTION MODELS

Powell, Owen[1]; McLean, Greg[2]; Brider, Jason[3]; Saddigh, Joe[1]; Technow, Frank[2]; Tang, Tom[3]; Totir, Radu[1]; Messina, Carlos D[2]; Hammer, Graeme[3]; Cooper, Mark[3]

[1]Queensland Alliance for Agriculture and Food Innovation (QAAFI), Centre for Crop Science, The University of Queensland, Brisbane, QLD, Australia;; [2]The University of Queensland, ARC Centre of Excellence for Plant Success in Nature and Agriculture, Brisbane, QLD, Australia;; [3]Corteva Agriscience, Johnston, IA, USA; 4 - University of Florida, Department of Horticultural Sciences, Gainesville, FL, United States

Maximising response to selection for complex traits in breeding programs requires accurate estimates of the average effects of allele substitutions. However, biological interactions underpinning complex trait variation generate context-dependencies across different genetic backgrounds (GxG) and environments (GxE) which can reduce the accuracy of estimating average effects. Despite this, crop and livestock improvement programs continue to be optimised around genomic prediction methods that prioritise main effects and marginalise complex trait variation due to interaction effects. The development of methods with improved predictive ability of genetic effects (G), environmental effects (E), and their interactions (GxG and GxE) controlling complex traits would create many opportunities to unlock improvements in agricultural productivity and sustainability. Hierarchical genomic prediction models, which contain symbolic networks of biological interactions, provide a framework to: (1) improve the predictive accuracy of complex traits across diverse genetic backgrounds and production environments; and (2) reduce the contribution of the conditional interaction effects during the estimation average effects. We present a novel hierarchical genomic prediction methodology, APSIM-WGP, that

links a symbolic network of prior knowledge of GEI for sorghum (APSIM sorghum crop growth model) with a conventional genomic prediction algorithm (BayesA). We demonstrate the potential value of hierarchical genomic prediction models, in an unpublished study, by comparing APSIM-WGP to a multi-trait RKHS genomic prediction approach using a simulated multi-environmental trial dataset representative of the sorghum production regions in Australia. Improvements in prediction accuracies of grain yield up to 0.8 were driven by the ability of the symbolic network of prior biological knowledge (APSIM) to deconvolute genotype-by-environment and trait-trait interactions prior to the estimation of average effects. Finally, we demonstrate and discuss the nascent opportunities for hierarchical genomic prediction models to improve the identification of casual genomic variants controlling complex trait variation.

CAN HYBRIDIZATION ALLOW THE EMERGENCE OF A SUPER-GENOTYPE IN ARABIS FLOODPLAIN SPECIES?

Rahnamae, Neda[1]; Özoglan, Yasar[2]; Metzger, Lukas[1]; Hördemann, Lea[2]; Saboor Khan, Abdul[1]; Ali, Tahir[2]; Tellier, Aurélien[1]; de Meaux, Juliette[2]

[1]Institute of Plant Sciences, University of Cologne, Cologne, Germany.; [2]Department of Life Science Systems, Technical University of Munich, Freising, Germany

Deciphering the genetic basis of ecological differences among hybridizing species is crucial for predicting their adaptive responses to climate change and human activities. Previous works identified a hotspot of hybridization on the banks of the Rhine River, revealing episodic gene flow between the close relatives Arabis nemorensis and A. sagittata. We genotyped a large interspecific F2 population (ca. 1000 individuals) resulting from the cross between sympatric individuals of these species, generating a high-density genetic map across 8 linkage groups. Quantitative trait loci (QTL) mapping for 24 traits revealed over 50 QTLs scattered along the genome, with flowering time exhibiting the strongest effect size QTL. Five QTLs indicating fertility trait incompatibilities. This study enhances our understanding of the genetic architecture of these species and provides insights into the potential existence of a "super genotype" capable of navigating complex ecological challenges in the presence of gene flow. Keywords: RAD sequencing, genetic map, QTL mapping, adaptation, introgression.

PREDICTION OF VARIANT EFFECTS BY FOUNDATION AI MODELS: IN VIVO VALIDATION AT NUCLEOTIDE AND HAPLOTYPE RESOLUTION IN PLANT POPULATIONS

Ramstein, Guillaume P[1]; Song, Baoxing[2]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark;; [2]Institute of Advanced Agricultural Sciences, Peking University, Shandong, China

In plant breeding, genomic technologies have been useful to rapidly screen populations by genomic selection (plant selection by DNA information), and make precise changes by precision breeding (targeted mutagenesis of plant

genomes). However, these applications are still limited by our ability to detect the most impactful genetic variants. Associations inferred by traditional quantitative genetics methods are useful to identify genetic effects, but they cannot point to the exact causal variants responsible for these effects. Artificial Intelligence (AI) models have introduced alternative methods to predict variant effects from biological sequences. Importantly, the latest generation of 'foundation' AI models, trained on large datasets across many species, should be applicable to plant species without expensive and time-consuming re-training. However, it is still unclear whether the accuracy and resolution of their predicted effects are high enough for genetic intervention on haplotypes (gene deletion/replacement/selection) or nucleotides (base editing). In this presentation, I will present methodologies to predict variant effects by foundation AI models, and report on their in vivo validation in different plant populations. In Brachypodium distachyon (a model species for cool-season grasses), we predicted phylogenetic conservation of nucleotide alleles by a protein language model (ESM-1v), and validated the effects of phylogenetically rare variants on fitness-related traits at nucleotide resolution, in 848 independent mutagenized lines. Evolutionary scores by ESM-1v pointed to mutations in central metabolism and predicted variant effects more accurately than observed nucleotide conservation (SIFT) or permuted ESM-1v scores. In maize, we used the US maize pangenome to infer protein variants in 25 bi-parental families. Then, we predicted the structure of protein variants by AlphaFold2, and we validated their effects on 32 phenotypic traits, by 'proteome-wide association studies' (PWAS) and genomic prediction. Our results show that protein structure prediction by AlphaFold2, combined with a suitable metric for quantifying structural similarity between protein variants, enables higher statistical power in PWAS and higher prediction accuracy in genomic prediction, compared to sequence-based similarity among protein variants. Altogether, our results suggest that foundation AI methods can point to biological effects of protein-coding variants. Therefore, they may improve genomic prediction, for selection of complex traits, and target selection, for precision breeding, e.g., by TILLING or CRISPR. To formally test whether AI methods are applicable to precision breeding at high precision (low false positive rate), future work is needed to create lines carrying candidate causal mutations for comparison with wild-type controls.

INTEGRATING SINGLE KERNEL PHENOMIC SELECTION WITH GENOMIC SELECTION: APPLICATIONS IN CORN BREEDING.

Resende, Marcio F. R.; Graciano, Rafaela P.; Peixoto, Marco

University of Florida, Gainesville, Florida 32611.

Phenomic Selection (PS) has recently been proposed as a cost-effective method for predicting complex traits and enhancing genetic gain in breeding programs. This new technique maintains the statistical procedure used in GS-based prediction models but replaces the molecular markers data (e.g., SNP data) with variables obtained from a multi-variate phenotyping method (e.g. , near-infrared spectroscopy (NIR) data). This current study, to our knowledge, is the first to to

explore the application of PS using single kernel NIR (skNIR). We evaluated the tool in an applied sweet corn breeding program. Here, we focused on predicting field-based traits of economic importance, including ear traits and plant traits. First, on a diversity panel, three models were employed: G-BLUP and skNIR BLUP models, which utilized relationship matrices based on SNP and skNIR data, respectively, and a third model that uses both kernels associated with independent random variables. The genomic relationship matrices were evaluated when the number of SNPs used to build the matrix varied from 500 to 200,000 SNPs. In a second approach, we utilized the skNIR BLUP model trained on the diversity panel to select doubled haploid (DH) lines for germination before planting. Our findings reveal that PS generated good predictive ability (e.g., 0.46 for plant height.). Also, it effectively distinguished between high and low germination rates in DH lines. This highlights the potential of skNIR to enable the selection of DH candidates before planting. Although GS generally outperformed PS, the model combining both information (PS+GS) yielded the highest predictive ability. Furthermore, accuracies of the PS+GS model were considerably higher than GS when low marker densities were used. This indicates skNIR's potential to maintain/improve accuracy together with SNP-based information while reducing marker density, which could decrease genotyping costs in the breeding program. In conclusion, PS is a promising low-cost tool that could help to optimize the sweet corn breeding program.

WARNING: SELECTION FOR DECREASED VARIABILITY IN MILK YIELD MAY LEAD TO ASOCIAL COWS!

Rönnegård, L.[1]; Fikse, W.F.[2]

[1]Department of Animal Biosciences, Swedish University of Agricultural Sciences, Box 7023, SE-750 07 Uppsala, Sweden.; [2]School of Information and Engineering, Dalarna University, SE-791 88 Falun, Sweden.

Dairy cows have agonistic and affiliative interactions in free-stall barns that may affect an individual's milk production either positively or negatively. As for many other farm animals, these indirect effects between individuals may have a genetic component referred to as indirect genetic effects (IGEs). Competition and IGEs are known to result in a measurable genetic variation in variability [1,2] that can be estimated using a Double Hierarchical Generalized Linear Model (DHGLM) [3], DHGLMs produce estimated breeding values for variability (vEBV), and selection on vEBVs may affect the level of competition between individuals. In our study, we go one step further and investigate how selection on vEBVs may affect the number of contacts a cow has with other cows in a model including IGEs. For each scenario, 20 replicates with 100 farms with 100 cows each were simulated. Milk yield was simulated using three correlated genetic components: direct, indirect and number of contacts. The number of contacts per cow was simulated as a quantitative trait using the same social network algorithms as in [4] and was used to simulate the milk yield. The simulated milk yield was calculated as the direct genetic effect plus the sum of the IGEs over all the individuals with whom the cow had contact plus an iid residual. The simulated heritabilities were around 0.3 both for the direct and indirect genetic

effects. The simulated number of contacts a cow had was on average five with equal intensities for all contacts and the heritability for number of contacts was 0.5. A DHGLM was fitted to each replicate and the correlation between the vEBVs and the simulated true breeding values (TBVs) for number of contacts was calculated. Under a base scenario of no genetic correlations between the three simulated genetic components, the results showed a positive correlation (0.17) between the vEBVs and the TBVs for number of contacts. Consequently, selection for reduced vEBVs will reduce the number of contacts a cow has under the assumptions of the simulated model. The results were not sensitive to correlations between the three simulated genetic components. In conclusion, we have shown that sociability can be an explanation for estimated genetic variability and that careless selection for decreased variability in milk yield could potentially lead to asocial cows. References: [1] Marjanovic et al. (2018). Heredity 121: 631-647. [2] Marjanovic et al. (2022). Evolutionary applications 15: 694-705. [3] Rönnegård et al. (2010). Genetics Selection Evolution 42:1-10. [4] Fikse, et al. (2022). In Proceedings of 12th WCGALP, pp. 490-492.

INTERPRETABLE GENOMIC PREDICTIONS VIA EFFECT PROPAGATION IN GENE REGULATORY NETWORKS.

Ruzickova, Natalia; Hledik, Michal; Tkacik, Gasper

ISTA

As their statistical power grows, genome-wide association studies (GWAS) have identified an increasing number of loci underlying quantitative traits of interest. These loci are scattered throughout the genome and are individually responsible only for small fractions of the total heritable trait variance. The recently proposed omnigenic explains these observations by postulating that numerous distant loci contribute to each complex trait via effect propagation through intracellular regulatory networks. We formalize this conceptual framework by proposing the "quantitative omnigenic model'' (QOM), a statistical model that combines prior knowledge of the regulatory network topology with genomic data. The QOM predicts expression levels by explicitly modelling the propagation of genetic effects through an experimentally reconstructed transcriptional regulatory network. In particular, we describe indirect (trans) genetic effects as a result of the propagation of direct (cis) effects. This has the potential to uncover causal regulatory mechanisms while reducing the number of parameters by orders of magnitude compared to traditional GWAS-type approaches, such as the polygenic risk score (PRS). In a yeast model, the QOM reaches performance comparable to PRS, while obtaining interpretable results, thereby supporting the omnigenic hypotheses. Furthermore, we demonstrate the relevance of QOM for systems biology, by employing it as a statistical test for the quality of regulatory network reconstruction and linking it to the propagation of non-genetic, environmental effects. Last but not least, by jointly predicting expression levels of all transcription factors, the QOM allows us to explore the implications of regulatory network architecture on genetic architecture, polygenicity and the evolution of gene expression traits.

ALTERED PRIOR MEAN OF ALLELIC EFFECTS: AN APPROACH FOR ADEQUATELY CONSIDERING GENE EDITED VARIANTS WITHIN GENOMIC PREDICTIONS.

Schrauf, M.F.; Wientjes, Y.C.J.; Vandenplas, J.

Animal breeding and Genomics, Wageningen University and Research, Wageningen 6708 PB, The Netherlands

Gene editing technologies offer a potential way to introduce novel genetic variation into livestock populations. However, the initial absence of phenotypic data from individuals carrying the edited alleles poses a significant challenge for estimating the genetic effects of these alleles in the population, which can then be lost due to drift if the selection criteria is not adjusted. In this study, we explored Bayesian approaches aimed at incorporating prior knowledge on gene edited variants into genomic prediction models to facilitate their integration in selection schemes. Specifically, we focused on altering the prior means for the random allelic effects of the edited loci, a novel approach compared to previous efforts that primarily focused on altering only their prior variance (usually formulated as a reweighting of allelic effects). Concretely, we compared selection on conventional genomic estimated breeding values (GEBVs) with those obtained using altered prior variances, altered prior means, or a combination of both. The conventional prior considered for allele effects had mean zero and an equal variance for all markers. In the altered priors, we only modified the prior effect of a single marker, in full linkage with the edited locus. For the altered prior mean we used the true effect of the edited allele, and for the altered prior variance we used the contribution of the edited locus to the genetic variance in the population, in the first generation after the gene edit. We simulated 20 replicates of a population of 1000 individuals from which, every generation, 100 sires and 100 dams were selected. At the 10th generation of selection, 5 of the selected sires were randomly chosen to be gene edited on a previously monomorphic locus. The effect of the edited allele was simulated to be positive and of half the magnitude of the genetic standard deviation, which resulted in a contribution from the edited locus on the genetic variation of the selected trait of approximately 0.6% in the 11th generation. Our results show that selection on conventional GEBVs is at high risk of losing the gene edited alleles over successive generations, while selection on GEBVs with altered prior variance only improved the situation marginally. On the other hand, selection on GEBVs with the altered prior means effectively preserves the alleles and allows a fast increase of their frequency within the selected population. Notably, using these GEBVs based on altered prior means consistently leads to the edited allele becoming predominant in under 10 generations. This study highlights the potential of altering the prior means of allelic effects to integrate prior knowledge on gene edited alleles, thereby ensuring the effective utilization of gene edited variants in genomic prediction schemes, which cannot be otherwise accurately estimated. Further exploration of this approach in diverse and more complex scenarios can provide information on its practical limitations and optimal applications.

MODELLING THE GENETIC ARCHITECTURE OF COMPLEX TRAITS VIA STRATIFIED HIGH-DEFINITION LIKELIHOOD

Lan, Ao; Shen, Xia

Sun Yat-sen University (AL, XS); Fudan University (AL, XS); Karolinska Institutet (XS); University of Edinburgh (XS)

The distribution of heritability across the genome describes how the underlying genetic architecture of complex traits is shaped. Using genome-wide summary association statistics, state-of-the-art techniques analyze the heritability distribution by stratifying the genome based on functional annotations. We present a new stratified high-definition likelihood (sHDL) method, an advanced statistical model that integrates genomic functional annotation to infer complex trait genetic architecture. sHDL improves upon the stratified linkage disequilibrium score regression (sLDSC) method by offering 1.4 to 7.4-fold higher estimation efficiency for heritability enrichment parameters with reduced bias, validated through simulations and real-data analyses. For 30 traits and 133 binary annotations, we identified 178 significant heritability enrichment cases by both sHDL and sLDSC, and among the 499 discordant findings, 480 were discovered only by sHDL and 19 only by sLDSC. Incorporating gene expressions specific to cell types, the sHDL method identified additional brain cell types linked to psychiatric disorders, intelligence, educational attainment, and height. By integrating cancer epigenetic information, sHDL was also able to discern associations between complex traits related to cancer and epigenetic profiles specific to cancer. sHDL advances our understanding of polygenic contributions to complex traits, providing a robust and versatile approach for complex trait genetic analysis.

RECONCILING LINKAGE AND ASSOCIATION STUDIES OF COMPLEX TRAITS

Sidorenko, Julia[1]; Couvy-Duchesne, Baptiste[2]; E. Kemper, Kathryn[3]; Moen, Gunn-Helen[4]; Bhatta, Laxmi[5]; Olav Åsvold, Bjørn[6]; Mägi, Reedik[7]; Ani, Alireza[8]; Wang, Rujia[9]; M.Nolte, Ilja[10]; Gordon, Scott[11]; Hayward, Caroline[12]; Campbell, Archie[13]; J.Benjamin, Daniel[14]; Cesarini, David[15]; M.Evans, David[16]; E.Goddard, Michael[17]; S.Haley, Chris[18]; Porteous, David[19]; E.Medland, Sarah[20]; G. Martin, Nicholas[21]; Snieder, Harold[22]; Metspalu, Andres[23]; Hveem, Kristian[24]; Brumpton, Ben[25]; M.Visscher, Peter[26]; Yengo, Loic[26]

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia; [2]QIMR Berghofer Medical Research Institute, Brisbane, Australia; [3]Sorbonne University, Paris Brain Institute – ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France; [4]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway; [5]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; [6]The Frazer Institute, The University of Queensland, Woolloongabba,

QLD 42, Australia; [7]HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, 7600 Norway; [8]Department of Endocrinology, Clinic of Medicine, St Olavs Hospital, Trondheim, Norway; [9]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia; [10]A full list of members and affiliations appears in the Supplementary Information; [11]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands; [12]Department of Bioinformatics, Isfahan University of Medical Sciences, Isfahan, Iran; [13]MRC Human Genetics Unit, Institute of Genetics & Cancer, University of Edinburgh, Western General Hospital, Edinburgh EH2XU, United Kingdom; [14]Centre for Genomic and Experimental Medicine, Institute of Genetics & Cancer, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom; [15]5HumanGeneticsDepartment,DavidGeffenSchoolofMedicine,UniversityofCaliforniaLosAngeles,CA90095,USA; [16]Behavi Decision Making Group, Anderson School of Management, University of California Los Angeles, CA 90095, USA; [17]National Bureau of Economic Research, Cambridge, MA 0238, USA; [18]Department of Economics, New York University, New York, NY 02, USA; [19]Center for Experimental Social Science, New York University, New York, NY 02, USA; [20]0 MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8BN, United Kingdom; [21]Centre for AgriBioscience, Agriculture Victoria, Bundoora, Victoria, Australia; [22]Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Victoria, Australia; [23]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 XU, UK; [24]Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH5 9RG, UK; [25]5CouplandCraftCider,Coupland,Northumberland,UK; [26]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Family-based genetic linkage studies have been successful in detecting and mapping loci underlying single gene disorders. However, the application of this experimental design to complex polygenic traits and diseases mostly failed to deliver replicable results. In contrast, population-based genome-wide association studies (GWAS) have identified robust associations between tens of thousands of SNPs and complex traits yet capture less than half of total genetic variance. While polygenicity has been hypothesized to be the cause of the lack of replication of linkage studies, this hypothesis, so far, lacks empirical support. Here we conduct a large-scale linkage study of height and body mass index (BMI) in 119,000 sibling pairs and show that linkage signals colocalize with GWAS-identified loci. Furthermore, we demonstrate that by combining linkage and GWAS data, we can get new insights into genetic architecture of the still missing heritability of height. We show that while locus-specific linkage is reduced by adjusting phenotypes for polygenic scores, it still colocalizes with GWAS hits, suggesting that causal variants for height not captured by current

GWAS are also enriched within GWAS-detected height-associated loci. Finally, we show that recombination rate (RR) dependent genetic architecture can lead to biases in IBD-based heritability estimates and provide a simple approach, RR-stratified IBD regression, to correct these biases. We estimate heritability for height and BMI to be 0.76 ± 0.05 and 0.55 ± 0.07, respectively, consistent with the estimates obtained from pedigree-based analyses and imply that a substantial fraction thereof has yet to be accounted for by GWAS-identified loci.

TRANSPOSABLE ELEMENT ABUNDANCE SUBTLY CONTRIBUTES TO LOWER FITNESS IN MAIZE

Stitzer, Michelle C.[1]; Khaipho-Burch, Merritt B.[2]; Hudson, Asher I.[3]; Song, Baoxing[1]; Valdes-Franco, Jose Arcadio[2]; Ramstein, Guillaume P.[3]; Feschotte, Cedric[1]; Buckler, Edward S.[2]

[1]Institute for Genomic Diversity, Cornell University; Department of Entomology and Plant Pathology, NC State University;; [2]National Key Laboratory of Wheat Improvement; Peking University Institute of Advanced Agricultural Sciences, Weifang, China;; [3]Center for Quantitative Genetics and Genomics, Aarhus University; Aarhus, Denmark; USDA-ARS, Ithaca, NY

Transposable elements (TEs) have long been shown to have deleterious effects on the survival and reproduction of their host organism. As TEs are mobile DNA that jump to new positions, this deleterious cost can occur directly, by inserting into genes and regulatory sequences. Classical population genetic theory suggests copy-number dependent selection against TEs is necessary to prevent TEs from expanding so much they take over a genome. Such models have been difficult to interpret when applied to large genomes like maize, where there are hundreds of thousands of TE insertions that collectively make up 85% of the genome. Here, we use nearly 5000 inbred lines from maize mapping populations and a pan-genomic imputation approach to measure TE content. Segregating TE content gives rise to 100 Mb differences between individuals, and populations often show transgressive segregation in TE content. We use replicated phenotypes measured in hybrids across numerous years and environments to empirically measure the fitness costs of TEs. For an annual plant like maize, grain yield is not only a key agronomic phenotype, but also a direct measure of reproductive output. We find weak negative effects of TE accumulation on grain yield, nearing the limit of the efficacy of natural selection in maize. This results in a loss of one kernel (≈0.1% of average per-plant yield) for every additional 14 Mb of TE content. This deleterious load is enriched in TEs within 1 kilobase of genes and young TE insertions. Together, we provide rare empirical measurements of the fitness costs of TEs, and suggest that the TEs we see today in the genome have been filtered by selection against their deleterious consequences on maize fitness.

ISOLATING ADAPTIVE VARIATION FROM NATURAL FOREST TREES

Nurmisto, Anni[1]; Akulova, Vasilina[2]; Arizpe, Alex[3]; Slusarz, Lucyna[4]; Weidlich, Lisa[5]; Polacek, Miroslav[6]; Guevarra, Paige[1]; Riefler, Julia[2]; Steindl, Sonja[3];

Wittmann, Nora[4]; Yen, Chun-Chieh[5]; Nemeth, Krisztian[6]; Akulov, Kirill[1]; Vallebueno, Miguel[2]; Cervenka, Jaroslav[3]; Seidl, Rupert[4]; Svoboda, Miroslav[5]; Swarts, Kelly[6]

[1]Gregor Mendel Institute,; [2]Sumava National Park,; [3]Berchtesgaden National Park,; [4]Charles University, Prague,; [5]Umea Plant Sciences Center,; [6]Swedish Agricultural University

Conifers are ecologically dominant and economically important, but are succumbing to drought, disease, early-budding and other challenges globally because the climate has changed so that mature trees are no longer adapted to their local environment. If we could predict how individual tree genotypes would respond to different environments, we could — given environmental predictions — plant the right tree in the right space. While agronomic approaches such as reciprocal transplant experiments and provenance trials can effectively estimate genotypic responses to common environments, with a 40 year generation time both the number of genotypes and the number of environments that can be evaluated are limited. We introduce a new approach that takes advantage of tree increment core samples to observe annual growth from natural forest trees. For each genotype, we thus have a life-time's worth of experienced year-environments. Following agricultural models, this allows us to partition growth variation into generalizable environmental responses for years with historical information from weather stations, satellites or historical records, using ecological approaches to control for correlated responses. We focus on Norway spruce (Picea abies), native to central Europe and Scandinavia but grown for millenia across western Europe and now globally, parsing variation in annual growth from increment cores into that explained by genotype, environment and genotype-by-environment interactions (GxE). We test the approach in a genetically diverse set of 700 trees backed by hundreds of thousands of year-tree observations from a multiscalar sampling design covering 11 plots located in three national parks across Europe. Resulting GxE estimates are highly heritable, with narrow-sense heritability estimated from the kinship generally greater than 0.7 for most environments tested. These refined estimates are used to map the genetic basis of adaptive response to environment in genome-wide association studies (GWAS). We find significantly associated genomic variants, even with a relatively small population size, but, more importantly, we show that our approach limits the confounding effects of population structure compared to a naïve model designed similarly to human population studies. We also use these estimates to predict genetic responses to novel environments in a cross-population prediction framework, identifying shared trait architecture across geographically, genetically and environmentally diverse popuations. Shifting the unit of observation from single genotype to longitudinal, annual growth measurements enables us to quickly infer the genetic basis of adaptive response in any population, providing the means to evaluate a tree's performance in any modeled environment. As environments shift under climate change, this provides a powerful tool to select parents for healthy, resilient forests.

LEVERAGING INTERACTOME AND TRANSCRIPTOME TO ENHANCE GENOMIC PREDICTION IN PLANT BREEDING

Tong, Hao[1]; Nikoloski, Zoran[2]

[1]Bioinformatics, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany;; [2]Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany

Genomic prediction (GP) has revolutionized plant breeding. Modern GP models depend not only on genomic data but also increasingly integrate various omic data, such as those obtained from transcriptomic and metabolomic technologies. The interactome, describing the molecular interactions, while offering a rich source of information has not yet been used in conjunction with GP models. With existing knowledge about plant interactome, such as protein-protein interaction (PPI) networks, it is important to explore whether including these interactome can result in improved performance of GP models. Here, we present a novel approach, termed interactomeGP, designed to increase trait prediction accuracy by integrating PPI networks with transcriptomic data into GP models. The interactomeGP approach consists of four key steps: (i) constructing genotype-specific PPI networks using transcriptomic data, (ii) transforming each PPI network into a matrix using network embedding techniques, (iii) assembling a tensor matrix of all genotypes, followed by (iv) trait prediction using tensor regression models. To evaluate interactomeGP, we applied it to a diversity panel of Arabidopsis thaliana consisting of 271 accessions and a PPI network comprising 2655 interactions and 1333 genes, along with the corresponding transcriptomic data. Comparing accuracy of predicting five yield-related traits using classical GP models from transcriptomic data, our interactomeGP approach demonstrated a notable improvement in prediction accuracy. This approach also enables the identification of modules of protein-protein interactions associated with the predicted traits. The proof-of-concept study explores the potential of integrating interactomic data into GP models, providing valuable insights into refining plant breeding strategies. It highlights the utility of incorporating molecular interaction networks to enhance the accuracy of trait prediction.

IBS VERSUS IBD - NEW INSIGHTS FROM WHOLE GENOME SEQUENCE DATA

Warburton, Christie[1]; Costilla, Roy[2]; Goddard, Mike[1]; Hayes, Ben[2]; Meuwissen, Theo[2]

[1]University of Queensland, AgResearch, University of Melbourne,; [2]University of Queensland, Norwegian University of Life Sciences

An implicit assumption in most methods for genomic prediction is that identical by state (IBS) genome regions identified by high density, genome wide markers will also be identical by descent (IBD). That is, chromosome segments with identical SNP alleles will also carry the same mutation (QTL allele) affecting the complex trait. With the availability of whole genome sequence on many individuals, we can now directly test this assumption. Here we describe a methodology for doing this and test this methodology in cattle, with IBS haplotypes tested for IBD within and across breeds, and across sub-species Bos

taurus indicus (Bos indicus) and Bos taurus taurus (Bos taurus). The aim of this research is to determine the number of SNP required to accurately identify IBD haplotypes and maintain LD phase between SNP and QTL within and across genetically diverse Bos indicus and Bos taurus cattle breeds. Using the 1000 bull genomes data base, we first identify pairs of haplotypes that are IBS at SNP in the commonly used Illumina Bovine HD Array, within and across breeds. We then count the number of variant alleles that are different between the haplotypes in the sequence data, and compare this number to expectations from theory given the length of the chromosome segment, effective population size and mutation rates. Within breeds, the number of variant alleles in the sequence data that were different for 250kb IBS haplotypes were very small. Across Bos taurus breeds, this number increased slightly, and was consistent with previous estimates of the number of SNP required for multi-taurus breed predictions of 360,000. This number was also consistent with theoretical predictions given effective population size and other parameters. In comparison, in populations consisting of Bos indicus and Bos taurus breeds of cattle, our results indicate that approximately 1.5 million SNP are required to ensure 250kb IBS haplotypes are also IBS. It is also woth noting that the age of the split between indicus and taurus cattle is such that many QTL will only segreate within sub-species, and our multi-species QTL mapping efforts support this. Both these findings suggest for multi-sub species predictions in cattle, a higher density of markers is required than the current HD arrays, and predictions from whole genome sequence should be valuable.

REDUCED RANK FACTOR ANALYTIC MODELS FOR CAPTURING GENOTYPE BY ENVIRONMENT INTERACTIONS IN LIVESTOCK

Waters, Dominic L.; van der Werf, Julius H.J.; Clark, Sam A.

School of Environmental & Rural Science, University of New England, Armidale, NSW, 2351, Australia

Genotype by environment (GxE) interactions occur when the genetic correlation between a trait measured in different environments is less than one, or when the genetic variance of a trait changes between environments. This is often captured using a multi-trait model with an unstructured genetic covariance matrix, where performance in a different environment is considered as a separate but correlated trait. Such an approach becomes computationally infeasible with a large number of environments; an analysis with n=30 environments would require the estimation of n[n+1]/2 or 465 genetic parameters. Hence, we need methods that enable the estimation of GxE interactions with fewer parameters. Factor analytic models approximate the multi-trait model by assuming the pattern of GxE across environments can be described by the regression of genetic effects on latent common factors. The latent common factors are estimated from the data such that they explain the maximum amount of covariance between environments. These models are potentially more flexible and less prescriptive compared to other methods such as reaction norms commonly used in livestock genetics. This study analysed post-weaning body weights from 15,908 lambs across 31 flock-years. The flocks

were linked via common sires artificial insemination, while the years were linked via dams used across years. Each flock-year had at least 350 lambs. A reduced-rank factor analytic model with two latent common factors for the additive genetic effects and genetic group effects, respectively, provided the best fit to the data based on a log-likelihood ratio test (LRT) and the AIC. The 465 pairwise genetic correlations between environments that were derived from the factor analysis ranged between -0.69 and 1.00, with an average of 0.68. Of these, 22% were significantly less than 1, while 12% were significantly less than 0.80. An alternative approach using a reaction norm model that regressed over the mean performance was also investigated. It was unclear which model was preferred; the reaction norm was significantly poorer than the reduced-rank factor analytic models based on the LRT and AIC but were preferred based on the BIC. However, when the underlying GxE interactions are multi-dimensional, factor analytic models present appealing formulation.

CHANGES IN ALLELE FREQUENCY AND GWAS RESULTS ACROSS YEARS IN TWO PIG POPULATIONS UNDER SELECTION

Wientjes, Y.C.J.[1]; Calus, M.P.L.[2]; Bijma, P.[1]; Huisman, A.E.[2]; Peeters, K.[2]

[1]Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, the Netherlands;; [2]Hendrix Genetics B.V., P.O. Box 114, 5830 AC Boxmeer, the Netherlands

Genomic selection has been implemented in all major animal and plant breeding programs around the world, because it can vastly increase genetic gain. The implementation of genomic selection has also resulted in faster changes in the genome compared to previous selection methods. Understanding and monitoring those changes is important to get more insights into the long-term consequences of genomic selection. Our aim here was to investigate the changes in allele frequency and in the results of genome-wide association studies (GWAS) in two commercial sow lines undergoing genomic selection from 2015 to 2021. For line A, genotypes of 44,054 segregating markers were available for 2,616 to 7,689 animals per birth year, with a total of 40,075 animals. For line B, genotypes of 44,000 segregating markers were available for 921 to 4,995 animals per birth year, with a total of 23,487 animals. Phenotypes for eight traits under selection were available on a subset of the genotyped animals (Line A: 738 – 6,423 animals per trait per birth year; Line B: 406 – 3,965 animals per trait per birth year), including general production and reproduction traits. Animals were selected based on a broad selection index combining production and reproduction traits, and for all genotyped animals estimated breeding values for the index were available. Over the seven birth years included in the dataset, absolute allele frequency changes up to 0.35 were observed in each line, with a few clear peaks. The regions with the largest changes in allele frequency did, however, not overlap between lines. While the largest observed allele frequency changes were not exceeding the expectation under drift, the average change in allele frequency was larger with selection than pure drift. Several significant regions were observed in the GWAS for the traits under selection. Many of those significant regions showed pleiotropic, and often antagonistic, effects on traits

included in the index. This reduces the selection pressure on those regions, which can explain why those regions are still segregating, even though the traits have been under selection for several generations. No significant regions were identified for the selection index, indicating that the index is affected by many loci with a small effect. Across the years, only small changes in GWAS results were observed, indicating that the genetic architecture was reasonably constant. Surprisingly, no significant markers were found related to any of the traits in the regions with the largest changes in allele frequency. Moreover, the correlation between significance levels of markers and changes in allele frequency was close to zero, even for the index. Altogether, our results indicate that selection acted on a very polygenic index, which spread selection pressure across the genome and limits allele frequency change. Therefore, the impact on allele frequency changes was larger for drift than for selection.

GENOME-WIDE FINE-MAPPING IMPROVES IDENTIFICATION OF CAUSAL VARIANTS

Wu, Yang; Zheng, Zhili; Thibaut, Loic; E.Goddard, Michael; R.Wray, Naomi; M.Visscher, Peter; Zeng, Jian

The University of Queensland

Fine-mapping refines GWAS signals with the aim to identify causal variants for complex traits. However, current methods focus on genome-wide significant loci only or consider one genomic region at a time, in isolation from the rest of the genome, which may result in miscalibration and compromise power. In addition, current strategies do not inform the power of fine-mapping for prospective studies. In this study, we showed advantages of conducting fine-mapping using a genome-wide Bayesian mixture model (GBMM), SBayesRC, which models all SNPs simultaneously with functional genomic annotations, requiring summary statistics only. We compared our GBMM to the existing fine-mapping methods, including FINEMAP, SuSiE, FINEMAP-Inf, SuSiE-Inf, and PolyFun+SuSiE. In the simulations across various genetic architectures, our GBMM had superior calibration in the ior probability of causality and increased the mapping precision quantified by the distance between the causal variants and identified SNPs. In addition, we showed through UK Biobank traits analyses that our GBMM improved replication rate of fine-mapping discoveries in an independent sample and enhanced prediction accuracy using identified variants within and between ancestries. In addition, leveraging the genetic architecture estimated from our GBMM, we developed a method to predict the power of a prospective fine-mapping study, thereby estimating the sample size required to identify a desired proportion of causal variants or to identify those that can explain a desired proportion of genetic variance. Applying our GBMM to 51 traits identifies 2,597 variants and 22,253 credible sets (CSs), estimated to capture 0.7% of all causal variants and 15.6% of the SNP-based heritability per trait. Take human height as an example. Using the GWAS summary statistics at 10 million SNPs from the UK Biobank (UKB), we identified 235 SNPs and 956 credible sets (mean credible size of 8.8), collectively explaining 27% of the genetic variance. These estimates are consistent with our analytical prediction given the UKB sample size (n=350k). Based on the genetic architecture estimated from the UKB data, we

predict that, when the sample size increases to 5 million, the number of fine-mapping discoveries would be ~10,000 considering significant PIPs (>0.9) only or ~30,000 considering both significant PIPs and CSs, explaining up to 95% of the genetic variance. This prediction is roughly consistent with the finding of a recent GWAS with 5 million individuals (Yengo et al Nature 2022). In conclusion, our study provides a robust and versatile genome-wide fine-mapping framework for identifying causal variants, highlighting the advantages of this approach over existing region-specific fine-mapping methods. With its capacity to enhance and inform mapping power, we believe genome-wide fine-mapping using a state-of-the-art GBMM will become the method of choice.

EMERGING MARKER ASSISTED SELECTION AND GENOMIC SELECTION

Zhang, Zhiwu

Washington State University

Marker Assisted Selection (MAS) had been a pivotal tool in molecular breeding, harnessing genetic markers linked to specific traits. However, the landscape had evolved with the emergence of Genomic Selection (GS), a revolutionary approach that utilizes genome-wide markers to predict the genetic merit of individuals. GS encompasses both the strategy of summing all marker effects into breeding values, as introduced in 2001 by Meuwissen et al. using Ridge Regression and Bayesian methods, and directly predicting individuals' breeding values using kinship derived from all markers, as introduced by Rex Bernardo in 1994, now known as genomic Best Linear Unbiased Prediction (gBLUP). The resurgence of MAS in the era of genomic selection is propelled by advancements in Genome-wide Association Studies (GWAS), facilitated by improved marker density, sample size, and statistical models. GWAS, a primary method for dissecting genetic loci underlying phenotypes, emphasizes the explanatory aspect of genomic research, complementing the predictive power of GS. While improved prediction doesn't always guarantee better explanation, superior explanation contributes to enhanced prediction. A study by Bernardo in 2014 demonstrated that incorporating known major genes can boost the precision of GS. However, the effectiveness of this incorporation depends on the discovery of causal genes through GWAS, with two-thirds of 200 simulated traits not showing significant benefits in incorporating GWAS using the Q+K model into GS using Ridge regression, which is equivalent to gBLUP. This study revisits the efficacy of newer models introduced post the Q+K model and illustrates their impact on enhancing the accuracy of genomic selection. The Q+K GWAS model, introduced in 2005 by Yu et al., aimed to fully control inflated P values, building upon the Q model by Richard in 2001 using General Linear Model (GLM). Subsequent developments focused on mitigating both false positives and false negatives to enhance statistical power. Strategies to minimize confounding between kinship and testing markers in the Q+K mixed linear model (MLM) include replacing individuals with corresponding groups in the compressed MLM. Additional enhancements, such as the enriched compressed MLM, SUPER, multiple loci mixed model (MLMM), FarmCPU, and BLINK, contribute to refining the capabilities to enhance genomic selection. The comparison of the above

seven GWAS models revealed notable disparities in their efficacy. Early models like GLM and MLM could only detect a limited number of causal SNPs, around 5 or 6, among 20 simulated causal SNPs underlying a trait with 75% heritability. In contrast, advanced methods like BLINK demonstrated a significant improvement, doubling the findings. When the number of causal SNPs increased by 150% to 50, GS, represented by gBLUP, achieved an accuracy of 55%. The accuracy of gBLUP was further enhanced by integrating covariates from associated SNPs identified in GWAS using the seven models. The magnitude of improvement aligns with the respective capabilities of GWAS models in identifying causal SNPs. The most substantial accuracy enhancement occurred with the incorporation of GWAS results using BLINK, reaching 68%, compared to Q+K at 60%. This comprehensive investigation illuminates the advancements post-Q+K and their potential to enhance the accuracy of genomic selection.

DECIPHERING THE GENETIC MECHANISMS OF COMPLEX TRAITS IN CHICKEN AIL POPULATIONS USING MULTI-OMICS DATA

Zhu, Xiaoning[1]; Li, Chong[2]; Luo, Chenglong[3]; Zhou, Huaijun[4]; Qu, Hao[2]; Fang, Lingzhao[3]; Hu, Xiaoxiang[4]; Wang, Yuzhe[1]

[1]State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing, China; [2]State Key Laboratory of Livestock and Poultry Breeding, Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou, China;; [3]Department of Animal Science, University of California, Davis, CA, USA;; [4]Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark

Integrative analysis of multi-omics data can elucidate valuable insights into genetic mechanisms for complex traits. Here we reported an F16 advanced intercross line (AIL) for QTL fine-mapping, characterized by sufficiently randomized the founder genomes, rapid linkage disequilibrium decay and abundant haplotype diversity. Utilizing 7.9 million SNPs and 75 phenotypes from five categories of about 1200 individuals, a total of 682 QTL were identified in 43 phenotypes. The 60.76% of these 682 QTL were loci associated with more than one trait and 25.78% were multi-domain loci. Gene-level mapping resolution was achieved at about 154 loci, of which 65 (involved 53 unique genes) were associated with growth and development phenotypes, had been identified in gene-edited mice. Next, we integrated the molQTL of the ChickenGTEx (~5000 transcriptome samples from 28 tissues) and the epigenetic annotation of the functional annotation of animal genomes (FAANG). We identified a wide range of multiple cause-single effect genes and one cause-multiple effect genes, while narrowing the range of candidate genes. For example, the Body weight at 8 weeks of age (BW8) phenotype is jointly regulated by 28 genes, and these genes act on the BW8 phenotype through 22 tissues. At the same time, with the occurrence of recombination, we detected new significant signals affecting EW in GGA4 of the F16 generation. The candidate SNP is located in the intestinal-specific enhancer, and this SNP affects both the promoter activity of NCAPG and NCAPG gene expression in the intestine, we believe that this SNP affects NCAPG gene expression and ultimately

affects the EW phenotype. In addition, we used hundreds of chicken samples from the world to illustrate the origin and transmission of causative haplotype, likely by combining standing variants from the red jungle fowl during the 1000s of years of chicken domestication, before they were rapidly accumulated in the high-weight chicken breeds during intense artificial selection. These results integrated molQTL and FAANG annotation information to determine the functional genes and causative mutations of complex traits, which have good guiding significance for in-depth analysis of the genetic structure of complex traits.

5 Poster Presentations

Please note that the first person listed is the presenter at the conference.

HOW ACCURATE IS GENOMIC PREDICTION ACROSS WILD POPULATIONS?

Aase, Kenneth[1]; A. Burnett, Hamish[2]; Jensen, Henrik[3]; Muff, Stefanie[3]

[1]Department of Mathematical Sciences, Norwegian University of Science and Technology.; [2]The Gjærevoll Centre, Norwegian University of Science and Technology.; [3]Department of Biology, Norwegian University of Science and Technology

Genomic prediction (GP) is a toolbox of methods that use high-density marker data to estimate and predict individual breeding values and phenotypes. Such methods have revolutionized plant and animal breeding and are also making headway in medical genetics. But despite its promise, GP has not yet been widely applied in evolutionary ecology or conservation contexts, mainly due to a lack suitable data sets. Recently, due to the increasing availability of genomic data for wild animal populations, a handful of GP studies have been performed in the wild, showing that GP is a valuable tool for predicting evolutionary change across time in wild populations. So far, results are promising when applying genomic models within a given population. But one major area of application, especially in conservation efforts, would be across-population GP, i.e., training the model on data from one population and predicting breeding values in another population. With existing methods there are severe limitations on the accuracies that can be expected in across-population GP. While this problem has been recognized in plant and animal breeding, as well as for human genomics, we posit that it is not yet sufficiently appreciated by evolutionary ecologists. In this work, we present results from standard GP models applied within and across two separate house sparrow metapopulations. We use phenotypic data from over 5700 house sparrows which were also genotyped on a 70K SNP array. To our knowledge, this is the largest multi-population GP study that has been performed on wild animal data until now. We show that existing GP models make very accurate predictions within populations. However, our results confirm and illustrate the challenges with across-population GP using current methods, namely the generally lower and more unpredictable prediction accuracy compared to within-population GP. Hence, although we confirm the promise of GP to increase our understanding of within-population evolutionary processes, further work is needed, both in terms of gaining a deeper understanding of the underlying issues limiting across-population accuracy, and in further adapting our models to the peculiarities and spatial structure of most wild populations.

GENETIC AND EPIGENETIC VARIANTS UNDERPINNING WITHIN-SPECIES TRANSCRIPTIONAL POLYMORPHISM IN A MAJOR FUNGAL PATHOGEN

Abraham, Leen[1]; Oggenfuss, Ursula[2]; L Tran, Nhu[1]; de Meaux, Juliette[1]; Croll, Daniel[3]

[1]AG de Meaux lab, Institute for Plant Sciences, University of Cologne,; [2]Department of Microbiology and Immunology, University of Minnesota Medical School, Minneapolis, Minnesota, USA,; [3]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, CH-2000 Neuchâtel, Switzerland

In agricultural ecosystems, outbreaks of diseases are frequent and pose a significant threat to food security. A successful fungal pathogen undergoes a complex and well-timed sequence of regulatory changes to avoid detection by the host immune system, hence well-tuned gene regulation is essential for survival. However, how regulatory adaptation enables pathogens to overcome host resistance and cause damage is poorly understood. Here, we used Zymoseptoria tritici, one of the most important pathogens of wheat, to generate a genome-wide map of genetic and epigenetic regulatory polymorphism governing gene expression. For this, we performed expression quantitative trait loci (eQTL) mapping on 146 con-specific strains. We identified cis-eQTLs for 65.3% of all genes and the majority of all eQTL were within 2kb of the transcription start site. Core genes were more likely to segregate eQTLs compared to accessory genes. We also found that insertion-deletion polymorphisms are more likely to act as a cis-eQTL and had a higher effect size than SNPs. Next, we contrasted the amount of cis-eQTL mapped across categories of pathogenicity-related genes. Effector genes were less likely to present cis-eQTLs compared to other genes including genes encoding CAZymes. This suggests that regulatory variation in effector genes is governed rather by epigenetic factors than by genetic polymorphism. This is consistent with pathogenicity genes tending to overlap regions of heterochromatin compared to other gene categories. To better understand epigenetic variation in the genome, we analyzed the transcriptional activity of individual copies of transposable elements (TEs) across isolates. We found 23 TE insertion loci with regulatory variation explained by cis-eQTLs. Furthermore, TE insertion polymorphism was associated with variation in pathogenicity traits among isolates. Our study establishes the first genome-wide map of genetic and epigenetic variation underpinning transcriptional plasticity and trait variation in a fungal pathogen. The extensive regulatory polymorphism is likely to fuel rapid adaptation to resistant hosts and environmental changes.

PREDICTING ADAPTIVE POTENTIAL FROM GENOMIC DATA AND ITS IMPLICATIONS FOR CONSERVATION

Abson, Katie[1]; Zijmers, Lillith[2]; Mittell, Lizy[1]; Eyre-Walker, Adam[2]; Hadfield, Jarrod[2]

[1]Institute of Ecology and Evolution, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL.; [2]School of Life Sciences, University of Sussex, Brighton, BN1 9QG.

Adaptive potential, the capacity for a population to adaptively respond to a shift in selective pressure, is necessary to avoid extinction in a rapidly changing world.

It is therefore critical that the adaptive potential of populations can be easily and accurately measured in order to inform conservation strategy. A widely used approach has been to use molecular genetic diversity as a cheap and unintrusive proxy for adaptive potential. However, previous empirical research has cast doubt on the efficacy of this approach, finding that the relationship between microsatellite diversity and additive genetic variation – a theoretically well-motivated proxy for adaptive potential – is very weak at best. More recently, the significant reduction in sequencing costs has facilitated a shift from the use of microsatellite diversity to nucleotide diversity and variation in functional regions. We plan to systematically re-assess the relationship between quantitative and molecular genetic variation using >125 species for which published estimates of heritability and evolvability exist. In doing so, we aim to comprehensively evaluate the utility of current genetic metrics to determine whether they are valuable predictors of evolutionary potential in a conservation context. Preliminary results will be presented.

UNLOCKING BARLEY ROOT GENETICS USING MACHINE LEARNING AND DRONE-MEASURED VEGETATION INDEXES

Alahmad, Samir[1]; Smith, Daniel[2]; Katsikis, Christina[3]; V. Meer, Sarah[4]; Meijer, Lotus[5]; Chenu, Karine[1]; Chapman, Scott[2]; B.Potgieter, Andries[3]; Wasson, Anton[4]; Baraibar, Silvina[6]; Godoy, Jayfred[1]; moody, David[2]; T. Hickey, Lee[3]; Robinson, Hannah[4]

[1]Centre for Crop Science, Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland (UQ), Brisbane, QLD, Australia; [2]Centre for Crop Science, QAAFI, UQ, Toowoomba, QLD, Australia; [3]Centre for Crop Science, QAAFI, UQ, Gatton, QLD, Australia; [4]Commonwealth Scientific and Industrial Research Organisation, Brisbane, QLD, Australia; [5]InterGrain Pty Ltd, Perth, WA 616, Australia # Presenting author; [6]InterGrain Pty Ltd, Perth, WA 6163, Australia # Presenting author

Root systems of future barley cultivars could be optimised to enhance soil resource capture and improve productivity and sustainability of production. However, for several reasons, most barley breeding programs are reluctant to apply selection for root traits. Direct selection is laborious and there is a lack of reliable and robust markers as well as a high degree of unexplored plasticity under complex genetic and environmental control. Furthermore, despite considerable phenotypic diversity for root traits identified using phenotyping under controlled conditions, the growth and distribution under field conditions are yet to be explored. In this study, we evaluated 395 Australian barley breeding lines for the first time in a field experiment at Gatton, Queensland, Australia. Integrated root and shoot phenotyping were performed to 1) characterise root distribution using the 'core break' method and shoot biomass for a subset of the 20 most diverse lines 2) explore the potential of using unmanned aerial vehicle (UAV) as a phenotyping tool for rapid and extensive capture of canopy traits 3) gain new insights into the relationships between above- and below-ground development. Data captured from extensive root

coring and above-ground biomass of the 20 selected lines revealed significant variability in root distributions and canopy development. Machine learning approaches were employed, utilising canopy traits and vegetation indexes to predict above-ground biomass and root distribution for the 20 genotypes. The trained model was then applied to predict root distribution for the 395 Australian barley breeding lines. A spline curve was fitted in the predicted above-ground biomass and root counts over depth. The overall area under the root count curve was calculated and used as a proxy for root size. The area under the root count curve for different depths was also calculated to investigate variation in root distribution at different depths. To validate the utility of this machine learning prediction model we used haplotype local GEBV approach to determine genetic drivers for root distribution and biomass development. Novel chromosomal regions for both above- and below-ground were identified, and the similarities between genetic drivers for above and below-ground were investigated. These findings underscore the potential for utilising UAV-derived canopy traits as 'proxy traits' to facilitate indirect selection for root traits and canopy development. This approach has the potential to accelerate the development of barley varieties with improved canopy and root systems that are better adapted to future climates.

EXPLORING GENETIC VARIATION FOR ROOT ARCHITECTURE IN GLOBAL AND AUSTRALIAN BARLEY

Aldiss, Zac[1]; Lam, Yasmine[2]; Massel, Karen[3]; Crisp, Peter[1]; Godwin, Ian[2]; Borrell, Andrew[3]; Robinson, Hannah[1]; Hickey, Lee[2]

[1]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St. Lucia, QLD, Australia; [2]School of Agriculture and Food Sustainability, The University of Queensland, Brisbane, QLD, Australia; [3]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, QLD, Australia 4InterGrain Pty Ltd, Perth, WA, Australia

Climate variability coupled with increased frequency and duration of drought events is a core challenge for the future of Australian barley production. Drought adaptation is complex, made up of several component traits, some interacting, and all influencing crop water-use across the growth cycle. Research in other cereal crops suggests that optimising root system architecture can lead to improved water-use efficiency, particularly under marginal production scenarios. However, prior to developing an effective breeding strategy for the selection of root architectural traits for improved adaptation through barley breeding, we first need to better understand the genetic architecture underlying its control. To date, research in Australian barley has focused on bi-parental populations or small breeding populations with constrained diversity that is somewhat unrepresentative of the broader Australian diversity. In this study, we used a multi-experiment approach to explore genetic variation for seminal root angle in a panel of 799 barley lines consisting of wild, landrace, elite Australian breeding lines and commercial cultivars. Seminal root angle was selected as the trait of interest because it is widely regarded as a proxy for mature root system architecture in related cereal crops. Three experiments were conducted to

reliably assess seminal root angle performance of the panel and a linear mixed model was used to model the variance-covariance structure across experiments. All individuals were genotyped with 12,562 markers from the Wheat Barley 40K XT SNP chip. Using a haplotype-based mapping approach, key chromosomal regions associated with the trait were identified. Through the exploration of the scaled haplotype variance, five haploblocks were detected on chromosomes 3H, 4H, 5H, and 7H. The haploblocks were compared to previously mapped QTL and inferences were made between diverse and elite breeding lines. Here we present the first comprehensive foundational study exploring the haplotype architecture across globally diverse and Australian germplasm and provide insight into future breeding strategies to improve adaptation of Australian barley.

MULTI-BREED GENOME-WIDE ASSOCIATION FOR BIRTH WEIGHT IN BEEF CATTLE

Aliloo, H.[1]; Walmsley, B.J.[2]; Donoghue, K.A.[3]; Clark, S.A.[4]

[1]School of Environmental and Rural Science, University of New England, Armidale, NSW 235; [2]Animal Genetics Breeding Unit, University of New England, Armidale, NSW, 2351; [3]3NSW Department of Primary Industries, Livestock Industries Centre, Armidale, NSW, 2351; [4]4NSW Department of Primary Industries, Agricultural Research Centre, Trangie, NSW, 2823

Birth weight (BTW) is an economically important trait in beef cattle as it directly impacts the survival and future growth of calves. Accurate genomic evaluations for BTW that work across breeds may benefit from the inclusion of quantitative trait loci (QTL) or variants in high linkage disequilibrium with them that are shared across breeds. A multi-breed genome-wide association study (GWAS) in this context, is a powerful method to identify such causative mutations. The aim of this study was to perform a multi-breed GWAS for BTW using records generated from the Australian Southern Multi-Breed project which has enabled the head to head comparison of animals from five different breeds. A total of 4106 animals from the five most common temperate beef breeds in Australia, including 1410 Angus, 490 Charolais, 976 Hereford, 605 Shorthorn and 625 Wagyu cattle were recorded for BTW across 3 years in 47 cohorts. Animals were genotyped using medium density panels and genotypes were then imputed to a common set of 95,529 SNPs. Pre-adjusted records of BTW were used to perform a single SNP association analysis using the mlma option in GCTA software by fitting the fixed breed and random animal additive genetic effect in addition to marker covariates. A genomic relationship matrix was constructed and used to capture the (co)variance structure of the random additive genetic term. The obtained P-values of the SNPs were adjusted using the false discovery rate method. Candidate regions were identified by first locating the significant SNPs and then searching within the 500-Kbp interval downstream and upstream (1 Mbp window) of the significant SNP for SNPs that passed the suggestive significance threshold. To visualize the distribution of P-values across the genome, Manhattan plots were created. The cattle QTL database was used to compare our identified candidate regions to literature. The candidate regions were further investigated for identification of genes residing in them. Our results will contribute to the further understanding of the genetic architecture of BTW

across beef cattle breeds and can enhance the effectiveness of breeding programs by increasing the reliability of genomic prediction.

DEVELOPMENT OF AN INTER-SPECIFIC MAGIC POPULATION COMBINING SOLANUM LYCOPERSICUM VAR. CERASIFORME AND S. PIMPINELLIFOLIUM GENOMES FOR DISSECTING QUANTITATIVE TRAITS

Antar†, Oussama[1]; Arrones1†, Andrea[2]; Pereira-Dias, Leandro[1]; Solana, Andrea[2]; Ferrante, Paola[1]; Giuliano, Giovanni[2]; Prohens, Jaime[1]; José Díez, María[2]; Gramazio, Pietro[2]; Vilanova, Santiago[1]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 4, 46022 Valencia, Spain; [2]Agenzia Nazionale Per Le Nuove Tecnologie, L'energia e Lo Sviluppo Economico Sostenibile (ENEA), Casaccia Research Centre, Rome, Italy

In this work, we present a new tomato MAGIC population (Multi-parent Advanced Generation InterCross) called ToMAGIC. This population was designed to unlock the genetic potential of tomato wild relatives for dissecting quantitative traits of interest for breeding. It was crafted by crossbreeding four varieties of each of Solanum lycopersicum var. cerasiforme (SLC) and S. pimpinellifolium (SP), which are respectively the closest weedy and the wild relative of the cultivated tomato S. lycopersicum var. lycopersicum (SLL). The selection of the eight ToMAGIC founders was done based on the geographical distribution of the two taxa, the different environmental conditions they belong to, and their genetic diversity. The final MAGIC population encompasses 354 lines, which were genotyped using a new 12K tomato single Primer Enrichment Technology (SPET) generating 6,488 reliable SNPs. The genotyping data revealed a high level of homozygosity (93.69%), an absence of genetic structure and an equitable representation of the founder genomes, ranging from 11.62% to 14.16%. To assess the utility of the ToMAGIC population in tomato genetics and breeding programs, we performed a pilot study by phenotyping several quantitative traits related to fruit size, plant pigmentation, leaf morphology and earliness traits. Genome-wide association studies (GWAS) analyses revealed highly significant associations for the studied traits, highlighting both known and novel candidate genes near or within the linkage disequilibrium blocks. Domesticated alleles for fruit size were recessive and found at a low frequency in wild/ancestral populations. Our results confirm that ToMAGIC is a valuable resource for the genetic dissection of quantitative traits in tomato breeding. Moreover, some MAGIC lines display a pyramiding of traits of interest that could have direct applicability for integration into breeding pipelines providing untapped quantitative variation for tomato breeding. Keywords: Tomato, Solanum lycopersicum var cerasiforme, S. pimpinellifolium, inter-specific multi-parent advanced generation inter-cross (MAGIC), genome-wide association studies (GWAS)

CORRELATION OF PI3K P85/P110 ALPHA AND GLUCOSE TRANSPORT PROTEINS IN GESTATIONAL DIABETIC PLACENTAS

S, Aydemir[1]; D, Harmanci[2]; H, Cengiz[3]; Nevin, Ersoy[4]; S, Sözdinler[5]; UE, Dogan[1]; G, Güner[2]; B, Baykara[6]

[1]Dokuz Eylül University Medicine Faculty, Health of Science, Department of Histology and Embryology, Turkey, Izmir; [2]Dokuz Eylül University, Medicine Faculty, Health of Science, Department of Molecular Medicine, Turkey, Izmir; [3]Tınaztepe University Medicine Faculty, Department of Histology and Embryology, Turkey, Izmir; [4]Dokuz Eylül University, Medicine Faculty, Surgery Medicine Department of Gynecology and Obstetrics, Department of Reproductive Endocrinology, Turkey, Izmir; [5]Izmir Economy University, Medicine Faculty, Health of Science, Departmen of Medical Biochemistry, Turkey, Izmir; [6]Tınaztepe University Medicine Faculty, Department of Histology and Embryology, Turkey, Izmir

Activating the PI3K/Akt signaling pathway in placental tissue with GDM increases the translocalization of GLUT proteins to the plasma membrane. We investigated the putative relationship between the GLUT and PI3K, p85/p110α mRNA levels in the different placental areas, which are thought to be related to diabetes, and the changes in 75 g OGTT blood-glucose levels. GLUT 1 acts as the primary in the control group, GLUT 4 and GLUT 12 role in a synchronized manner in response to insulin, p110 p85/p110α complex plays a role. GLUT 3 may functionally replace the GLUT 1 that carries out primary glucose transport in the GDM. All these suggest that gene expression levels tend to be preserved in the FP area of the placenta, but this balance is impaired in MC areas on the maternal surface of the placenta. we compared the blood glucose concentrations between the control and GDM groups, we found that there were significant differences in gene expression levels between 160 and 200 mg/dL in the 1st hour. In the control group, a decrease in GLUT 12, p110α gene expression levels were observed in the group with GDM, while an increase in GLUT 3 gene expression levels was observed. If the 2nd-hour blood concentration level was between 140 and 200 mg/dL, a decrease in GLUT 1 gene expression level was observed, especially in the GDM group compared to the control group, and an increase in GLUT 3 expression level was observed. Considering these data, when we compare blood glucose concentration levels with placental GLUT and PI3K p85/p110α gene expression levels, we observed differences in levels as more than 100 mg/dL at 0h, 160 mg/dL at 1st hour, and 140 mg/dL at 2nd hour. Therefore, evaluating the possible relationship between blood glucose concentration and especially GLUT gene expression levels can contribute to GDM diagnostic criteria levels.

A SINGLE-LOCUS QUANTITATIVE GENETIC MODEL TO INCLUDE DNA METHYLATION INFORMATION

Ayres, L.; L. Calus, M. P.; Bovenhuis, H.

Animal Breeding and Genomics, Wageningen University & Research, The Netherlands

After 1900, the rediscovery of Mendel's work sparked great interest among scientists, leading to the establishment of the study field of genetics. In 1918, a seminal paper by R.A. Fisher joined Mendelian genetics and biometry by introducing the infinitesimal model, laying the foundations for quantitative genetics. Notably, he presented the foundation for decomposing the genotypic value and the phenotypic variance—central concepts in quantitative genetics. Further developments that resulted from studies of this paper introduced the notions of genotypic value, breeding value, average excess, average effect, additive genetic variance, dominance deviation, dominance variance, and heritability. These basic concepts are used today in applications where many loci are included in genetic analyses. Here, we focus on the case of a single locus and extend the notion to include DNA methylation information to the gene as a variable which has a repressive effect on gene expression and, consequently, on gene action. In this manner, we introduce the concepts of expressed breeding value and whole breeding value and show the equivalence between the whole breeding value and the classic breeding value concept as the conditional expectation of a least-squares fit to the three possible genotypes. In the absence of methylation, our proposed model reduces to the traditional decompositions of the genotypic value and phenotypic variance. Graphically, the expressed breeding values appear as intersection lines between a regression surface and three parallel planes, each corresponding to a genotype. The regression surface is estimated from a multiple linear regression with an extra predictor variable and an interaction term.

GENOME-WIDE ANALYSIS OF MULTI-ANCESTRY SUMMARY STATISTICS USING VECTOR APPROXIMATE MESSAGE PASSING

Combining genome-wide association studies across genetically diverse cohorts of ancestries can potentially leverage the shared information and increase polygenic risk score portability. Current state-of-the-art cross-ancestry summary statistics methods rely on several regularized linear regressions or variational inference schemes. In contrast, an individual-level method called gVAMP was recently developed based on the approximate message passing framework, which is hypothesized to be Bayes optimal. gVAMP allows for the joint inference of genetic effects accounting for Bayesian prior while achieving similar performance to the top sampling-based methods in only a fraction of time. In this work, we first adapt gVAMP to the summary statistics setup in order to propose a novel method called summary gVAMP (sgVAMP). We demonstrate that compared to other popular summary statistics methods, sgVAMP achieves state-of-the-art out-of-sample prediction accuracy across several traits in the UK biobank using the 2.17 million SNP set. Secondly, we extend sgVAMP to a multi-cohort setting, which allows for the joint estimation of shared and population-specific signals across multiple ancestries. Finally, we use sgVAMP for the comprehensive joint analysis, combining UK biobank and Estonian biobank datasets using 500K and 200K individuals, respectively.

MODIFIED TWO-PART STRATEGY TO RAPIDLY IMPROVE TARGET TRAITS

Bancic, Jon[1]; Rhode, Antje[2]; Cavanagh, Colin[3]; Cocks, Nicole[4]; Gorjanc, Gregor[5]; Tolhurst, Daniel[5]

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana,; [2]Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain; [3]The Roslin Institute and Royal (Dick) School of Veterinary Studies; [4]University of Edinburgh, Easter Bush, United Kingdom 3; [5]BASF, Technologiepark-Zwijnaarde 101, 9052 Gent, Belgium

Plant breeding is vital for developing high-yielding crop varieties to feed the growing world population. A simulation study by Gaynor et al. 2017 demonstrated that a two-part breeding strategy that splits the program into population improvement and product development components can more than double genetic gains compared to conventional line breeding methods. These gains are driven by rapid recurrent genomic selection in the population improvement, with one or more crossing cycles per year, which rapidly increases the frequency of favourable alleles. Despite its clear advantage, the adoption of the two-part strategy has been slow due to practical concerns relating to (i) extensive program restructuring requirements and potential risks such as the loss of elite germplasm, (ii) challenges with management of multiple must-have traits in the population improvement, and (iii) the impact of genotype by environment (GxE) interaction. For this talk, we use simulation to propose a transition two-part strategy that maintains a conventional breeding pipeline with a partial allocation of resources to the population improvement component. This strategy is applied to a commercial hybrid canola breeding program and compared against the two-part strategy of Gaynor et al. 2017 and other conventional hybrid breeding methods. We simulate and monitor genetic progress of three must-have traits in canola representing grain yield, protein content and disease resistance with different genetic correlations and availability of training data for genomic prediction in the population improvement. We integrate our newly developed framework for simulating realistic GxE into breeding simulation to evaluate the performance of all strategies under low, moderate and high GxE levels. Our findings show that the transition strategy not only mitigates risks of losing elite germplasm but also maintains the gain advantages of the two-part strategy of Gaynor et al. (2017). Both two-part strategies demonstrate significant advantages over conventional methods across different GxE levels. Additionally, our findings highlight the importance of proactive management of must-have traits in the population improvement with rapid cycling to ensure they align with breeding objectives and the target growing area. This study leverages the principles of quantitative genetics to develop an alternative breeding strategy, investigate the effect of multiple trait selection and GxE to offer practical insights to breeders for its implementation.

GENOMIC SELECTION IN PERENNIAL FORAGE SPECIES: THE EXAMPLE OF ALFALFA

Barre, Philippe; Pégard, Marie; Julier, Bernadette

Most perennial forage species have an outcrossing reproduction mode and crossing are not easily controlled at a large scale. As a result, varieties are synthetics created by panmictic multiplication of selected plants in a polycross

that contains genetically distinct plants. Phenotypic selection is first conducted on spaced plants for traits such as vigour, disease resistance, or heading date, and the progenies of the best plants are tested on dense canopies (swards) forage yield and biochemical composition. In both steps, phenotyping takes about three years. When including one year for crossing of the selected plants, a cycle of phenotypic selection takes seven years and could be reduced to four years if selection is based solely on phenotypic data from swards. In order to increase the genetic progress, the challenge is to integrate genomic selection into this breeding program and assess the potential gains. We propose to explore this with alfalfa, focusing on forage yield, protein content, and ligno-cellulose content, using data from the European project EUCLEG. We suggest developing a genomic predictive equation using phenotypic data from swards sown with half-sib families or synthetics and genomic data from pooled individuals (allele frequencies). This equation would be applied to a large number of individually genotyped seedlings, enabling selection from the first year. Including one year for crossing, a cycle of genomic selection would take a maximum of two years. Concurrently, the equation would be updated annually with phenotypic data from swards of selected candidates evaluated in multi-site trials before registration. To evaluate the potential gain of genomic selection, 395 highly diverse alfalfa populations were phenotyped in two locations (France and Serbia) and genotyped using genotyping by sequencing (GBS) on pools, resulting in about 100,000 markers. The phenotypic standard deviations were 1.1 t/ha/year for forage yield, 41% of dry matter for protein content, and 40% of dry matter for ligno-cellulose content. Broad-sense heritability values were 0.26, 0.29, and 0.22 for forage yield, protein content, and ligno-cellulose content, respectively. Predictive ability values were 0.58, 0.66, and 0.50 for these traits, respectively. With a selection pressure of 5% in phenotypic and 5 ‰in genomic selection, the genetic gain per year would be more than six times greater with genomic selection than with phenotypic selection on swards for the three studied traits. This result is very promising for implementing genomic selection in alfalfa breeding. However, it must be approached cautiously, with several steps needing validation in practice. Specifically, the predictive ability must be assessed in breeding material, the application of the genomic predictive equation on individuals—although built on pools—must be validated, and the cost and feasibility of the required logistics for genomic selection must be evaluated. These issues are currently being investigated in the European project BELIS, which involves several breeding companies.

Breeding simulations with efficient haplotype tracking – putting ARGs into AlphaSimR

Becher, Hannes; Gorjanc, Gregor

The Roslin Institute, University of Edinburgh

In selective breeding and other population genomic simulations, it can be desirable to track the genomes of the entire population over the course of many generations. Stored naively, this information are highly redundant, because individuals inherit parental haplotypes from which, in theory, need not be stored again. The concise tree sequence, a type of ancestral recombination graph, can

store these haplotype relationships without redundancy by focusing on DNA events instead of DNA sequence itself, thereby the reducing disk/memory footprint of simulations. We are extending our popular selective breeding simulator AlphaSimR so it can work with tree sequences. We demonstrate a prototype that imports founder haplotypes in tree sequence format and that then grows the tree sequence as the simulation proceeds forwards in time. The resulting tree sequence may contain a full representation of all genomes that existed during the course of the simulation or can be simplified to store only certain, e.g. present-generation, individuals and ancestral haplotypes shared by these individuals. The tree sequence format is a natural choice for storing simulated genome information. The high-performance tskit library allows for efficient querying of tree sequence objects and to perform genealogy-based analyses like genealogically nearest neighbours, genealogy-based estimates of population genetic statistics, relatedness, and linear algebra operations. As the tree sequence format is increasingly adopted by other software, this will also increase the interoperability between AlphaSimR and the wider population genomics ecosystem.

GENOMIC PREDICTIONS AND GENOME-WIDE ASSOCIATION STUDIES ON DNA POOLS TO CHARACTERIZE TRADITIONAL MAIZE LANDRACES AND IDENTIFY GENOMIC REGIONS ASSOCIATED WITH AGRONOMIC TRAITS AND ENVIRONMENTAL ADAPTATION

Ben Sadoun, S.[1]; Galaretto, A.[2]; Roux, A.[1]; Gouesnard, B.[2]; Charcosset, A.[1]; Moreau, L.[2]; Madur, D.[1]; Nicolas, S.[2]

[1]Université Paris-Saclay, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Centre National de la Recherche Scientifique (CNRS), AgroParisTech, Génétique Quantitative et Evolution (GQE) Le Moulon, 91190, Gif sur; [2]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

Genetic resources are potential sources of diversity that can be mobilized to enrich plant breeding germplasm. They are therefore important to ensure long-term genetic gain and face new breeding objectives. Traditional maize landraces have a large genetic diversity and show various environmental adaptations. Using these landraces as a source of adaptative alleles for variety development could help to cope with climatic change and agroecological transition. However, traditional maize landraces remain relatively unexploited. To identify promising landraces and favorable alleles likely to enlarge the genetic diversity of modern varieties, we studied 262 traditional maize landraces from the French national collection. Each landrace was evaluated for different agro-morphological and seed composition traits, and was genotyped with a SNP array using pooled DNA samples. In addition, bioclimatic variables of the population collection sites were extracted from the WorldClim database. Genetic structure analysis showed that this collection is composed of 4 genetic groups. Genome-wide association studies (GWAS) identified several genomic regions associated with agronomic traits and environmental adaptation. Some of the identified genomic regions colocalize with known genes related to the studied traits. We also showed that genomic

predictions (GP) can be used to accurately predict several agronomic traits of landraces, which could be a way to predict the potential of large collections of unphenotyped landraces maintained in genebanks. To conclude, our results suggest that pool genotyping of traditional landraces is highly efficient to conduct GWAS and GP.

MEASURING LINKAGE DISEQUILIBRIUM AND IMPROVEMENT OF PRUNING AND CLUMPING IN STRUCTURED POPULATIONS.

Standard measures of linkage disequilibrium (LD) are affected by admixture and population structure, such that loci that are not in LD within each ancestral population appear linked when considered jointly. The influence of population structure on LD can cause problems for downstream analysis methods, in particular those that rely on LD pruning or clumping. To address this issue, we propose a measure of LD that accommodates population structure using the top inferred principal components. We estimate LD from the correlation of genotype residuals and prove that this LD measure remains unaffected by population structure when analyzing multiple populations jointly, even with admixed individuals. Based on this adjusted measure of LD, we can perform LD pruning to remove the correlation between markers for downstream analysis. Traditional LD pruning is more likely to remove markers with high differences in allele frequencies between populations, which biases measures for genetic differentiation and removes markers that are not in LD in the ancestral populations. Using data from moderately differentiated human populations and highly differentiated giraffe populations we show that traditional LD pruning biases Fst and PCA but that this can be alleviated with the adjusted LD measure. In addition, we show that the adjusted LD leads to a better PCA when pruning and that LD clumping retains more sites with the retained sites having stronger associations.

GENOMIC PREDICTION FOR GRAIN AND PANICLE ARCHITECTURE IMAGE-BASED PHENOTYPING

Berro, Inés; Lado, Bettina; Gutiérrez, Lucía

I. Berro, L. Gutierrez, Department of Agronomy, University of Wisconsin – Madison, I. Berro, B. Lado, L.Gutierrez, Statistics Department, Facultad de Agronomía, Universidad de la República.

The grain and inflorescence architecture are some of the key determinants of grain quality and yield in cereals. While the genetic makeup behind these traits is often complex, a large body of research has focused on understanding, mapping, and predicting these traits. However, the grain and panicle architecture of oats is less understood than that of other cereals such as wheat, barley, or maize. The goal of this research was to study the genetic architecture of panicle and grain traits in a diverse oat population. Specifically, we identified genomic regions associated with panicle, grain, and field traits through a large genome-wide association study (GWAS), and we compared models for genomic prediction, including single-trait, the use of de novo marker-trait association

(MTA), and the use of multi-trait models. Four large and highly replicated, multi-environment field evaluations followed by image-based phenotyping traits in oat grains and panicles were used. Mostly, there were no differences among the five Bayesian single-trait models in terms of their predictive ability, with the Bayesian Lasso performing worse for a few traits. Including the MTA improved the predictive ability for most traits with MTA that explained 5 % to 7 % of the variance. Furthermore, the use of auxiliary traits in multi-trait genomic prediction improved the prediction of correlated traits. Our results, in combination with the impressive developments and availability of high-throughput image-based phenotyping systems, could provide researchers with new tools to study and understand complex quantitative traits such as grain and panicle architecture and, in turn, find more efficient systems to improve crop performance.

GENETIC RELATIONSHIP BETWEEN NUMBER OF MORBIDITIES AND LIFE EXPECTANCY.
The relationships between genetic risk for a number of morbidities and lifespan have been reported. However, the genetics of the number of morbidities and their association with lifespan remains not thoroughly investigated. The aim of this study was to estimate the genetic heritability of the number of morbidities individuals have and to assess the predictive power of genetic risk for the number of disorders on lifespan. In light of this, a genome-wide association study (GWAS) was conducted on the number of morbidities, and genetic heritability was derived using UK BioBank individuals. To assess the predictive power of polygenic risk score (PRS) for the number of morbidities on the lifespan, the PRS was derived from UK BioBank individuals with records of age at death, independent of the individuals in the GWAS. The heritability of the number of morbidities was estimated to 0.096 (SE: 0.0034). The PRS exhibited a significant association with lifespan. Individuals with the highest 10% PRS were estimated to have a 0.9-year shorter lifespan than those with the lowest 10% PRS. The number of morbidities showed a significant global genetic correlation with parental lifespan (rg: -0.79, se: 0.041, P = 9.57E-79) as well as 44 significant local genetic correlations. This study elucidates the genetics of the number of morbidities and their association with lifespan, providing insight into genetic background of morbidities on lifespan.

GENOMIC DIVERSITY, ANCESTRY AND INBREEDING IN NEW ZEALAND FERAL KAIMANAWA HORSES
Bielke, Arne; Mohandesan, Elmira

University of Vienna
By revolutionizing mobility, agriculture and warfare, domesticated horses (Equus ferus caballus) have influenced the history of humankind significantly. As a result of intensive selective breeding, aimed at enhancing various performance traits, appearance, and temperament, has led to a concerning decline in genetic diversity within the genomes of domestic horses, declining by over 16% in the past two centuries. At a time where no truly wild ancestor of today's domestic

horse remains and genetic diversity in modern domestic breeds is rapidly declining, feral horse populations and local non-breed horses represent invaluable genomic resources for future horse breeding programs. Initially introduced in 1815 by European settlers, New Zealand's Kaimanawa mountain range now harbours the fourth-largest feral horse population in the world. To counteract their significant population decline (due to unregulated hunting, farming, and forestry), the New Zealand government passed a spatially restricted protective law and annually reduces the population to ~ 300 animals. Over the last 150 years, independently of artificial selection, different demographic forces have left their footprints in the genome of Kaimanawa Horses. Thus, this population offers a natural, ongoing laboratory for investigating the effects of founder events and different conservation strategies on genetic diversity and inbreeding. Since our knowledge about the genetics of this unique feral horse population is limited to uniparental markers (mtDNA, Y-chr) exclusively,, we generated the first comprehensive genomic dataset for feral Kaimanawa horses, containing more than 250 horse samples. Preliminary data indicates that approximately 53% of the individuals in the KH population have one or more second-degree or closer relatives (pi-hat ≥ 0.25) in the sampled population. Employing SNP genotyping (80K), we further investigated the relatedness, revealing that the feral Kaimanawa Horse population is highly inbred and characterized by significant genetic divergence compared to other domestic breeds. Additionally, we estimated the effective population size in Kaimanawa Horses and examined whether and to what extent past and recent demographic events (historical bottlenecks, founder effects vs. recent bottleneck) impacted the genome of the current feral Kaimanawa Horses. The wide range of heights, body patterns and colors in Kaimanawa Horses indicate a diverse ancestry in this population. Preliminary data suggests a major historical contribution of imported stallion horses with Thoroughbred, Arabian, and British (Welsh) ancestry into today’s Kaimanawa Horses’ gene pool. Using 80K SNP, data we were able to validate this. Admixture analyses not only validated these initial findings but also provided evidence of the involvement of other notable horse breeds in shaping this feral horse population. Altogether, this project not only provides a scientific backbone for future conservation management plans to preserve genomic resources represented by this feral horse population. It will also help to gain a better understanding of the complex interaction between different demographic forces. The generated results will be beneficial for feral horses in New Zealand, and, mutadis mutandis, to feral horses all over the world.

Exploring the effects of the training set properties on Phenomic selection’s performance: a case study on a large sorghum BCNAM population

Bienvenu, Clément[1]; Salas, Nicolas[2]; Vincent, Garin[3]; Thera, Korotimi[4]; Diallo, Chiaka[5]; MohamedLamine, Tekete[6]; Baptiste, Guitton[7]; Dagno, Karim[8]; Diallo, Abdoulaye[9]; Mamoutou, Kouressy[10]; Leiser, Willmar[11]; Fred, Rattunde[12]; Sissoko, Ibrahima[1]; Toure, Aboubacar[2]; Nebie, Baloua[3]; Samake, Moussa[4]; Kholova, Jana[5]; Frouin, Julien[6]; Vaksmann, Michel[7]; Weltzien, Eva[8]; Teme, Niaba[9]; Rami, Jean-François[10]; Segura, Vincent[11]; De Verdal, Hugues[12]; Pot, David[12]

[1]AGAP/GIV/BIOS, Cirad, Montpellier, France ;; [2]ICRISAT, Patancheru, India ;; [3]IER, Bamako, Mali ;; [4]ICRISAT, Bamako, Mali ;; [5]AGAP/DDSE/BIOS, Cirad, Montpellier, France ;; [6]Agronomy Department, University of Madison, Madison, Wisconsin, United States ;; [7]International Maize and Wheat Improvement Center, Dakar, Senegal ;; [8]Université des Sciences des Techniques et des Technologies de Bamako - Faculté de Bamako, Mali ;; [9]Department of Information Technologies, Faculty of Economics and Management, Cze, Prague, Czech Republic ;; [10]Cirad, Montpellier, France ;; [11]Agronomy Department, University of Wisconsin, Madison, Wisconsin, United States ;; [12]AGAP/DAAV, INRAE, Montpellier, France

The concept of phenomic selection has been formalized by Rincent et al. in 2018. It is based on the idea that spectral information (Near Infra-Red Spectroscopy (NIRS), Hyperspectral Imaging…) acquired from animal or plant tissues contains genetic information that can be used to predict the genetic values of candidates to selection. Together with this genetic information, spectra also capture information linked to the environment and genotype by environment interaction effects, that can also prove to be useful to optimize the prediction of individual's performances in different environmental contexts. Furthermore, because spectral information corresponds to intermediate phenotype (endophenotypes) located between the genome and phenotypes of interest, it can also capture interaction effects between genes that are typically difficult to obtain from DNA polymorphism information. This novel approach of phenomic prediction has proven to be relevant in several plant species, achieving higher genetic gains in a variety of contexts than with classical phenotypic or genomic selection approaches. In this study, we have made the first evaluation of this method on sorghum, an important cereal crop. More specifically we investigated seven different traits in a large multiparental BCNAM (backcross nested association mapping) population based on 2 recurrent parents and 22 donors encompassing a total of 2458 BC1F3:5 families. While phenomic selection boasts advantages in cost and throughput compared to genomic selection, the factors influencing its accuracy remain unclear and need to be studied. We have thus compared genomic and phenomic selection with different scenarios according to two main factors known as influencing genomic predictive abilities: training set size and genetic relatedness between training and testing sets. In addition, we have tested the ability of phenomic selection to predict genetic values while excluding the phenotypic data from the NIRS acquisition environment, in order to minimize any potential proxy effect due to correlations between the target and biochemical traits. Our results show that phenomic and genomic selection can reach similar predictive abilities for similar prediction scenarios of training and testing sets over a wide range of scenarios. Interestingly, phenomic selection seems less impacted by the genetic relatedness between training and testing populations than genomic selection. Our results also show that fewer individuals are needed in the training set for phenomic selection to reach its maximal predictive ability. Finally, our results suggest that NIRS is capable of capturing genetic information and that high prediction accuracies do not rely on a proxy-like correlation between NIRS and the measured trait.

WITHIN AND ACROSS POPULATION GENOMIC PREDICTIONS INCORPORATING FUNCTIONAL GENOMIC ANNOTATIONS

Bonifazi, R.[1]; Heidaritabar, M.[2]; D. Barlow, L.[3]; C. Bouwman, A.[4]; Plastow, G.[5]; Stothard, P.[1]; Chen, L.[2]; Basarab, J.[3]; Li, C.[4]; Karaman, E.[5]; Moreira, G.C.M.[3]; BovReg consortium, the[4]; Gredler-Grandl, B.[5]

[1]Wageningen University & Research Animal Breeding and Genomics, Droevendaalsesteeg 1, 6700 AH Wageningen, the Netherlands;; [2]Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, AB T6G 2HI Edmonton, Canada;; [3]Lacombe Research and Development Centre, Agriculture and Agri-Food Canada, AB T4L 1W1 Lacombe, Canada;; [4]Center for Quantitative Genetics and Genomics, Aarhus University, 8000, Aarhus C, Denmark;; [5]Unit of Animal Genomics, GIGA Institute, University of Liège, Liège, Belgium; [6] https://www.bovreg.eu/project/consortium/

The inclusion of causal variants underlying phenotypic trait variation can improve the accuracy of genomic predictions. However, identifying causal variants for complex traits is challenging. Thus, large consortia are established in livestock and humans to combine datasets across different populations and countries and improve the statistical power of causal variant identification. The BovReg project has developed a multi-dimensional functional genomic annotations (FGA) map of the bovine genome in a diverse catalogue of tissues and for different traits. Biology-driven genomic prediction models can account for the identified FGA from different omics levels. FGA can be incorporated in such models through pre-selection and weighting bi-allelic variants (i.e., single nucleotide polymorphisms, SNPs) based on their functional impact. These biology-driven genomic predictions have the potential to improve the accuracy of both single and multi-population genomic predictions for lowly related individuals, especially for complex and scarcely recorded traits associated with biological efficiency, such as feed efficiency. In this study, we aimed to implement and validate biology-driven genomic predictions that incorporate FGA identified in the BovReg project for cattle feed efficiency within and across populations. Imputed WGS and feed efficiency data were available for dry matter intake for ~3,600 Holstein-Friesian dairy cows from the Netherlands and ~5,500 composite crossbred beef cattle from Canada. We developed a pipeline to extract, from each population, the SNPs associated with FGA at different omics levels: genomics (meta-GWAS QTLs), transcriptomics (eQTLs), and chromatin accessibility (ATAC-seq). Different scenarios were compared: 50K: genomic predictions using a commercial 50K SNP panel (used as a benchmark to investigate the benefits of moving towards biology-driven genomic predictions); 50K+QTL: as 50K but modelling a separate additional layer of SNPs associated with QTLs; 50K+eQTL: same as 50K but modelling a separate additional layer of SNPs associated with eQTLs; Multi_GF: same as 50K but modelling a separate additional layer of SNPs associated with QTLs and/or eQTLs, and including ATAC-seq information. A forward-in-time validation was used to estimate genomic prediction accuracy and dispersion within each scenario. Genomic predictions were implemented

using both a SNPBLUP (i.e., uniform genetic variance across all SNPs) and a BayesR approach, except for Multi_GF, which used a BayesRCπ approach to account for overlapping FGA. Genomic predictions were implemented both within populations, using separate univariate models, and across populations, using a bivariate model accounting for the heterogeneity of variances between populations. For Canada, relative to the 50K scenario, prediction accuracy increased by 24% in the 50K+QTL scenario and by 27% in the Multi_GF scenario. For other scenarios, no benefits were observed relative to 50K. Similar results were obtained for dispersion. For the Netherlands, all scenarios had similar accuracy and dispersion as in the 50K scenario. Finally, we observed no differences between SNPBLUP or BayesR approaches. Multi-population results will be presented. Our results highlight the need to pinpoint causal variants for biology-driven genomic prediction models and show how FGA from (open-access) consortium databases can be used to efficiently identify and pre-select SNPs with the potential to increase the accuracy of genomic predictions.

LEVERAGING ENVIRONMENTAL COVARIABLES IN HISTORICAL WHEAT DISEASE TRIALS TO ENHANCE GENOMIC PREDICTION ACCURACY

Brault, Charlotte[1]; Conley, Emily J.[2]; Gill, Harsimardeep S.[1]; Fiedler, Jason D.[2]; Anderson, James A.[2]

[1]Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN, United States; [2]USDA-ARS Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND 58102

Plant breeding is a crucial lever to increase the yield and quality of crops. In the U.S. Midwest, Fusarium Head Blight (FHB) is a disease posing a major threat to wheat due to yield loss and mycotoxin accumulation. The severity of FHB outbreaks is influenced by genotype susceptibility, environmental conditions, as well as genotype-by-environment interactions (GEI). Deciphering the basis of genotype-by-environment interactions (GEI) is an old question in quantitative genetics. Weather and soil data can be incorporated into prediction models to help explain GEI and genotype response. Our objective is to predict the genotype adaptation (i.e., GEI) in new environments to provide better prediction accuracy for our breeding pipeline, as well as a deeper understanding of genotype performance. In this study, we leveraged nearly three decades of phenotypic trials to assess FHB susceptibility in the Midwest. Over 700 genotypes were evaluated across 4 to 5 locations annually in inoculated misted trials, resulting in nearly 150 unique environments. Phenotypic traits related to FHB susceptibility were scored, and a subset of 230 genotypes were genotyped using a 3k genotypic array. Daily environmental covariates (ECs) were collected for each environment during the growing season . GEI were widely studied with the Finlay-Wilkinson (FW) regression, which correlates genotype performance with an environmental index derived from mean phenotype values. More recently, ECs were used to build an environmental relationship matrix, analogous to the genomic relationship matrix. We compared three approaches: (i) the FW regression, where we used the mean phenotype and the environmental relationship matrix; (ii) a reaction norm with an environment index was defined

based on ECs through partial least square regression to weight their performance in relation to the target phenotype, associated with a window search to find the best timing of data to extract; and (iii) a mixed model integrating genotype, environment and GEI effects through kernel matrices. We compared these approaches by assessing the predictive ability within environments when predicting unknown genotypes in unknown environments. For the FW and reaction norm regression, we extracted intercept and slope in response to the environment. The narrow-sense heritability values were very high for intercept and slope (>0.9) for the FW regression, and lower for the second approach. We applied genomic prediction in cross-validation to study the predictability of these parameters and found that, overall, predictive ability was higher and less variable for intercept than for slope. The prediction of the phenotypic value based on predicted parameters remained challenging with highly heterogeneous predictive ability and was higher for the reaction norm than for the FW approach. The mixed model approach demonstrated the highest overall accuracy. By providing these methodologies and prediction models to breeders, we aim to facilitate the selection of genotypes tailored to specific regions or adaptive to changing climatic conditions, ultimately enhancing crop resilience and productivity.

IDENTIFICATION OF MODERATE EFFECT SIZE GENES IN AUTISM SPECTRUM DISORDER THROUGH A NOVEL GENE PAIRING APPROACH

Caballero, Madison[1]; Satterstrom, F Kyle[2]; Buxbaum, Joseph D.[3]; Mahjani, Behrang[4]

*[1]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 2Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 3Program in Medical and Population Genetics, Broad Institut; [2]5Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA 6Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 7The Mindich Child Health and Deve; [3]Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA.; [4]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. 12Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. **

Autism Spectrum Disorder (ASD) arises from complex genetic and environmental factors, with inherited genetic variation playing a substantial role. This study introduces a novel approach to uncover moderate effect size (MES) genes in ASD, which individually do not meet the ASD liability threshold but collectively contribute when paired with specific other MES genes. Analyzing 10,795 families from the SPARK dataset, we identified 97 MES genes forming 50 significant gene pairs, demonstrating a substantial association with ASD when considered in tandem, but not individually. Our method leverages familial inheritance patterns and statistical analyses, refined by comparisons against control cohorts, to elucidate these gene pairs' contribution to ASD liability.

Furthermore, expression profile analyses of these genes in brain tissues underscore their relevance to ASD pathology. This study underscores the complexity of ASD's genetic landscape, suggesting that gene combinations, beyond high impact single-gene mutations, significantly contribute to the disorder's etiology and heterogeneity. Our findings pave the way for new avenues in understanding ASD's genetic underpinnings and developing targeted therapeutic strategies.

INTEGRATING MULTI-OMICS AND INTERPRETABLE ARTIFICIAL INTELLIGENCE MODELS FOR FLAVOR-ASSISTED SELECTION IN BLUEBERRY

Casorzo, Gonzalo[1]; Ferrão, Felipe[2]; Benevenuto, Juliana[1]; Azevedo, Camila[2]; Munoz, Patricio[2]

[1]University of Florida, Department of Horticultural Science, Gainesville, Florida, USA.; [2]Federal University of Viçosa, Department of Statistics, Viçosa, Brazil

Blueberries are a commonly consumed fruit known for their health benefits. Modern breeding programs are focusing on genetically improving fruit quality traits of blueberries to increase consumption. Flavor, a complex trait integrating taste, mouthfeel, and aroma, significantly influences consumption interest. Traditionally, blueberry flavor assessment relied on sugar-acid ratios, overlooking the aroma profile due to the high cost and complexity of analyzing volatile organic compounds (VOCs). However, recent findings indicate that VOCs account for 42% of consumer preference in blueberry, making it an important target for breeding programs. This study aims to integrate the use of genomics, metabolomics, and fruit chemical information using artificial intelligence (AI) models to assist in the selection of flavor-related traits in a breeding program. We hypothesize that AI models can extend their utility beyond enhancing predictive abilities, also offering interpretability by elucidating how the different input features influence these predictions. To this end, we paired sensory analyses to multi-omics data from 1061 individuals of a Southern Highbush Blueberry breeding population. Genotypic information was sequenced using the Illumina CaptureSeq method, yielding approximately 60000 SNP markers after filtering. Metabolomic data was obtained using two-dimensional gas chromatography (GC×GC-TOFMS) approach, resulting in the measurement of 56 VOCs. Fruit chemical data, including soluble solids and acidity, was measured using a refractometer and an automatic titrator. Employing linear mixed models and Extreme Gradient Boosting (XGBoost), we developed multi-omic prediction models for flavor liking and aroma intensity sensory scores. We also estimated the features' importance and their effects using Shapley Additive Explanations (SHAP). Interestingly, we found that several VOCs had a high feature importance in predicting sensory traits. A positive impact on prediction accuracy was observed when integrating metabolomic and chemical data to genomic selection models. Remarkably, using individual VOCs for model training resulted in similar prediction accuracies to those achieved with the entire metabolomic profile. This implies that a cost-effective and high-throughput strategy for predicting organoleptic traits within breeding programs could de developed.

UNVEILING THE GENETIC LANDSCAPE OF SHEEP METHANE PRODUCTION: INSIGHTS FROM RNA-SEQUENCING AND NETWORK ANALYSIS.

Chacko Kaitholil, Steffimol Rose[1]; Mooney, Mark[1]; Rezwan, Faisal[1]; Aubry, Aurelie[2]; Cristobal-Carballo, Omar[2]; Razban2, Vahid[3]; Shirali, Masoud[3]

[1]Queens University, Belfast, UK1; [2]Agri-food & Biosciences Institute, NI, UK2; [3]Aberystwyth University, Wales, UK3

Background: As the third-largest emitter among ruminant species, sheep contribute significantly to total enteric methane emissions from livestock, and with a worldwide population exceeding 1.2 billion, the environmental impacts of sheep methane production necessitate urgent attention. Reduced methane production is a heritable trait and could be originated from differentially expressed genes (DEGs). RNA-Sequencing based studies on sheep with varying emissions could help selection and breeding of sheep with reduced methane production enhancing productivity and profitability within sheep industry. Objective: This study aimed to identify potential DEGs associated with methane production in sheep and unravel the key signalling pathways driving this process. Results: RNA-Sequencing analysis of twenty-four lambs were performed on the Illumina NovaSeq Platform and analyzed by differential gene expression (DGE) analysis and weighted gene co-expression network analysis (WGCNA). The study identified key DEGs associated with sheep methane production. Notable downregulated genes included SPRR1A (Small proline rich protein 1A/cornifin-A), SPRR4 (Small proline rich proteins 4), and CACHD1 (Cache domain containing 1), while upregulated genes comprised ASNSD1 (Asparagine synthetase domain-containing protein 1-like) and NME4 (Nucleoside diphosphate kinase 4). Gene ontology (GO) analysis resulted in significant enrichment in twenty-nine terms, indicating involvement in nucleotide metabolism and ribonucleotide biosynthetic processes. Consistent with GO results, KEGG pathway analysis revealed enrichment in various pathways including purine and pyrimidine metabolism, suggesting roles in essential metabolic processes that could affect methane production. WGCNA identified a specific module encompassing sixty-one genes strongly associated with sheep methane production. Through GO analysis, notable enrichments emerged in biological processes associated with regulation of DNA replication suggesting potential involvement of cellular mechanisms governing DNA synthesis in regulating methane production in sheep. Furthermore, KEGG analysis showcased enrichment of these genes in diverse pathways, such as transcriptional misregulation in cancer and viral carcinogenesis, associated with dysregulated gene expression which may reflect broader molecular alterations associated with sheep methane production. Additionally, enrichments were observed in pathways such as systemic lupus erythematosus which are linked to immune dysfunction and inflammation indicating interactions between methane production, host immunity and metabolic processes. Lastly, five genes, namely histone H3.1, histone H2A type 1-B, histone H4, histone H2B type 1-C/E/F/G/I, and histone H2A type 1, emerged as hub genes characterized by a module membership > 0.8 and gene trait significance > 0.5. These genes exhibited high

interconnectedness with other genes within the module, indicating their central role in regulating methane production-related processes. Conclusion and future research implications: This study lays a foundation for addressing environmental challenges posed by sheep methane emissions through exploring candidate genes like NME4 and histone protein variants identified through DGE and WGCNA could serve as potential biomarkers used within targeted breeding strategies aimed at reducing methane emissions from sheep.

UNVEILING THE GENETIC LANDSCAPE OF SHEEP METHANE PRODUCTION: INSIGHTS FROM RNA-SEQUENCING AND NETWORK ANALYSIS.

Chacko Kaitholil, Steffimol Rose; Mooney, Mark; Rezwan, Faisal; Aubry, Aurelie; Cristobal-Carballo, Omar; Razban, Vahid; Shirali, Masoud

Queens University, Belfast, UK1, Agri-food & Biosciences Institute, NI, UK2, Aberystwyth University, Wales, UK3

Background: As the third-largest emitter among ruminant species, sheep contribute significantly to total enteric methane emissions from livestock, and with a worldwide population exceeding 1.2 billion, the environmental impacts of sheep methane production necessitate urgent attention. Reduced methane production is a heritable trait and could be originated from differentially expressed genes (DEGs). RNA-Sequencing based studies on sheep with varying emissions could help selection and breeding of sheep with reduced methane production enhancing productivity and profitability within sheep industry. Objective: This study aimed to identify potential DEGs associated with methane production in sheep and unravel the key signalling pathways driving this process. Results: RNA-Sequencing analysis of twenty-four lambs were performed on the Illumina NovaSeq Platform and analyzed by differential gene expression (DGE) analysis and weighted gene co-expression network analysis (WGCNA). The study identified key DEGs associated with sheep methane production. Notable downregulated genes included SPRR1A (Small proline rich protein 1A/cornifin-A), SPRR4 (Small proline rich proteins 4), and CACHD1 (Cache domain containing 1), while upregulated genes comprised ASNSD1 (Asparagine synthetase domain-containing protein 1-like) and NME4 (Nucleoside diphosphate kinase 4). Gene ontology (GO) analysis resulted in significant enrichment in twenty-nine terms, indicating involvement in nucleotide metabolism and ribonucleotide biosynthetic processes. Consistent with GO results, KEGG pathway analysis revealed enrichment in various pathways including purine and pyrimidine metabolism, suggesting roles in essential metabolic processes that could affect methane production. WGCNA identified a specific module encompassing sixty-one genes strongly associated with sheep methane production. Through GO analysis, notable enrichments emerged in biological processes associated with regulation of DNA replication suggesting potential involvement of cellular mechanisms governing DNA synthesis in regulating methane production in sheep. Furthermore, KEGG analysis showcased enrichment of these genes in diverse pathways, such as transcriptional misregulation in cancer and viral carcinogenesis, associated with dysregulated gene expression which may reflect broader molecular alterations

associated with sheep methane production. Additionally, enrichments were observed in pathways such as systemic lupus erythematosus which are linked to immune dysfunction and inflammation indicating interactions between methane production, host immunity and metabolic processes. Lastly, five genes, namely histone H3.1, histone H2A type 1-B, histone H4, histone H2B type 1-C/E/F/G/I, and histone H2A type 1, emerged as hub genes characterized by a module membership > 0.8 and gene trait significance > 0.5. These genes exhibited high interconnectedness with other genes within the module, indicating their central role in regulating methane production-related processes. Conclusion and future research implications: This study lays a foundation for addressing environmental challenges posed by sheep methane emissions through exploring candidate genes like NME4 and histone protein variants identified through DGE and WGCNA could serve as potential biomarkers used within targeted breeding strategies aimed at reducing methane emissions from sheep.

AGE-RELATED CHANGES IN METHYLOME NETWORKS ARE ASSOCIATED WITH KEY BIOLOGICAL PROCESS DRIVING HUMAN AGING

Numerous studies have explored age-related alterations in the mean or variance of the methylome at individual cytosine-phosphate-guanine dinucleotide (CpG) sites across the genome, yielding valuable insights into aging-related mechanisms and disorders. However, a notable gap remains in our understanding in the changes of specific co-methylation networks with advancing age and the resulting biological implications. This study examined changes in DNA co-methylation patterns at 753k CpG sites in the whole-blood sample from 7,532 unrelated individuals in the Generation Scotland (age range 18 to 99 years) cohort. DNA methylation data were pre-corrected for mean effects on age, and imputed smoking status and cell compositions. The dataset was then stratified into eight age groups, each comprising approximately 950 individuals, with the youngest and oldest age groups being defined as individuals under 36 years and over 66 years, respectively. Using unsupervised weighted correlation network analysis (WGCNA), we identified co-methylated CpG networks within the youngest age group and assessed their preservation in older age groups. We compared modules showing extreme (top 5%) changes or stability, based on the distribution of the proportion of variance explained and correlation preservation statistics. The modules exhibiting the largest changes across age groups ("changing modules") demonstrated a substantial decline in average correlation between their members, indicating a progressive erosion of methylation networks with advancing age. Similarly, the proportion of variance explained by the eigennode for changing modules decreased, signifying a breakdown in module density over time. Conversely, at the other extreme, "stable modules" exhibited largely unaltered connectivity with age. Genomic regions associated with changing modules were notably enriched in apoptosis and developmental pathways, while stable modules exhibited enrichment in immune system-related pathways. In addition, both stable and changing modules were strongly associated with DNA binding and protein activity suggesting the importance of these regulatory processes in modulating gene expression changes and cellular stability during aging. Overall, these findings

offer new insights into the link between changes in DNA methylation networks and biological mechanisms contributing to human aging.

GENETIC MAPPING OF DEPOT-SPECIFIC QTLS FOR OBESITY IN MICE

Cheverud, James M; Naila, Nafia; Andrade Oliveria, Fernando Cipriano

Loyola University Chicago

Fat is held in discrete adipose organs and as ectopic fat in various organs, such as the liver and skeletal muscle. These different depots have different physiological roles, for example, visceral fat in humans has important effects on insulin resistance, coronary heart disease, and liver physiology, while subcutaneous fat either has no effects or protects against these factors. However, mapping QTLs for fat depot-specific effects has been lacking. We use data from the F16 and F34 generations of the LG,SM advanced intercross line of mice. Four discrete fat pads, reproductive, renal, mesenteric, and inguinal, were weighed at necropsy at 20 weeks of age. The sample analyzed includes 960 animals from 76 full-sib families from the F16 generation and 1133 animals from 137 families in the F34 generation. We scored 1536 SNPs in the F16 and an additional 1536 SNPs in the F34 generation. Marker genotypes not scored in the F16 were imputed using the F34 map positions so both generations are fit to a common map. We analyzed the data using a linear mixed model with family membership as a random effect. Two genetic models were considered, a Full model including all genotype by diet and sex interactions (a*d, d*d, a*s, d*s, a*s*d, and d*s*d) and a Reduced model estimating 'a' and 'd' without those interactions. Interactions were included whenever a log-likelihood test indicated that the Full model had a significantly better fit to the data than the Reduced model. When we chose the Full model, we ran the mapping analysis on each of the four sex-diet cohorts separately to determine which cohorts show significant effects. We mapped 58 general obesity QTLs in the population and then we mapped again, this time including the sum of the other fat pads as a covariate in the model. We detected 54 depot-specific QTLs, 12 of which were also mapped as general obesity QTLs. We mapped 21 loci for the reproductive fat pad, 9 for the renal fat pad, 14 for the mesenteric fat pad, and 14 for the inguinal fat pad. Four of the loci fit two traits instead of one making for 58 QTL effects at the 54 loci. Fifty-four percent of these QTLs were fit with the Full model while 46% showed no significant improvement when interactions were included so those QTLs were fit using the Reduced model. The additive effect('a') is the most common genetic factor occurring at 30% of the locus*trait (4*54) combinations while dominance values are only significant at 11% of the potential number of traits and QTLs. Additive interactions are common with 20 a*s, 20 a*d, and 19 a*d*s interactions. In contrast, dominance interactions occurred at fewer locations with 9 d*s, 17 d*d, and 13 d*s*d interactions. Interactions involving diet (n = 59) are more common than those involving sex (n = 42). Averages of significant standardized additive and dominance genotypic effects were at 0.15 and 0.20 standard deviation units, respectively. We will further use bioinformatic analysis to identify positional candidate genes in these locations.

MODEL OF GENOTYPIC EFFECTS ACCOUNTING FOR MULTIPLE ALLELIC QTL AND DIFFERENT PLOIDY LEVELS

Chu, Thinh Tuan[1]; Jensen, Just[2]

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark;; [2]Vietnam National University of Agriculture, Faculty of Animal Science, Trâu Quỳ, Gia Lâm, Hanoi, Vietnam;

Stochastic simulation software is commonly used to assist breeders in designing cost-effective breeding programs, and to validate statistical models used in genetic evaluations. A common assumption of these simulation software is the bi-allelic state of quantitative trait locus (QTL). This assumption might come from current, common genetic models that mostly assume substitution effects of one allele to its alternative. While the bi-allelic state of marker loci is due to the common choice of genotyping technology of single nucleotide polymorphism (SNP) chip, the assumption may not hold for the linked QTL. Furthermore, given the existence of four DNA nucleotides, it is inherently illogical to limit QTLs to a bi-allelic state. Although some multi-allelic models have been developed for genomic prediction in diploid species, there is a notable absence of studies exploring multi-allelic models incorporating additive, dominance, and epistatic genetic effects for simulation purposes. Particularly lacking are simulation software used for modeling polyploids. Therefore, we propose a simulation model capable of simulating genotypic effects for multiple allelic models across different ploidy levels. This model accommodates genotypic effects of additive, dominance, and epistasis. When assuming bi-allelic QTL, the generalized model becomes identical to the model assumption in common simulation programs, and in genetic textbooks. We tested the simulation model through a small example that studied the effects of multi-allelic versus bi-allelic assumptions on accuracy of prediction in a single-population breeding program. The example used diploid and tetraploid genome structures of potato that have genome size of 888 cM. Linkage disequilibrium (LD) between markers and QTL was generated through a long historical population spanning 1000 generations. While assuming 10k bi-allelic markers, different levels of multi-allelism were assumed for 2k QTLs. Interestingly, we found that varying levels of multi-allelic assumptions for QTLs did not significantly impact the accuracy of predicted breeding values based on bi-allelic markers in this specific example. This could be attributed to the high marker density, enabling multiple markers to link to alleles in a QTL. Consequently, the effects of all alleles at the QTL with multi-allelic states could be estimated using bi-allelic markers. For instance, three closely linked bi-allelic marker loci could collectively represent up to eight different alleles of the QTL. This could be the reason that regardless of multi-allelism in QTL, bi-allelic markers with reasonably high densities could effectively predict breeding values in many genomic selection programs. While we did not anticipate comparable accuracy between different levels of multi-allelic assumptions for QTL, these assumptions remain particularly relevant for genomic prediction studies involving multiple breeds and populations. In conclusion, we have developed a simulation model capable of simulating genotypic effects generalized for multiple

allelic models and different ploidy levels. With a reasonable density of bi-allelic markers, genomic models can effectively predict breeding values despite the presence of multi-allelic QTL.

LME4BREEDING: ENABLING GENETIC EVALUATION IN THE ERA OF GENOMIC DATA.

Covarrubias-Pazaran, Giovanny Eduardo, Principal Investigator/Group Leader

International Rice Research Institute, Pili Drive, Los Baños, Laguna 4031, Philippines

Mixed models are a cornerstone in quantitative genetics to study the genetics of complex traits. The classical quantitative genetic model assumes some effects to be a random sample of a population (e.g., individuals) correlated based on their identity by descent and state. In addition, other relationships arise in the genotype by environment interactions (i.e., covariance structures). Open-source mixed model routines are available but do not account for complex covariance structures and are able to fit big genomic models. The proposed lme4breeding library performs first the eigen decomposition of the relationship matrix and is coupled as a second step with the Cholesky factor of the diagonal relationship matrix (eigen values) to speed up the REML computation, especially in multi-trait scenarios. This approach expands the powerful lme4 capabilities to fit random regressions and sparse problems to solve many of the plant and animal breeding problems of genetic evaluation. As shown, thousands of genotyped individuals tested in many environments can be fitted in seconds due to the sparse parameterization of the problem and surpasses the capabilities of similar software. The package is available in R (https://CRAN.R-project.org/package=lme4breeding).

Key words: quantitative genetics, R package, covariance structure, generalized linear mixed model.

MODELLING THE INFLUENCE OF MITOCHONDRIAL INHERITANCE ON QUANTITATIVE TRAITS: PLEIOTROPIC EFFECTS OF HAPLOTYPES ON MILK PRODUCTION IN DAIRY COWS

Genetic improvement of quantitative traits in animal production has mainly focussed on the use of polygenic additive variation in the nuclear genome, while

little attention has been paid to the mitogenome. Our study investigates the influence of mitogenome variation on milk production traits in Holstein cattle from Croatia. We used strategically generated next-generation sequencing (NGS) data from 109 maternal pedigree-lineages resulting in 7115 milk production data from 3006 cows (only data from the first five lactations were analysed). Since little is known about the biology of the relationship between mitogenome variation and production traits, our quantitative genetic modelling was complex. Thus, we employed five different models (cytoplasmic or maternal lineage model, haplotype or complete mitogenome sequence model, amino acid model, evolutionary or BEAST phylogenetic model, and mitogenome SNP model) to estimate the proportion of phenotypic variance attributable to mitogenome inheritance. In addition, the polygenic autosomal and X chromosome additive genetic effects based on pedigree were modelled, together with the effects of herd-year-season interaction, permanent environment, location, and age at first calving. Our results show that the mitogenome makes a substantial contribution (4-7% phenotypic variance) to all three milk traits, except in the evolutionary model. It should be emphasised that the estimated haplotype effects showed high linear correlations between the milk production traits (rMILK-FAT=0.83; rMILK-PROTEIN=0.98, and rFAT-PROTEIN=0.85), suggesting pleiotropic behaviour of the non-recombinant mitochondrial haplotypes. In the calculation of m2SNP, the variance between mitogenome effects included all genic/SNP locus variances as well as both intragenic covariances (between SNP loci within defined mitogenome genes/regions) and intergenic covariances (between SNP loci between defined mitogenome genes/regions). In the calculation of m2SNP (SNP model) we were able to decompose genetic variance into variance components within and between mitogenome regions. This approach was novel and opened new perspectives for analyzing the effects of non-recombining mitogenome SNP polymorphism on quantitative traits. Our research not only demonstrates ideas and methods for modelling mitochondrial inheritance effects on quantitative traits, but also highlights the potential of mitogenome information to improve milk production especially as the acquisition of complete genome sequences becomes cost-effective.

TREE SEQUENCES EFFICIENTLY STORE AUTOTETRAPLOID HAPLOTYPE INFORMATION FROM A MULTI-PARENTAL BREEDING POPULATION OF POTATO

Da Silva Pereira, Guilherme[1]; Pires Nogueira, Nathállia[2]; Mendes, Thiago[3]; Kante, Moctar[4]; Lindqvist-Kreuze, Hannele[1]; De Siqueira Gesteira, Gabriel[2]; Mollinari, Marcelo[3]; Zeng, Zhao-Bang[4]; Becher, Hannes[3]; Gorjanc, Gregor[4]

[1]Department of Agronomy, Federal University of Viçosa, Brazil.; [2]The Roslin Institute, The University of Edinburgh, UK.; [3]Bioinformatics Research Center, North Carolina State University, USA.; [4]The International Potato Center, Peru.

Because each parent randomly passes on half of their chromosomes to any one offspring, pedigrees are an imperfect tool for indicating the flow of genetic information that can show substantial variation due to recombination and segregation. In autopolyploids, this variation is higher than the common diploid setting, as more chromosomal combinations and crossover products are

possible. While encoding haplotype information seems trivial, doing this in a principled way that enables efficient storage and downstream computations is challenging. The succinct tree sequence is a table-based data structure for efficiently encoding an ancestral recombination graph of observed haplotypes and their history via coalescence/branching, mutation, and recombination events. Herein, we use tree sequence encoding in autotetraploid potato (2n=4x=48) data from the International Potato Center. A multi-parental population (7 females × 3 males) composed of 10 full-sib families (per family = 56~145, median = 120, total = 1,173 individuals) was genotyped with ~2.5k single nucleotide polymorphisms (SNPs). Based on allele dosages and reference genome ordering, phasing was carried out for each full-sib family using a hidden Markov model framework as implemented in the mappoly2 R package. The resulting consensus map of 2,114 SNPs spanning 1,585 cM was used to compute probabilities of individual haplotypes, which were ultimately recorded in the tree sequence format. Analyses directly applied to the tree sequence have revealed the expected underlying population structure, indicating a successful application. This pilot study is opening a potential to track the ancestry of DNA fragments based on pedigree and genome-wide dosage data for quantitative trait loci mapping or genomic selection as part of the increasingly available genotypic data from every recurrent selection cycle and observational trial stages of plant breeding programs. Tree sequences offer a succinct way of storing and reconciling haplotype data across generations in autopolyploids for statistical genetic analyses.

ADAPTIVE SEED-SIZE VARIATION IN ARABIDOPSIS THALIANA

While many studies on both wild and domesticated species have established that local adaptation is ubiquitous in plants, the mechanisms remain elusive: typically, we do not know which traits are important nor which genes are involved. In this study, we focus on seed and seedling survival, fundamental life-history traits that are also of enormous importance in agriculture. We utilize the advantages of the model plant A. thaliana to study what is likely to be a universally important mechanism: resource allocation to individual seed. We observe that seed size is highly variable, affects seedling growth, has a strong genetic basis, and shows geographic correlations consistent with local adaptation. In particular, we find that beach populations in southern Sweden produce extremely large seeds, suggesting that this trait provides a competitive advantage during seedling establishment. To investigate this, we conduct dense sampling of A. thaliana to study population structure and estimate migration rates, with particular emphasis on polymorphisms affecting seed size. We also measure seed size in the global "1001 Genomes" collection of A. thaliana lines, use GWAS to gain insight into its genetic basis, and look for ecological correlates that may explain the global distribution of this phenotype. In addition, we use crosses and reciprocal-transplant field experiments to test whether seed size influences fitness in beach vs inland environments. Our results demonstrate that the observed differences in seed size have important consequences for seedling growth and survival, and are likely to be involved in local adaptation.

COMPSIM, A SHINY APP TO BUILD AND COMPARE SIMILARITY MATRICES

David, Ingrid; Guibert, Julien; Marie-Etancelin, Christel

GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326 Castanet Tolosan, France

The last 20 years have been marked by major biotechnological advances that have made it possible to acquire 'omics' and high-throughput phenotyping data on a large number of individuals. Various models/methods have been proposed in the literature to integrate these data into genetic mixed models in order to i) estimate variance components ii) improve predictions of phenotypic, genetic/genomic or transmissible potential of individuals. One of the key steps in integrating this information is the construction of a similarity matrix between individuals based on the new information available. There are different types of similarity matrices (based on distances or co-variances between animals) and many ways of constructing them. We propose Compsim (https://forgemia.inra.fr/ingrid.david/compsim), an R Shiny application to facilitate the comparison of similarity matrices. The application permits to compare 10 different methods to build similarity matrices based on distance (Bray-Curtis, Jaccard, Euclidian), kernels (linear, polynomial, gaussian and arc cosine), ordination (Multidimensional Scaling, Dentrented Correspondence Analysis) and Poisson log-normal model. The application is organized into 3 pages. The first page is dedicated to the construction of the similarity matrices. Input data (individuals in line, measures in column) in csv format is loaded from this page. The application allows data to be imputed (constant or GBM) and/or transformed (log, clr, ilr, alr) if needed. Once the method is chosen, a partial visualization of the similarity matrix generated is provided and it is possible to export the entire similarity matrix in a csv file. The second page is dedicated to the comparison between matrices obtained using the different methods chosen by the user (from 2 to 10). It consisted in 3 panels: i) correlation plots of the different matrices ii) plot of off-diagonal elements between matrices iii) table providing the average value of off-diagonal elements for each chosen matrix, correlation between off-diagonal elements of different matrices and ratio of their range. The last page provides explanation of the different methods used to build similarity matrix. For the time being, therefore, comparisons between matrices are essentially graphical. In the long term, our aim is to improve the application to provide ad'hoc comparison statistics according to the purpose for which the matrices are used.

EVALUATING GENOMIC AND BIOCONTROL STRATEGIES FOR MANAGING SEPTORIA TRITICI BLOTCH IN WHEAT

Septoria tritici Blotch (STB) poses a significant threat to wheat production worldwide, causing notable yield losses due to the evolving behavior of the fungal pathogen Zymoseptoria tritici. Traditional management strategies, such as synthetic fungicide application and breeding for qualitative resistance, are increasingly less effective. Therefore, new, environmentally friendly and durable

approaches are necessary to control STB and mitigate yield losses. Breeding for quantitative resistance via quantitative trait loci and developing biocontrol strategies are crucial to achieve this goal. In this study, we evaluated 100 isolates of Z. tritici and their interaction with 200 wheat accessions and two microorganisms used as biocontrol agents. Using whole genome sequencing, we performed a genome-wide association study and identified two markers significantly associated with reduced fungicide activity of one of the endophytes. These markers were located upstream of MYCGRDRAFT_10686 on chromosome 1 and in the second exon of MYCGRDRAFT_69029 on chromosome 2. The proteins encoded by these genes, related to RNA splicing, maturation, and chitin synthesis, could explain the endophyte's fungicidal activity. We also tested the prediction accuracy of genomic prediction models that included pathogen marker information. Predictions were strongly influenced by the inoculated isolates, and no significant improvement was observed compared to existing models using only wheat or pathogen genotypic information. Although further testing and model refinement are still necessary, this project is a step towards a more sustainable disease management against STB.

POPULATION GENOMICS UNRAVELS THE EVOLUTION OF MUTATION BURDEN IN BREAD WHEAT AND ITS RELATIVES

Bi, Aoyue[1]; Xu, Daxing[2]; Dong, Jiayu[3]; Kang, Lipeng[1]; Guo, Yafei[2]; Song, Xinyue[3]; Zhang, Jijin[1]; Zhang, Zhiliang[2]; Qiu, Xuebing[3]; Xu, Jun[1]; Xu, Song[2]; Jiang, Liping[3]; Li, Yiwen[2]; Yin, Changbin[3]; Wang, Jing[1]

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, China.; [2]Key Laboratory of Seed Innovation, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.; [3]University of Chinese Academy of Sciences, Beijing, China. 4CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.

Mutations constantly arise in the genomes of living organisms. Although they are the ultimate source of evolutionary adaptation, most are deleterious. As a paragon of agricultural success, bread wheat has evolved from a weedy grass in the Fertile Crescent to the most widely grown crop worldwide. Deciphering how deleterious mutations were eliminated or controlled in the wheat genome is pivotal to advance crop breeding. In this study, we constructed the second generation of whole-genome genetic variation map of wheat (VMap 2.0) by re-sequencing 1,062 diverse wheat accessions. Through investigating 39,404 identified nonsynonymous deleterious mutations, we found that bread wheat harbors 2.33% and 5.39% more deleterious alleles than its wild progenitors, wild emmer and Aegilops tauschii, respectively. Nevertheless, selection during domestication and improvement has successfully purged deleterious mutations within certain genomic regions. Our introgression analysis showed that the increase in mutation burden from recent introgressions offset the decrease from

ancient ones, resulting in a 3.4% increase in mutation burden compared with regions without introgression. Our evaluation of deleterious mutations in the wheat genome sheds light on the complex dynamics of evolutionary forces at play and their collective impact on the fitness of bread wheat, offering valuable insights for future breeding endeavors.

BACK TO THE WILD: DISENTANGLING SHARED SELECTION BETWEEN NATURAL AND EXPERIMENTALLY EVOLVED POPULATIONS

Experimental evolution is a powerful tool for studying evolutionary dynamics in a more controlled setting. It can help us understand how populations respond to specific stressors. For quantitative traits, even when the stressor is known, genomic responses to selection are more difficult to disentangle from other evolutionary forces in natural populations. Therefore, the use of replicated laboratory populations can help us understand these processes. Here, we use experimentally evolved Drosophila populations derived from a natural population in which the genetic redundancy of 99 selected haplotype blocks indicated a polygenic adaptive architecture. While 74 haplotype blocks had a low frequency (< 0.1) in the founder population, the remaining 25 had a rather high initial frequency. Since natural Drosophila populations typically have very low linkage disequilibrium, the high frequency of large haplotype blocks (average: 14 kb) was rather unexpected. Consistent with ongoing selection in the natural population, these high-frequency haplotype blocks show reduced nucleotide diversity and substantial linkage disequilibrium. We discuss different evolutionary scenarios to explain why the selection response in natural populations continues in the drastically different laboratory environment.

EXPLORING THE RELATIONSHIP BETWEEN HOST GENOME, RUMEN MICROBIOME, AND METHANE EMISSION PRODUCTION

Methane emission production is a key contributing factor to climate change. According to the UK government, agriculture was estimated to be the source of 48% of the UK's methane emissions in 2020, which is an increase of 1.9% from 2019. Dairy cows alone produced 20 million tonnes of CO2 equivalent in 2020 due to the activities of their rumen microflora. Understanding the biological factors that regulate methane emissions is essential for reducing the environmental impact of animal production. This project aims to use machine learning to investigate the relationship between the host genome of ruminants, the rumen microbiome, and ruminant methane emission production. The goal is to understand the biological mechanisms between these factors to control methane emission production. Literature Reviews Two systematic literature reviews have been conducted as part of this project. The first review focused on the role of rumen microbial features in modulating methane emission production in sheep and bovine hosts. Over 400 papers were retrieved, and 117 were selected for final review using the PRISMA guidelines. The review revealed that Archaea have a potential, but controversial, role in methane emission, despite their low abundance. Bacterial genera commonly found in the rumen, such as Bacteroidetes, can have a variable effect on methane emissions, with Prevotella spp. being associated with lower methane emissions. Protozoa were found to be

associated with higher methane emissions, despite their low abundance in the gut microbiota. Fungi, although less explored, were also found to be associated with higher methane emissions. Additionally, studies that considered heritability found moderate heritability of gut microbiota, ranging from 0.12 to 0.40. The second literature review focused on machine learning methods used to predict microbiota changes in both human and animal hosts. The review identified 24 papers, of which 11 were selected for the text review stage using PRISMA filtering. Forty-five percent of the papers focused on the effect of genetics on microbe abundance in the context of human disease. One animal study focused on predicting the onset of subclinical ketosis using host genetics and the microbiome. These literature reviews provide a solid foundation for further research in this project. Future Work The next phase of this project involves using machine learning techniques to investigate the relationships between methane emission production, host genome, and rumen microbiome. The Northern Ireland Farm Animal Biobank (N.I.FAB) data will be used, which includes over 100 genotyped dairy cattle with rumen meta-genome sequence information and direct methane production records. The goal is to create models based on animal genome and rumen meta-sequence data to predict methane production in dairy cattle. Conclusion In conclusion, this project aims to understand the biological mechanisms behind methane emission production in ruminants by investigating the relationship between the host genome, rumen microbiome, and methane emissions. The systematic literature reviews conducted so far have provided valuable insights into the role of microbial features and machine learning methods in this context. Future work will involve using machine learning techniques to further explore these relationships using data from the N.I.FAB,

GENETIC DETERMINANTS OF VEGETATIVE GROWTH TRAITS OF LUCERNE USED AS LIVING MULCH FOR CEREAL PRODUCTION

"Lucerne (Medicago sativa), a drought-tolerant forage legume, is increasingly used in agroecological systems as a service plant. Employed as a perennial crop in intercropping with annual cash crops such as cereals, they form a living mulch system. This method is supposed to show multiple benefits: it controls weeds effectively through the living mulch effect, enriches the soil with nitrogen, and supports agricultural practices that reduce tillage, thus saving energy and preserving biodiversity. Recent studies have identified a significant challenge: current lucerne varieties are overly competitive, limiting the productivity of intercropped crops such as wheat and consequently negatively impacting yields. This issue appears because these varieties, selected for forage production, are characterized by a tall, erect growth habit unsuitable for mixed cropping systems. The aim of this study is to determine the genetic determinism of traits that limit the intercropping of lucerne with wheat. To address this, our research focuses on the genetic and phenotypic diversity of 30 lucerne populations from diverse geographical origins and characteristics, including variations in subspecies (sativa, falcata, or xvaria), ploidy levels (diploid or tetraploid), autumn dormancy (rated from 1 to 10), and plant architecture ranging from wild prostrate to cultivated erect forms. Conducted in a complete design with four

replications from April 2021 to December 2023, each population is represented by 40 individuals. Additionally, a 245-genotypes offspring from the polycross of three populations comprising 'Krasnokutskaya' (a falcata variety), 'Mezzo' (a vigorous sativa variety), and 'LPIII' (a sativa breeding population selected for seed production) is being studied. These individuals were phenotyped, focusing primarily on vegetative plant growth: height (measured every 15 days from January to June), lodging, leaflet size, and plant diameter, measured during the 2022 and 2023 growing seasons. Genotyping-by-Sequencing (GBS) was conducted using plant leaf samples, with a double-digest GBS approach, utilizing the restriction enzymes PstI and MseI. In processing the genetic data, we applied stringent quality controls by retaining only SNP with less than 20% missing values and excluding plants with more than 60% missing data. This rigorous filtering resulted in a final genotyping matrix with less than 3% missing values, and 61K SNP covering all eight chromosomes of the lucerne genome. Linkage disequilibrium was very short, below 500 pb. A high heritability of the traits was observed. GWAS analyses was performed with the MLMM method and QTL were identified for the traits, considering the 30 populations, the F1 progeny or the whole design. These results provide a basis for the genetic improvement of lucerne used as a living mulch, to ensure that it effectively contributes to sustainable agricultural practices."

GENETIC AND ADAPTIVE ARCHITECTURE OF BODY SIZE VARIATION UNDER TRUNCATING SELECTION IN DROSOPHILA

Kofler, Robert; Masri, Siraj; Kofler, Robert; El Masri, Siraj

University of Veterinary Medicine Vienna

Most medically, evolutionary and economically significant traits are polygenic.Yet, despite advances in mapping these traits, many open questions remain about the genetic architecture and the adaptation dynamics of them. Our project aims to elucidate the genetic and adaptive architecture of body size in Drosophila simulans, a typical complex trait.We established a population of 100,000 flies from inbred lines outcrossed to each other; this population will be allowed to breed and recombine freely over ten generations. We will conduct a Genome-Wide Association Study on a sample of 1000 flies from this base population to identify loci underlying variation in body size. Additionally we will perform Evolve and Resequence (E&R) study with truncating selection for increased body size in female D.simulans. Previous simulations indicated an enhanced power to identify loci under selection with a gradual increase of selected individuals throughout the experiment. We will thus use such a regime to identify adaptive loci and shed light on the Adaptive architecture of the trait. This regime will be compared to a traditional constant selection regime. We will track changes in allele and haplotype frequencies. Hypothesizing that not all loci identified by GWAS will respond in the (E&R), we will compare results from both approaches. Finally, after 20 generations of selection, another GWAS will be performed to detect causative alleles that were initially rare. This research will advance our understanding of the genetic and adaptive architecture of quantitative traits.

GENOMIC LANDSCAPE AND GENETIC DETERMINISM OF RECOMBINATION IN THE DOMESTIC GOAT

Etourneau, Alice; Rupp, Rachel; Servin, Bertrand

GenPhySE, Université de Toulouse, INRAE, INPT, ENVT

Recombination is a fundamental process in the viability of eukaryotic species which perform sexual reproduction. One possible approach to characterizing the recombination process is to study allelic segregations in large families. Analysing allelic transmissions between parents and offspring allows to detect and localize crossovers on the genome. These recombination data can then be used to study the intensity and the distribution of recombination along the genome and their variation between individuals. With this approach, the genetic architecture of recombination phenotypes has been characterized and several QTL linked to the recombination process have been detected in many species. Having access to similar datasets in related populations and species should bring information on the evolution of recombination traits. In this respect, recombination can be considered a good model for evolutionary quantitative genetics. In livestock, the development of cheap genotyping tools and genomic selection programs have produced large, densely genotyped pedigreed datasets which can be leveraged for recombination studies. In ruminants, several populations of cattle and sheep have been characterized for recombination phenotypes. Here, we present new results in a phylogenetically close ruminant species, the goat. We used dense 50k genotyping data in large pedigrees (7,588 samples total) from two breeds - Alpine and Saanen - in order to characterize the recombination process and its variability in this new species. To characterize the recombination landscape, we modelled the number of observed crossovers in a genomic interval by using a new Bayesian model accounting for differences between sexes, breeds and informativity of markers within families. Consistent with results in sheep, we found no significant effect of the breed on the genetic map, large sex differences (heterochiasmy) and that the part of the genome exhibiting heterochiasmy is typically found at chromosome extremities with males recombining more than females. We used the same data to derive a phenotype of recombination intensity, the genome-wide recombination rate (GRR), combining the number of crossovers per meiosis and the specific informativity of the genotypes of a parent-offspring pair. We showed it is highly variable between sexes but also between breeds. In both breeds, significant heritable variation for GRR was found in male but not in females. After genotype imputation, sequence-based GWAS on male breeding values for GRR revealed different genetic architecture of the traits in the two breeds with low genetic correlations and few overlapping QTLs. The QTLs detected in our study match previously discovered QTLs in ruminants, but some are also novel loci. Our study on the intensity and landscape of recombination and on their genetic determinisms in a new ruminant species opens perspectives to understand the evolution of this fundamental complex trait.

DUAL-TRAIT GENOMIC ANALYSIS IN HIGHLY STRATIFIED POPULATIONS USING GENOME-WIDE ASSOCIATION SUMMARY STATISTICS

Feng, Xiao; Shen, Xia

Greater Bay Area Institute of Precision Medicine, State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, School of Life Sciences, Fudan University, China

Genome-wide association study (GWAS) is a powerful tool to identify genomic loci underlying complex traits. However, the application in natural populations comes with challenges, especially power loss due to population stratification. Here, we introduce a bivariate analysis approach to a GWAS dataset of Arabidopsis thaliana. We demonstrate the efficiency of dual-phenotype analysis to uncover hidden genetic loci masked by population structure via a series of simulations. In real data analysis, a common allele, strongly confounded with population structure, is discovered to be associated with late flowering and slow maturation of the plant. The discovered genetic effect on flowering time is further replicated in independent datasets. Using Mendelian randomization analysis based on summary statistics from our GWAS and expression QTL scans, we predicted and replicated a candidate gene AT1G11560 that potentially causes this association. Further analysis indicates that this locus is co-selected with flowering-time-related genes. The discovered pleiotropic genotype-phenotype map provides new insights into understanding the genetic correlation of complex traits.

THE ROLE OF SPARSE DESIGNS AND ADVANCED MODELING IN IMPROVING GENOMIC SELECTION FOR PLANT BREEDING

In plant breeding, traditional field trials have relied on balanced designs with high replication to ensure reliable breeding value predictions. The advent of genomics, however, introduced the genomic relationship matrix, which links different genotypes and reduces the need for extensive replication. This shift has made augmented, sparse designs more effective than classical ones, despite the statistical challenges posed by low replication and imbalance. We conducted a simulation study using datasets from maize, sunflower, and oat to explore the interaction between experimental design and modeling across various field sizes. We compared classical and sparse designs, created through randomization or optimization, and developed a faster implementation of the coefficient of determination optimization criterion to overcome computational limits. Additionally, we evaluated different modeling strategies, contrasting single-stage models with unweighted and fully efficient two-stage analyses, and tested spatial analyses incorporating blocking structures versus high-resolution 2D P-splines. Our results confirmed that sparse designs achieved significantly higher accuracies than classical designs, with accuracy gains ranging between 3% and 50% depending on the scenario. Improved statistical analyses were crucial for maximizing accuracy in sparse designs. Optimization usually resulted in an approximate 1% accuracy gain compared to randomized augmented designs, while the 2D P-splines increased accuracy by about 4%, but only when paired with single stage or fully efficient models. These findings highlight the

importance of advanced experimental designs and statistical methods in enhancing the accuracy of genomic selection in plant breeding, providing valuable insights for future research and practical applications.

GENOMIC PREDICTION OF INDIVIDUAL INBREEDING LEVELS FOR THE MANAGEMENT OF GENETIC DIVERSITY IN POPULATIONS WITH SMALL EFFECTIVE SIZE

Soledad Forneris, Natalia[1]; Bosse, Mirte[2]; Gautier, Mathieu[3]; Druet, Tom[3]

[1]1Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium;; [2]Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands; Amsterdam Institute for Life and Environment (A-LIFE), Section Ecology and Evolution, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands;; [3]CBGP, INRAE, CIRAD, IRD, L'institut Agro, Université de Montpellier, Montpellier, France

In populations of small effective size (Ne), such as those in conservation programs, companion animals or livestock species, inbreeding management is key. In that context, homozygosity-by-descent (HBD) segments are valuable as they allow efficient estimation of the inbreeding coefficient, provide locus-specific information and their length is informative about the "age" of inbreeding. Therefore, our objective was to evaluate tools for predicting HBD in future offspring based on parental genotypes, a problem equivalent to identifying segments identical-by-descent (IBD) among the four parental chromosomes. In total, we reviewed and evaluated 16 approaches using both simulated and real data with small Ne, including a sequenced dairy cattle pedigree and genotyped Mexican wolves, a population that faced extinction in the wild. Methods included model-based approaches, mostly hidden Markov models (HMM), that considered up to 15 IBD configurations among the four parental chromosomes (corresponding to the so-called identity states), as well as more computationally efficient rule-based approaches such as those developed to analyze entire biobanks. The accuracy of the methods for predicting genome-wide and locus-specific HBD levels in offspring based on parental genotypes was then evaluated. Comparisons were also done using low-density marker panels or genotyping-by-sequencing data and on small groups of individuals, features typically found in such populations. We found that two HMMs (based on ibd_haplo and ZooRoH) performed consistently well, and that two rule-based approaches (based on phasedibd and ROH) were also efficient for genome-wide predictions. The model-based approaches were particularly efficient when information was reduced (e.g. low marker density, locus-specific estimation). We identified a number of less efficient methods that should not be applied to similar populations. Methods using phased data proved to be more efficient, while some approaches relying on unphased genotype data proved to be sensitive to the allele frequencies used. In some settings, pedigree information was competitive in predicting recent inbreeding levels. Finally, we showed that our evaluation is also informative about the accuracy of the methods for estimating relatedness and identifying IBD segments between pairs of individuals.

MOLECULAR DISSECTION OF GENETIC GISK PARTITIONING STRATEGIES TO EXPLORE INFLAMMATORY BOWEL DISEASE (IBD) PATIENT HETEROGENEITY

Gaite-Reguero, Adrian[1]; Sanchez-Mayor, Milagros[2]; Mars, Zoeline[3]; Lao Grueso, Oscar[4]; Bujanda, Luis[1]; M. Marigorta, Urko[2]

[1]Integrative Genomics Lab, Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Derio, Spain.; [2]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain.; [3]Biodonostia, Gastrointestinal Disease Group, Universidad del País Vasco (UPV/EHU), 20014, San Sebastián, Spain.; [4]IKERBASQUE, Basque Foundation for Sciences, Bilbao, Spain. 5 Institute of Evolutionary Biology, CSIC–Universitat Pompeu Fabra, Barcelona, Spain.

Genome-wide association studies (GWAS) excel at discovering loci associated with the risk of complex disease. They have been particularly successful in inflammatory bowel disease (IBD), with up to 290 independent loci discovered over the last 20 years. However, IBD is characterized by ample heterogeneity in disease presentation and symptomatology across patients. Because large-scale GWAS primarily compare cases and controls, extant GWAS have not served to dissect the role of genetic factors in IBD heterogeneity. Current successful approaches to this question focus on phenotypic-based partitioning of genetic risk (e.g. discovery of 5 subtypes of type 2 diabetes). However, this approach is not feasible for diseases that cannot anchor on associated biomarkers or comorbidity with other diseases, such as IBD. Given the biological complexity underlying the disease and the high heterogeneity observed among IBD patients, we hypothesized that different genetic archetypes converge to develop the disease. We hypothesized that these underlying genetic profiles capture relevant biological functions at the gene pathway level that control the risk of developing IBD. We developed a three-step strategy to characterize genetic variability across IBD patients and obtain polygenic risk scores (PRS) to predict patient subtypes. First, we carried out a PheWAS strategy to classify genome-wide significant variants into non-overlapping clusters of IBD signals. This phenotypic-based approach rendered seven independent clusters associated with a number of traits, including two clusters associated with inflammatory conditions and behavioural traits, respectively. Although promising, the partitioned polygenic risk scores constructed from each cluster do not serve to stratify the circa 4,000 IBD patients in the UK Biobank. Second, we performed an alternative approach based on gearing functional genomics evidence to obtain molecularly-based partitioned PRSs. Specifically, we used four approaches to classify SNPs into gene and tissue of action, identifying clusters of genes involved in IBD that share similar functional profiles. This includes classifications based on i) similarity in enhancer landscapes, ii) expression changes in IBD across tissues according to TWAS, iii) patterns of co-expression in relevant tissues, and iv) publicly available gene ontologies and biological pathways. Overall, we obtained 18,470 different pathways through this strategy. Notably, 25% of these pathways show significant differences between IBD patients and healthy controls (i.e. AUC > 0.54). However, none of them achieves the classification power of

the PRS based on genome-wide effects (i.e. AUC=0.66). For PRS partitioning strategies, this result uncovers a trade-off between molecular resolution and overall classification power. Third, we followed a combination-based approach to accommodate the diverse molecular functions that underlies the polygenicity in IBD. We developed a genetic algorithm to identify molecularly-based partitioned PRSs that jointly assign the circa 4,000 IBD patients from the UK Biobank into independent clusters that maximize the genetic differences among them. Through this work, we have evaluated a diverse set of strategies to dissect the inter-patient heterogeneity observed in IBD. This approach sheds light on the complex genetic architecture of IBD and offers insights into potential pathways for future therapeutic interventions.

LANDSCAPE BREEDING

García-Gil, MR[1]; Holmgren, J[1]; Olofsson, K[2]; Niemi, J[2]; Nordström, A[3]; Hall, D[3]

[1]Faculty of Forest Sciences, SLU, Sweden, SLU; 2Skogforsk, Sweden; [2]Skogforsk, Sweden; [3]Faculty of Forest Sciences, SLU

Introduction In conventional forest tree breeding, tree selection follows a recurrent scheme with repeated rounds of crossing, testing and selection, the so-called breeding cycle. This is a long process sustained by costly logistical and organizational commitments and constrained by factors such as the experimental size of the test trials, the accuracy of the measurements, and the ability to model individual genetic effects across sufficient environments. The genomic revolution opened the possibility to develop molecular tools to infer accurately the realized proportion of the genome shared among individuals - the pedigree. In conifer trees, this is the bases of an alternative approach to crossing-based conventional breeding, the so-called Breeding without Breeding (BWB). Originally, BWB approach was proposed to circumvent artificial mating, but it remained constrained by the need for establishing structured open-pollinated (OP) trials at a limited number of environments. Recent developments in remote sensing technology offer high-throughput, accurate, and spatially explicit means for tree phenotyping that can operate at large landscape scales and thus account for climatic variables, water, and soil data, among other variables to model ecological or environmental zones. Furthermore, remote sensing phenotyping also allows assessing a myriad of tree properties otherwise difficult or impossible to measure without expensive sampling, such as tree quality and tree health, among others. Methods In Sweden, up to 70% of the commercial forests of Norway spruce have been regenerated with improved OP progenies from known seed orchards. This represents the perfect scenario to integrate molecular marker-based pedigree reconstruction with remote sensing tree phenotyping to build a novel digitized breeding strategy that can operate at a landscape scale on existing commercial forests without the need for expensive crossing and mating-designs. This novel method, we named as Landscape Breeding, will allow monitoring of the level of genetic diversity, improvement of existing seed orchards, and selection of outstanding trees directly from existing commercial forests. Such an approach aims to accelerate forest improvement for economic and adaptive traits simultaneously to meet the increasing demand

for sustainable forest biomass. Results and Conclusion We have developed a remote sensing method to combine genetic, genomic, and ground&airborne remote sensing data. The method has been utilized to scan five Norway spruce stands and a total of 6000 trees. Those stands include one clonal archive, one progeny trial and three commercial forests. We are currently conducting data processing and the first set of remote sensing data is being utilized to conduct genetic analyses. We have proven our remote sensing protocol efficient in incorporating remote sensing phenotyping to assess genetics at the landscape levels.

NO STRONG HETEROSIS EFFECTS FOR MILK PRODUCTION AND CALVING INTERVAL OF CROSSES OF LOCAL ETHIOPIAN AND INTERNATIONAL DAIRY BREEDS KEPT UNDER ON-FARM CONDITIONS

Tadel Gebre, Kahsa[1]; Meseret, Selam[2]; Mrode, Raphael[3]; Gebreyohannes, Gebregziabher[4]; Mészáros, Gábor[2]; Okeyo Mwai, Ally[3]; Sölkner, Johann[3]

[1]Mekelle University, Department of Animal, Rangeland and Wildlife Sciences (ARWS), Enda-Eyesus campus, P.O Box 23, Mekelle, Ethiopia,; [2]BOKU University, Institute of Livestock Sciences, Gregor-Mendel-Strasse 33, 1180, Vienna, Austria,; [3]International Livestock Research Institute, Addis Ababa, P.O. Box 5689 and Nairobi P.O. Box 30709; [4]Mekelle University, Department of Animal, Rangeland and Wildlife Sciences (ARWS), Enda-Eyesus campus, P.O Box 231, Mekelle, Ethiopia,

International dairy breeds are utilized extensively in developing countries, such as Ethiopia, for crossbreeding with local cattle populations with the objective of improving milk production. In Ethiopia, Holstein Friesian and Jersey are the most widely used international breeds. The objective of this study was to investigate the breed difference and heterosis components for two traits in crossbred dairy cows in Ethiopia. Most previous studies have primarily focused on the impact of breed differences, yet heterosis is also a crucial aspect to consider in cows kept under on-farm conditions in the tropics. Milk yield (yield trait) and calving interval (fitness trait) were identified as traits that are expected to be differentially affected by heterosis. A total of 4,759 phenotype records and genotype data of dairy cows with 38,344 SNP were obtained from the African Dairy Genetic Gains (ADGG) project in Ethiopia. Local ancestry information was calculated using Efficient Local Ancestry Inference (ELAI) software, employing reference populations comprising Local cattle (Fogera (37) and Boran (44)) and international breeds (Jersey (37) and Holstein Friesian (29)). The prediction of heterosis effects was performed by calculating the additive, dominance, and epistasis levels from ancestry informative markers. For milk yield, a mixed model was employed, considering parity, stage of lactation and calving season as fixed effects, and additive, dominance, and epistasis as covariates, along with animal genetic plus permanent environmental as well as herd-test-day as random effects. In the case of calving interval, a similar model was used, excluding stage of lactation and replacing herd-test-day by herd. Genetic, permanent environmental and herd-test-day/herd variances were obtained by single-trait analysis using the AIREMLF90 software, and the heterosis effects were estimated

using BLUPF90+. Heritability was 0.28 for milk yield and <0.02 for calving interval. Significant breed differences were observed for milk yield and calving interval, with values of 7.85 kg and -38.77 days, respectively, indicating that the milk yield of pure international dairy cows is more than double that of local Ethiopian cows and calving interval is more than one month shorter for international dairy cows. Somewhat surprisingly, no significant dominance and epistatic effects were identified for either of the traits, indicating that F1 or F2 cows will be intermediate in performance compared to ancestral pure breeds. These results are not supported by a recent meta-study that found strong heterosis effects for milk production and many functional traits in similar types of crosses under tropical conditions.

ASRGWAS: AN R PACKAGE TO PERFORM COMPLEX GENOME-WIDE ASSOCIATION STUDIES (GWAS)

Most software for performing Genome-Wide Association Studies (GWAS) is restricted in many aspects, including limited or no additional fixed and/or random effects, accepting only a homogeneous error variances, and using a single record per genotype (unreplicated data). This limits the full use of the phenotypic records available, yielding in many cases to suboptimal analyses. This simplification is difficult to justify given the current availability of complex and powerful linear mixed model routines. In response to the above issue, at VSNi, we have developed a free R library, ASRgwas, to provide a complete tool to implement GWAS. This library assists with preparing data and matrices, and verifies that they are adequate to perform GWAS. In addition, it has a set of complementary functions to be used for post-GWAS analyses to help with interpretation, use of the output information, and obtaining graphical outputs. The main tasks considered within ASRgwas are: 1) preparing and auditing phenotypic and genomic data, 2) fitting GWAS models and identifying significant markers, and 3) evaluating results and generating tables and graphical output. The intent of this tool is to facilitate the execution of GWAS in a straightforward and efficient manner, along with providing full reproducibility of these analyses. We have used state-of-the-art approaches and algorithms to obtain reliable and fast estimation of marker effects. In addition, we have made use of parallelization, whenever possible, and fast matrix operation routines (e.g. C++) available within the software R. ASRgwas is designed to allow for any number of fixed and/or random structures, and heterogenoeus error variances. It also accepts raw and replicated data. All of this making full use of the linear mixed model (LMM) methodology as available in ASReml-R. In addition, this package accepts missing values in the marker information avoiding the need to implement marker imputation. As part of ASRgwas we have extended the analytical options within GWAS beyond a Normal distribution, by allowing the evaluation of phenotypic responses that follow a Binomial distribution using Generalized Linear Mixed Models (GLMM). ASRgwas is a free to use R library which can be downloaded from this web page: https://asreml.kb.vsni.co.uk/download-asrgwas/

VALIDATION AND SELECTION ON RESILIENCE INDICATORS DERIVED FROM LONGITUDINAL PERFORMANCE MEASUREMENTS

Resilience is defined as the ability of an animal to be minimally affected or quickly recover from a disturbance. Thanks to an increasing abundance in automated on-farm monitoring systems, longitudinal performance measures of individual animals become more routinely available. Several studies have proposed that statistical measurements of the deviation of an animal from its target trajectory (i.e., performance in ideal condition) provide useful resilience indicators (RI) and may be suitable for genetic selection. The aim of this study was to assess the ability of these RI to discriminate between different response types, their dependence on the quality of available data and correlated response to selection. The RI considered in this study were skewness, autocorrelation, integral, mean of squares and log-variance of performance deviations. Performance trajectories of three broad response types with respect to a short-term challenge were simulated which were, Fully Resilient (not affected by the challenge), Partially Resilient (affected but recovered after a period), and Non-Resilient (permanently affected by the challenge). The simulations included individual variation within response types. The ability of the RI to discriminate correctly between the response types was assessed assuming that target trajectories were unknown and using different methods to estimate these. Across all simulated scenarios, it was found that all RI could correctly distinguish Fully Resilient from Partially or Non-Resilient animals. However, only log-variance, integral and mean of squares correctly identified the Partially Resilient response type as more resilient than the Non-Resilient type, and this required data both within and outside the perturbation period. The results of this study highlight the potential risk of misclassifying animals based on the diverse RI, and method and data requirements to overcome these. Furthermore, response to selection based on log-variance on actual disease resilience of a sheep population was evaluated using an existing mechanistic host-pathogen interaction model for gastro-intestinal nematode infection under various scenarios. The results show that selection for logarithm of variance can reduce the production potential but could increase resistance to parasite (lower faecal egg counts - FEC). Selection on resistance, however, can increase the resilience but is less likely to reduce the production potential.

THE INTERPLAY OF TEMPERATURE AND LABORATORY ADAPTATION: A CASE STUDY USING REDUCED GENETIC VARIATION

Goel, Prerna[1]; Nolte, Viola[2]; Schlötterer, Christian[1]

[1]Institute für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz, 1210 Wien, Austria; [2]Vienna Graduate School of Population Genetics, Vienna

The interaction of different environmental stressors is well-understood on the phenotypic level, but not yet on the genetic level. Experimental evolution offers a valuable approach for studying evolutionary dynamics in the presence of multiple environmental stressors. In the present study, we explored the adaptive response of multiple stressors by modifying one of them, temperature. A founder population consisting of two genotypes was evolved for 48 generations in a hot

(29°C) environment. The evolved population was split into a population that continued to evolve in the same environment and a population that evolved in a cold environment (18°C). Using a Pool-Seq of populations evolved for about 75 generations in the different temperature regimes, we identified a surprisingly high similarity in the genomic response. Nevertheless, some genomic regions showed a temperature-specific selection response. The high overall similar selection response between the two selection regimes highlights the importance of laboratory adaptation even in two vastly different temperature regimes. It is, however, not clear to what extent the differences in genomic response reflect temperature-specific adaptation or an interaction of laboratory adaptation with temperature.

THE INTERPLAY OF TEMPERATURE AND LABORATORY ADAPTATION: A CASE STUDY USING REDUCED GENETIC VARIATION

Goel, Prerna[1]; Nolte, Viola[2]; Schlötterer, Christian[2]

[1]Institute für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz, 1210 Wien, Austria; [2]Vienna Graduate School of Population Genetics, Vienna

The interaction of different environmental stressors is well-understood on the phenotypic level, but not yet on the genetic level. Experimental evolution offers a valuable approach for studying evolutionary dynamics in the presence of multiple environmental stressors. In the present study, we explored the adaptive response of multiple stressors by modifying one of them, temperature. A founder population consisting of two genotypes was evolved for 48 generations in a hot (29°C) environment. The evolved population was split into a population that continued to evolve in the same environment and a population that evolved in a cold environment (18°C). Using a Pool-Seq of populations evolved for about 75 generations in the different temperature regimes, we identified a surprisingly high similarity in the genomic response. Nevertheless, some genomic regions showed a temperature-specific selection response. The high overall similar selection response between the two selection regimes highlights the importance of laboratory adaptation even in two vastly different temperature regimes. It is, however, not clear to what extent the differences in genomic response reflect temperature-specific adaptation or an interaction of laboratory adaptation with temperature.

THE EVOLUTIONARY BASIS OF SUSTAINED RAPID PHENOTYPIC EVOLUTION OF POLYGENIC TRAITS

Dadashi, Andisheh; Gulisija, Davorka

University of New Mexico

The unprecedented changes caused by global climate change are exerting increasing selective pressures on natural populations. To persist, many populations need to rapidly evolve or shift habitats. Polygenic traits have been shown to evolve rapidly in response to strong selective pressure. Understanding the evolutionary mechanisms underlying phenotypic evolution in complex traits is, therefore, crucial for predicting and mitigating the negative effects of global

climate change. In this study, we employ a theretical genetic model and forward-in-time computer simulations to investigate the whole-genome molecular basis of phenotypic response to rapid and sustained environmental change in finite populations. We examine the effects of a broad range of population and genetics parameters, including starting allele frequency distributions, strength of selection, and mutation rates on polygenic evolution in repose to uni-directional changes in environments. We find that shifts in phenotypes are primarily underlined by shifts in the distribution of allele frequencies, modulated by selection, and are rather robust to starting conditions. By elucidating the genetic mechanisms driving phenotypic evolution in polygenic traits, this research contributes to the understanding the long-term adaptation of populations to global climate change.

THE GENETIC ARCHITECTURE OF CLIMATE ADAPTATION UNVEILED BY 1,628 WHEAT GENOMES

Guo, Yafei[1]; Song, Xinyue[2]; Kang, Lipeng[3]; Xu, Jun[1]; Xu, Song[2]; Zhang, Zhiliang[3]; Zhang, Jijin[1]; Jiang, Liping[2]; Qing, You[3]; Wang, Jing[3]; Yin, Changbin[3]

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, China.; [2]University of Chinese Academy of Sciences, Beijing, China.; [3]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.

Originating from domestication centers, crops dispersed to new climate zones where concomitant genetic changes occur during the adaptive process. Bread wheat, a cereal crop domesticated in the Fertile Crescent, demonstrates a remarkable environmental adaptability across extensive longitudinal and latitudinal regions. Despite being the most widely cultivated cereal crop globally, bread wheat's genetic architecture of climate adaptation remains a puzzle. Through the analysis of whole-genome variants in 1,628 bread wheat accessions grown worldwide, we have identified loci associated with environmental adaptation and dissected adaptive differences influenced by geographic and climate factors, such as solar, temperature, precipitation and soil. Our findings provide evidence that the adaptation of bread wheat to latitudes is more intricate than to longitudes, despite its primary spread along the latter. We have unveiled a constraint and polygenic basis for the climate adaptability of bread wheat, identifying a suite of genes enriched with signatures of repeated local adaptation to climate, potentially linked to adaptations such as flowering time. These results reveal the genetic foundation of bread wheat's climate adaptation and can help develop new crops that adapt to specific geographical regions and changing climates.

CAN METABOLOMIC-GENOMIC PREDICTION IMPROVE ACCURACY OF PREDICTED BREEDING VALUES IN PIGS?

Accurate prediction of breeding values is paramount in modern animal breeding programs. Increasing the accuracy of predicted breeding values can be obtained by incorporating additional information sources, such as various omics data, and a promising source is metabolomic data, which is the last biological layer before phenotypes. Therefore, integration of metabolomic information into genomic evaluation has become a topic of interest in recent years. In this study, we explored the potential for increasing the accuracy of breeding value by using a metabolomic-genomic model (MGBLUP) to integrate metabolomic data into genomic prediction. We conducted our investigation using a dataset comprising 8,289 Duroc pigs born between 2020 and 2022, from the DanBred breeding system. Approximately half the pigs were males from a test station and the other half were females from breeding herds. All the pigs investigated were phenotyped for average daily gain in the interval 30kg to 100kg, genotyped for 33,960 genetic single-nucleotide polymorphism markers, and profiled for a wide array of Nuclear Magnetic Resonance metabolomic features (MFs) with 30,468 intensities ranged from 0 to 10 parts per million (ppm). The variance components were estimated by applying MGBLUP on the full dataset. To evaluate the MGBLUP model in a practical breeding setting, we employed two validation strategies: test station to breeding herd validation (TB) and 5-fold cross-validation (5F). Increase in accuracy of predicted breeding values from successive inclusion in the validation population (VP) of metabolomics data and further of phenotypes was investigated using the linear regression (LR) method. A genomic BLUP model (GBLUP) was carried out as a baseline so that the results from MGBLUP can be compared with this. Estimation of parameters with MGBLUP model shows that there is similar amount of total phenotypic variances can be explained by genomic and metabolomic effects, and comparing this to GBLUP results, it is seen that most of the metabolomic variance comes from a reduced residual variance. The population accuracy of predicted breeding values when only genomic data is available for VP (MGBLUPg), is similar to the accuracy when both genomic and metabolomic data is available for VP (MGBLUPmg), while when further including phenotypic data into VP (MGBLUPpmg), a substantial increase in accuracy is observed. The comparison between MGBLUPg and MGBLUPmg represents the level of improvement in genomic prediction when including metabolomic information, while the comparison between MGBLUPmg and MGBLUPpmg represents the level of improvement when including phenotypes. Therefore, these results show that metabolomic data from the VP is not improving genomic prediction.

TOWARDS THE OPTIMAL APPROACH TO ASSESS INDIRECT GENETIC EFFECTS IN DAIRY CATTLE

Hansson, I.[1]; Bijma, P.[2]; Fikse, W.F.[3]; Rönnegård, L.[4]

[1]Department of Animal Biosciences, Swedish University of Agricultural Sciences, Box 7023, SE-750 07 Uppsala, Sweden; [2]Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700 AH, Wageningen, The Netherlands; [3]Växa, Swedish University of Agricultural Sciences, Ulls väg 26, SE-

756 51 Uppsala, Sweden; [4]School of Information and Engineering, Dalarna University, SE-791 88 Falun, Sweden

The social interactions imposed by surrounding cows in a dairy herd may impact an individual's production traits, e.g., milk yield. These interactions are believed to have a genetic component that can be modelled in terms of indirect genetic effects (IGE). IGEs contribute to heritable variation in pigs and poultry, but studies on IGEs in cows are scarce, and the size and importance of these effects are unknown. There is also a lack of knowledge of appropriate methods to monitor social interactions in dairy cows. This study assesses whether we can estimate IGEs in cows based on the social contact structure in dairy herds. We performed a simulation study to investigate how herd size (50, 100, or 200 cows), the correlation between direct genetic effect (DGE) and IGE (-0.6, 0, 0.6), and the size of IGE (where the total social effect explained 60, 30 or 3% of the phenotypic variance, VP) could affect the estimation. For the basic scenario, we simulated 100 replicates of social networks in 100 herds. Each herd contained 100 cows with unrelated dams and sires randomly mated from a base population of 10,000 cows and 100 sires. The number of contacts per cow was randomly drawn from a Poisson distribution with a mean of 30. Phenotypes for milk yield were simulated with a fixed herd effect, a DGE, the sum of the IGEs and indirect environmental effects (IEEs) from the individuals a cow had contact with, plus a residual. A phenotypic standard deviation of 800 and a direct heritability of 0.3 were used, and the variance of IEEs was assumed to be equal to the variance of IGEs. The estimates of the variance components for the basic scenario were unbiased, with moderate to high accuracy and unbiased EBVs, and similar results were found when altering the herd size and the correlation between DGE and IGE. When reducing the social effects to explain only 3% of VP, we could estimate the size of the variance components quite well but with large standard errors and more difficulties in getting the models to converge. Our results show that we were able to estimate IGEs based on social contact data in dairy cows. We further aim to assess the importance of knowing the intensities and direction of contacts to provide guidance on the necessary monitoring strategy of social interactions between cows. This study will be a step forward in understanding the optimal approach to assess IGEs in dairy cattle.

OPTIMIZING BREEDING PROGRAM DESIGNS THROUGH EVOLUTIONARY ALGORITHM

Hassanpour, Azadeh[1]; Geibel, Johannes[2]; Simianer, Henner[3]; Rohde, Antje[4]; Pook, Torsten[5]

[1]1University of Göttingen, Department of Animal Sciences, Animal Breeding and Genetics Group, Albrecht-Thaer-Weg 3, 37075, Göttingen, Germany; [2]2Center for Integrated Breeding Research, Carl-Sprengel-Weg 1, 37075, Göttingen, Germany; [3]3Institute of Farm Animal Genetics, Friedrich-Loeffler-Institute, Höltystraße 10, 31535 Neustadt, Germany; [4]4BASF Belgium Coordination Center CommV, Technologiepark 101, 9052 Gent Zwijnaarde, Belgium; [5]5Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, Netherlands

Modern breeding programs aim to improve several breeding objectives (progress in performance traits, maintenance of genetic diversity, etc.) simultaneously. Typically, the considered objectives are in conflict with each other, and the design parameters in a breeding program (number of selected /phenotyped /genotyped individuals) affecting the breeding goals are highly interdependent. Stochastic simulations with programs such as MoBPS have become a useful tool for breeders to evaluate different scenarios and understand how changes in the design parameters of a breeding program may influence outcomes. The optimization of a breeding program design using stochastic simulation is complicated by the fact that the output of a simulation is only the realization of a stochastic process. Thus, multiple replicates would be necessary to increase the precision of the results. Since breeding programs often involve numerous parameters, it is not feasible to simulate all possible breeding designs many times, as simulating a real-world breeding scheme can be computationally expensive. Therefore, users are limited to examining only a restricted range of potential scenarios to identify the best possible outcome among the set of considered ones. Due to these limitations, there is a need for an efficient optimization approach to improve breeding program designs. For this purpose, we suggest the use of evolutionary algorithms. To do this, we initially choose parameter settings randomly from the entire range of potential breeding programs. Parameter settings with the best value of the objective function (e.g. genetic gain or loss in genetic diversity) will be selected as parents. The "offspring" settings are created either by "recombination", which involves taking two existing parents and combining their information to create a new parameter setting, "mutation", which introduces small, random changes to a single parent, or by introducing new parameter settings randomly to overcome the problem of convergence to a local maximum. Following that, the optimal parameter settings will undergo simulation, iterating through this process until convergence is attained. Our optimization pipeline benefits from the automation provided by the Snakemake workflow management system, which can easily integrate with our iterative optimization approach, simplifying the execution of a set of tasks that need to be run in a specific order regularly. Our algorithm proved successful in a toy breeding program, highlighting its effectiveness.

EFFECT OF INBREEDING ON HEIGHT IN BROWN SWISS CATTLE

He, Qiongyu; Kadri, Naveen; Pausch, Hubert

ETH Zurich, Institute of Agricultural Sciences, Animal Genomics Group

Low effective population size in domestic cattle breeds leads to inbreeding and an accumulation of deleterious alleles in homozygous state that reduce the fitness of inbred individuals referred to as inbreeding depression. Understanding inbreeding depression is challenging when causal variants underlying recessive diseases or reduced fitness are unknown. Small sample sizes can further complicate analysis. This study aims to determine how inbreeding affects height in Brown Swiss (BSW) cattle and examine which genetic variants underly inbreeding depression. We also explore whether an easy-to-measure trait like height can serve as a proxy to better understand the pathology underlying

inbreeding depression. The average genomic inbreeding coefficients for 15,306 BSW cattle, based on the fraction of an individual's genome in runs of homozygosity (FROH) calculated from imputed genotypes at 20,039,070 sequence variants was 0.369 (±0.022). We examined inbreeding depression through a linear mixed model where we regressed height on FROH with the top four principal components (PCs) of a genomic relationship matrix as covariates. We found inbreeding depression is significantly affecting height in BSW (p=4.45*e-10). We also observed negative effects significantly associated with inbreeding arising from long (> 2 Mb), medium (0.1 - 2 Mb), and short (< 0.1 Mb) ROH (p=4.19e-09, p=0.001 and p=0.006, respectively). The observation that long ROH exhibit the most significant p-values suggests that recent inbreeding is more likely the primary factor contributing to inbreeding depression than ancient inbreeding. Next, we examined if recessively acting alleles that contribute to inbreeding depression on height can be identified using non-additive genome-wide association analysis (GWAS). Association testing with imputed genotypes of 20,039,070 sequence variants identified a novel recessive QTL for height on BTA25 with the top associated SNP at 14,462,320 bp (p = 5.55e-40). The ABCC6 gene encoding a protein essential for cholesterol transport and lipid metabolism is nearby the QTL. Abnormal expression of ABCC6 had been associated with various disorders like Pseudoxanthoma elasticum (PXE), which leads to connective tissue mineralization in humans, as well as dyslipidaemia and atherosclerosis, causing lipid and lipoprotein metabolism disorder in both mice and humans. These disorders might indirectly affect growth. Our study shows inbreeding depression on height in BSW cattle. We also uncover a recessive allele that reduces height through non-additive association testing. Although no disease symptoms are apparent in cattle with two copies of the recessive allele, accumulating such deleterious alleles with mild effect may reduce population fitness and elevate future disease risks.

INDIRECT GENETIC MODELS INCLUDING EARLY LIFE SOCIAL EFFECTS FOR EAR DAMAGE IN PIGS

Hegedus, Bernadett[1]; Galoro Leite, Natália[2]; Elizabeth Bolhuis, J.[3]; Bijma1, Piter[3]

[1]Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg , 6700AH Wageningen, the Netherlands; [2]Adaptation Physiology, Wageningen University & Research, De Elst 1, 6708WD, The Netherlands; [3]Topigs Norsvin Research Centre, Meerendonkweg 25, 5216 TZ, Den Bosch, The Netherlands

Introduction As pigs are housed in groups, it is essential to consider the effect of social interactions on breeding decisions. Behavior traits are difficult to measure. Thus, damage signs observed on victims of harmful behavior can serve as proxies. However, those only account for the recipient aspect of the trait. Given the availability of damage observations and pen information, social interaction models including indirect genetic effects can be used to study the actor component of biting behavior. Indirect genetic effects are already known in the animal breeding community. However little attention has been given to

indirect effects that are due to early life experiences, such that litter mates share, and how these effects affect the estimates of indirect genetic effects. Scientific question Here, we analyzed ear damage scores provided by Topigs Norsvin. We estimated genetic parameters with a traditional animal model including pedigree relationships in ASReml. We extended this model with indirect genetic effects and also with indirect litter effects that represent the early life experiences that litter mates share. The aim was to investigate whether 1) there are heritable effects in the social interactions between the animals, 2) the early life experiences affect the damage inflicted on pen mates and 3) the indirect genetic effects are overestimated if the indirect litter effect is not included in the model. Results For ear damage we found significant direct and indirect genetic variances and a significant indirect litter variance. Traditional heritability (h2), i.e., the heritability of direct effects, for ear damage was 0.03. T2 represents the total heritable variation that also includes indirect genetic effects in proportion to phenotypic variance. T2 for ear damage was 0.50 when excluding the indirect litter effect, and 0.37 when including these in the model. This difference highlights that early life experiences can play a substantial role in behavioral development, and can cause bias in estimated indirect genetic effects when not included in the model. Nevertheless, results from the full model indicate that the total heritable variation available for breeding is more than 10 fold greater than suggested by the traditional h2 estimate. Implications Our results show that there is heritable variation in the social interaction between animals with regard to ear damage. By using the total heritable variation, breeders can chose animals that function better in groups. However, close attention should be paid to the model terms to not inflate this variation. Our results also highlight the importance of the early life experiences of pigs in shaping their social effects on phenotypes of others.

THE GENOMIC BASIS AND EVOLUTION OF IMMUNE TRAIT VARIATION IN SOAY SHEEP

Henderson, Gina[1]; Sparks, Alexandra M.[2]; Pilkington, Jill G.[3]; Pemberton, Josephine M.[1]; McNeilly, Tom N.[2]; Nussey, Daniel H.[3]; Johnston, Susan E.[3]

[1]Institute of Ecology and Evolution, University of Edinburgh, Charlotte Auerbach Road, Edinburgh, EH9; [2]FL, United Kingdom. 2School of Biosciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom.; [3]Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik, EH26 0PZ, United Kingdom.

Immune responses involve trade-offs between increased resistance to infection and fitness-related traits, such as reproduction and survival. Quantitative genetic theory predicts that strong selection will lead to a reduction in genetic variation for immune traits, yet often this variation persists. Investigating the direct relationship between genetic variation and fitness traits will help us to better understand this phenomenon. Wild Soay sheep on St Kilda, Scotland, have high genetic variation underlying the antibody isotype IgA in response to a common gastrointestinal parasitic infection caused by Teladorsagia circumcincta (h2 = 0.57, IA = 0.38). In this study, new methods in genomic association studies are used to assess the genetic architecture of IgA in the Soay sheep (N

= 6,543 IgA measurements from 3,190 sheep). We integrate genomic data from 420K SNPs (all sheep) and whole genome sequence data (N = 134 sheep) to confirm that IgA variation is underpinned by two major effect loci, TNFRSF17 (BCMA) and the IGH complex. We also explore associations between genotypes and three fitness measures: juvenile survival, adult annual survival, and adult annual reproductive success. This study highlights the importance of using a wild dataset to investigate the selective importance of IgA levels, providing an insight into the adaptive significance of variation in immune responses.

MAPPING IMMUNE CELL GENE NETWORKS WITH SINGLE CELL DATA FOR HUMAN AND ANIMAL HEALTH

Hillis, Richard[1]; Shirali, Dr Masoud[2]; Overton, Dr Ian[2]

[1]School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Northern Ireland,; [2]Agri-food and Biosciences Institute (AFBI), 18a Newforge Lane, Belfast, Co Antrim, Northern Ireland.

The immune system functions through complex networks of interactions involving numerous components across a number of different immune cell types. These networks connect genotype to immune and disease phenotype. Mapping biochemical interactions in immune cells provides a framework to efficiently explore the underlying genetic factors, including epistasis, in immune response and individual disease susceptibility. Leveraging functional genomics data for different immune cell types, including single cell transcriptomics and Gene Ontology semantic similarity scores we are training statistical and machine learning models for network inference. Accordingly, our approach predicts context-specific biochemical interactions between genes across a range of immune cell types. The dairy industry is a major component of the global agricultural economy, and one in which disease continues to be a burden, both to the economic performance of dairy farms and to the welfare of their animals. The spread of disease among farm animals also presents risks to human health such as the potential for development of zoonotic and antibiotic resistant infections. To be able to alleviate these risks it is necessary to better understand the underlying mechanisms which define immune response and disease susceptibility. A single cell RNA sequencing dataset for cattle peripheral blood mononuclear cells was downloaded from the gene expression omnibus which, after quality control filtering, provided data for 26,141 single cells, with expression values across 14,500 genes. These cells were split into 53 subsets based on cell types identified in the datasets associated journal article, the majority of which are immune cell subtypes such as CD8 TCM1, CD8 TEM1, CD14 monocyte, NK3 and NK4 among others. For each cell type, Pearson's correlation scores were calculated for co-expression between all unique gene pairs. Gene ontology annotations for Bos taurus genes were used to calculate semantic similarity scores for the gene pairs across each of the three sub-ontologies using Resnik's method. The co-expression and semantic similarity data were combined into cell-type-specific functional genomics datasets, with gold standard data developed from KEGG pathways used to label gene pairs as positive or negative, based on whether or not both genes were members of the same pathway,

allowing for supervised machine learning to classify pathway co-membership. The first machine learning methodology tested was a Bayesian logistic regression approach and the models produced will serve as a baseline for future testing of other machine learning architectures, including CosNI developed by the Overton group. The networks produced using these models provide a biochemical reference for integrated analysis of pathway and protein-complex regulation within both differential expression and GWAS data, providing insights into fundamental biology. The resulting predictions allow identification of genes and pathways important to different diseases and traits. Building upon these results we plan to produce network-informed risk stratification models for prediction of bovine immune status, disease susceptibility, and risk. Ultimately the outcomes would help to inform veterinary practice and future research. Indeed, our methods are broadly applicable to the study of immune systems in other animals and in humans, with application to immuno-oncology envisaged in future.

UNVEILING THE CHALLENGES: THE DIFFICULTY IN MAKING RELIABLE INFERENCES OF BETWEEN-POPULATION MEAN GENETIC DIFFERENCES FROM GWAS

Hivert, Valentin[1]; Revez, Joana A.[2]; Zeng, Jian[3]; Zheng, Zhili[4]; Goddard, Michael E.[5]; Wray, Naomi R.[3]; Yengo, Loic[4]; Visscher, Peter M.[5]

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia;; [2]Agriculture Victoria Research, AgriBio, 5 Ring Road, Bundoora, Victoria, Australia;; [3]Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, Victoria, Australia;; [4]Department of Psychiatry, University of Oxford, Oxford, UK;; [5]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK

It is well known that human traits and diseases exhibit mean genetic and phenotypic differences between populations. Although it is theoretically possible to infer mean genetic differences using polygenic scores (PGS) derived from genome-wide association studies (GWAS), making reliable inferences remains challenging. Here, we use genotype data from unrelated individuals of European and African genetic ancestries in the UK Biobank (UKB) and perform a simulation study to assess the performances of three PGS methods: Pruning and Thresholding (P+T), approximate conditional and joint (COJO) multiple-SNP analyses, and SbayesR. We found COJO to outperform the other methods and show that robust inferences are limited by the power of the GWAS experiment and LD between variants, resulting in upwardly biased estimated mean genetic differences. We emphasise that interpretation of between-population genetic differences inferred using PGS must be made with respect to its prediction error variance, that is function of the proportion of trait heritability explained by PGS. Finally, we infer mean PGS in four different genetic ancestries in the UKB across 28 blood biomarkers traits as well as height and BMI, finding significant mean PGS differences across populations. However, these differences become non-significant when considering genetic variation that is unexplained by PGS. We

caution that careful interpretation is warranted when comparing mean PGS differences across populations.

INTEGRATING DYNAMIC MODE DECOMPOSITION WITH GENOMIC PREDICTION TO PREDICT PLANT DEVELOPMENT

Hobby, David[1]; C.Heuermann, Marc[2]; Mbebi, Alain[3]; Tong, Hao[1]; Altmann, Thomas[2]; Nikoloski, Zoran[3]

[1]Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466, Seeland OT Gatersleben, Germany; [2]Systems Biology and Mathematical Modeling Group, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany; [3]Bioinformatics Department, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

Understanding the intricate dynamics of plant growth is vital for optimizing breeding strategies. This study introduces a novel methodology, termed dynamicGP, which merges Dynamic Mode Decomposition (DMD) with Genomic Prediction (GP) to address this challenge. DMD is a powerful technique used to analyze complex time-series data by extracting temporal patterns and identifying coherent dynamic structures in time-resolved data. We applied this method to analyze time-resolved phenomes, comprising diverse growth-related traits measured, in a maize MAGIC population. DMD unveiled the complex temporal relationships underlying various growth traits. Using genomic data, we investigate the heritability and predictability of key components of the DMD algorithm, including the operator which encapsulates system dynamics. This operator can be used to predict the plant phenome in future time-points using the phenome in a preceding time point. However, predicting this operator from genomic marker data (e.g. single nucleotide polymorphisms (SNPs)) is a computationally challenging task. To address the computational challenge, we devise a shortcut method that directly predicts the intermediate components of the algorithm from SNP data. We show that this method substantially reduces computational costs without sacrificing prediction accuracy. We used two approaches to assess accuracy of the dynamic phenotypes obtained from the the predicted operator: (1) recursive, beginning with the measured phenome values at the initial time point, the subsequent phenomes are obtained using predicted phenomes at each intermediate time point; and (2) iterative, using the measured phenome values at each time point to predict the phenome in the following time point. We demonstrated that the recursive version of dynamicGP resulted in lower prediction accuracies compared to the iterative version, particularly for later time points. We also observed that traits exhibiting consistent heritability over time demonstrated higher mean prediction accuracies across time. Importantly, we showed that the iterative version of dynamicGP outperformed rrBLUP baseline predictions by 5.7-fold for traits whose average accuracy of prediction was at least 0.7; further, the increase for the recursive version of dynamicGP was 2.1-fold for traits whose average accuracy of prediction was above 0.5 This integrated approach holds promise for unraveling the genetic foundations of complex plant growth dynamics and

facilitating the development of more targeted breeding strategies for enhanced crop production. Moreover, dynamicGP sets the stage for future investigations into prediction of crop developmental traits.

MULTIVARIATE BASIS OF DROUGHT RESISTANCE STRATEGIES

Hoerdemann, Lea; de Meaux, Juliette

De Meaux group, Institute for Plant Sciences, University of Cologne, Cologne, Germany

The frequency of drought is increasing in times of global climate change. Consequentially, drought stress evolves to be a global prevalent environmental challenge for plants. To face these problematic developments, it is crucial to understand plant response strategies to water withdrawal. Studying differences in drought response strategies in ecologically specialized Arabidopsis species helps to understand how species balance costs and benefits of drought stress reactions. Here, we investigate the multivariate basis of natural variation in drought response strategies between the sister species A. halleri and A. lyrata, which deploy distinct strategies to face drought stress. For this purpose, an interspecific backcross population of 240 genotypes was generated, and a set of 20 morphological, physiological, and stress-related traits was investigated in well-watered as well as stress conditions. Furthermore, we performed QTL (quantitative trait loci) mapping to analyse the underlying genetic architecture of relevant traits. We utilized multivariate and additional statistical approaches like structural equation modelling to identify causal relationships between known functional traits and their role in the adaptive process to drought stress. Collectively, segregating variation was detected in all functional traits within the backcross population. Our analysis of multi-trait relationships showed covariation of functional traits, forming a trait network. Combined with the results of principal component analysis (PCA) we identified groups of traits within the network, altogether mediating the plant´s response to drought stress. Many small effect QTL and few large effect QTL (quantitative trait loci) contribute to interspecific differences in these traits. Further, structural equation modelling will be used to dissect the role of each QTL in the divergence of drought response strategies. An analysis of differential gene expression during stress will add decisive insights into the underpinnings of key functional traits contributing to drought tolerance strategies. By identifying and mapping multivariate correlation patterns between key functional traits involved in drought reaction strategy differences, our study provides valuable information on the genetic factors that drive adaptive processes in stressful environments.

CHARACTERISATION AND QUANTIFICATION OF DELETERIOUS GENETIC VARIANTS IN NON-MODEL ORGANISMS: FROM PRESENT TO EXTINCT SPECIES

Höglund, Julia[1]; Hu, Seyan[2]; Derks, Martijn[3]; Dalén, Love[1]; Bosse, Mirte[2]

[1]Dept. of Animal Breeding and Genomics, Wageningen University & Research, Wageningen, THE NETHERLANDS;; [2]Centre for Palaeogenetics, Dept. of Zoology,

Stockholm University, Stockholm, SWEDEN;; [3]Section Ecology & Evolution, Amsterdam Institute for Life and Environment (A-LIFE), Vrije Universiteit Amsterdam, THE NETHERLANDS

Animals have always been exposed to extinctions. Now, during what can be seen as the sixth mass extinction event, human activities have been one of the main reasons for dramatic decline. When a population becomes smaller, threats affecting the species' genome become larger. It becomes more prone to inbreeding, good genetic variation will be lost, and damaging variation will increase. This leads to a worse ability to adapt to changing environments and could lead to extinction. Hence, there is a crucial need to characterise and quantify damaging variation and its contribution to species decline, both for conservational purposes and understanding what drove former species to extinction. One way to characterise variation is to score genetic variants based on their predicted deleteriousness. By comparing simulated and derived variants based on comparison between a species and an ancestral state, a model can be trained to differentiate between benign and deleterious variants based on annotations. These types of scores are almost always species specific, and they are currently still largely unreliable. Currently, they are only developed for some model species. Hence, there is a crucial need to expand beyond model species. Here, we used the reference genome of domesticated pig as a starting point, and substitution rates and mutation rates were estimated from the most recent common ancestor of pig and cow, sheep, horse and elephant respectively to compare the impact of phylogenetic scope in annotation and scoring. Preliminary results show similar results of predicted variant effect distribution in derived and simulated variants across ancestral nodes. Using dense sequencing data, the scoring model will then be trained within domesticated pig, validated with data from wild boar and lastly tested in sequencing data from endangered pig species. In contrast to previous similar scoring models, this model extends across similar species already in the initial score estimation. This scores with then be utilised in quantifying genetic load, with the long-term goal of estimating how much genetic variation is contributing to animal extinction.

COMPARING GENETIC AND PHENOTYPIC VARIANCE MATRICES

Holstad, Agnes; Hansen, Thomas F.; Pélabon, Christophe

AH & CP: Department of Biology, Norwegian University of Science and Technology; Trondheim, Norway. TFH: Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo; Oslo, Norway.

Evolution depends on heritable trait variation. This can be measured by the additive genetic trait variance that captures the variation of genetic effects that are directly transmitted from generation to generation. However, traits seldom evolve in isolation due to genetic correlations among them. Hence, the additive genetic variance matrix, G, measures the multivariate potential for evolution. The estimation of G requires breeding designs or pedigrees that can be difficult or even impossible to achieve in some cases, for instance, for fossil organisms. We therefore investigate if and when the phenotypic variance matrix, P, that

measures the multivariate total trait variation, can be used as a substitution for G to study the potential for evolution. P is estimated with higher precision than G, but will be biased by the environmental variance matrix, E. We first show that if the guestimate of heritability (proportion of additive genetic variance) is within a few percent (≤ 10%) of the true heritability, the bias corrected variance components of P are estimated with higher accuracy for the true G than the estimated G based on a half-sib design including up to hundreds of sires. We then use average heritabilities for morphological and life-history traits to correct P-matrices for bias, before we compare empirical estimates of G- and P-matrices and their individual components to show that the similarity depends on estimation error, the precision of bias correction and type of traits.

WRAP IT ALL UP - RECENT ADVANCES IN VARROA RESISTANCE BREEDING FOR HONEYBEES IN THE BEEBREED.EU FAMILY

The invasion of the Western honey bee (Apis mellifera) by the parasite Varroa destructor has changed the face of beekeeping. As the parasite cannot be eradicated, breeding is the most sustainable path the tackle the problem. In the past decades, the main approach has been to improve behavioural traits similar to those of A. cerana, the original host of Varroa destructor, such as high brood hygiene. A large number of auxiliary traits have been introduced and bees have been selected for them: general brood hygiene as measured by the pin test, Varroa-specific hygiene as measured by VSH and SMR, bee infestation measurements, mite mortality, brood infestation and many more. In breeding programmes supported by the BeeBreed.eu breeding value platform, associations have selected for these traits, some of which have shown dramatic progress. Among them, we present the results of a selection scheme based on video evaluation in observation hives, where the behaviour of individual worker bees is the selection criterion. However, despite the large progress in auxiliary traits the original problem of colony losses caused by Varroa parasitism and related diseases has not yet been solved and it is time to bring the strands together and refocus on the real desired outcome. We propose a novel concept to measure the ultimate success of the process, which is the survival under the omission of any Varroa treatment with a scaled evaluation of the overwintering strength and the colony development in the spring after the survival test as selection traits. Auspicious early results indicate that this is a beacon for breeding strains of bees that do not require treatment for Varroa, which promises to revolutionise apiculture.

THE POTENTIAL OF SPATIAL MODELLING FOR QUANTITATIVE GENETIC ANALYSIS OF TANZANIAN SMALLHOLDER CROSSBRED DAIRY CATTLE

Houaga, Isidore[1]; Mrode, Raphael[2]; Okeyo, Mwai[3]; Ojango, Julie[4]; Nziku, Zabron[5]; Nguluma, Athumani[1]; Djikeng, Appolinaire[2]; Lavrencic, Eva[6]; Ekine-Dzivenu, Chinyere[4]; Pocrnic, Ivan[5]; Gorjanc, Gregor[5]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK ;; [2]Centre for Tropical Livestock Genetics and Health, The

Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, UK;; [3]Scotland's Rural College, UK; International Livestock Research Institute (ILRI), Nairobi, Kenya;; [4]Tanzania Livestock Research Institute, Tanzania; 6-Sokoine University of Agriculture, Tanzania;; [5]Biotechnical Faculty, , University of Ljubljana, Kongresni trg 12, 1000 Ljubljana, Slovenia; [6]Scotland's Rural College, UK; 4-International Livestock Research Institute (ILRI), Nairobi, Kenya;

Small and scattered herds with low genetic connectedness characterize African smallholder dairy production systems. Coupled with the lack of an appropriate recording system, quantitative genetic analyses are challenged to deliver meaningful inferences and accuracy of breeding values, limiting genetic progress. One key underlying mechanism of achieving high accuracy is an accurate separation of environmental and genetic effects, which is notoriously low with small and genetically disconnected herds. Here, we assessed the impact of accounting for spatial variation between herds on quantitative genetic analysis of milk production in Tanzanian smallholder crossbred dairy cattle. To this end, we applied GBLUP-based models to 19,375 test-day milk yield records of 1894 genotyped crossbred dairy cows from 1386 herds. The cows had 664,822 SNP marker genotypes after quality control. We modelled herds as a random independent effect or a random spatially-correlated effect between the herd locations using the Matern covariance function via the "SPDE" approach implemented in inlabru R package. The results show a large amount of spatial variation and of estimated breeding values. The results also strongly indicate that spatial modelling of herd effects more accurately separated genetic and environmental effects than independent herd effects and thus increased the accuracy of breeding values. Further studies that integrate genotype-by-environment interactions are needed to further model the staggering variation in African smallholder crossbred dairy production systems.

THE GENETIC BASIS OF COLD AND DROUGHT ADAPTATION IN GRASSES

Hsu, Sheng-Kai; Schulz, Aimee; Hale, Charlie; Costa-Neto, Germano; Stitzer, Michelle; Miller, Zack; Wrightman, Travis; AuBuchon-Elder, Taylor; A Kellogg, Elizabeth; Cinta Romay, M; S Buckler, Edward

Institute for Genomic Diversity, Cornell University

Grasses (Poaceae), encompassing major cereal crops such as wheat, rice, and corn, are foundational to global food security. This family is one of the largest and most important plant families, containing around 780 genera and approximately 12,000 species distributed globally. Their presence spans from arctic tundras to tropical forests, covering approximately 40% of the Earth's land surface, excluding Greenland and Antarctica. Understanding the genetic mechanisms that underpin this adaptability is crucial for enhancing agricultural resilience against climate change. In this study, we investigate the genetic basis of cold and drought adaptation in over 500 grass species by integrating genomic and biogeographic data. We employ cross-species association approaches to elucidate the genetic variation associated with adaptation to key environmental

factors like precipitation and temperature. Our analysis aims to identify specific genes and genetic pathways that confer tolerance to these stressors, potentially uncovering overlap in the genetic mechanisms of drought and cold resistance that have evolved in these species. By delineating these genetic factors, our research contributes to the strategic development of crop varieties better suited to the changing climate, thereby enhancing food security.

THERE IS NO SOLUTIONS, ONLY TRADE OFFS.

In plant breeding, there is no singular solution to improving crop varieties, but rather a continuous process of optimization and trade-offs. Each breeding objective, whether it be yield, disease resistance, or drought tolerance, presents unique challenges that require balancing multiple factors. Modern breeding techniques, such as Marker-Assisted Selection (MAS), Quantitative Trait Loci (QTL) mapping, Genome-Wide Association Studies (GWAS), and Genomic Selection (GS), provide powerful tools to navigate these complexities. MAS and QTL mapping enable the identification of specific genes associated with desirable traits, streamlining the selection process. GWAS extends this by examining the entire genome to find genetic variations linked to phenotypic traits. GS further enhances predictive accuracy by using dense marker data and statistical models to predict the genetic potential of breeding candidates. Additionally, mating plans, new-omics, and other tools , allows breeders to capture and analyze a vast array of plant traits with precision. Each method has its limitations and requires careful consideration of the breeding context. In this talk, I will focus on the solutions and trade-offs that plant breeding faces, highlighting how breeders can optimize their strategies to achieve the best possible outcomes.

INFLUENCE OF EFFECTIVE MICROORGANISM SUPPLEMENTATION ON THE WATER AND ON GUT MICROBIOME OF COMMON CARP

Jakimowicz, Michalina[1]

[1]Biostatistics Group, Department of Genetics, Wrocław University of Environmental and Life Sciences, Kożuchowska 7 51-631 Wrocław; [2]Institute of Ichthyobiology and Aquaculture, Polish Academy of Sciences in Gołysz, Kalinowa 2 43-520 Chybie PL; [3]

The purpose of the research study was to investigate the impact of feed and water supplementation with EM (effective microorganism) on the gut microbial communities of Common carp (Cyprinus carpio), on fish growth performance and on water microbial composition. The composition was determined by high-resolution sequencing of two hypervariable regions (V3 and V4) of the 16S rRNA gene. The experimental setup comprised six tanks, each serving a distinct purpose: two control tanks containing fish without probiotic supplementation, two tanks supplemented with water supplement W1 and feed supplement F1 (design 1), and two tanks supplemented with W2 and F2 (design 2). The modeling of abundance diversity was performed on the genus taxonomic level. Taxonomic analysis revealed a diverse microbial landscape in both intestinal and water samples. For intestinal samples, the most abundant genera were

Acinetobacter, Cetobacterium, Lectobaciellus, Latilactobaciellus, while for the water sample, the most abundant genera included Candidatus Koribacter, Cetobacterium, Nocardioides, and Rhodanobacter. Alpha diversity was quantified by the Shannon index. In intestinal samples, the Shannon index took the highest value (2.86) for the tank from control group (A3), while the lowest value (1.07) for the tank of design 1 (A5). For water, the highest value (3.67) was observed in tank from design 1 (A4), where the lowest value (2.71) was observed in tank control group (A3). To assess beta diversity, the Bray-Curtis distance was chosen. When looking for intestinal samples, the greatest distance (0.98) was observed between the tank of the control group (A2) and the tank of design 2 (A7), while the shortest distance (0.24) was observed between tank from design 1 (A5), and design 2 (A6). For water samples, the greatest distance (0.82) was between the control group (A3) and design 2 (A6), while the shortest distance was between two tanks from the control groups (A2 and A3). Differential abundance analysis of particular genera revealed significant alterations in microbial composition only in two comparisons in intestinal samples – control vs. design 1, and control vs. design 2. For water, the differences in abundance were observed only in comparison between the control group and design 1. When looking at intestinal samples, the most differences between control and design 1, showed a reduced abundance of genera in design 1 compared to the control group, while comparison between control setting and design 2 showed an increase in bacterial abundance. In particular, the Lactococcus genus exhibited a consistent increase in abundance between the supplemented groups compared to the controls. For water, only Polynucleobacter showed a significant increase in abundance compared to samples from the control group and design 1. Furthermore, our investigation extended to the growth performance of fish under different supplementation regimes. After a 94-day observation period, a significant weight increase was observed, highlighting the potential beneficial effects of EM supplements on fish growth.

INTEGRATING DEEP PHENOTYPING AND WHOLE-GENOME SEQUENCING TO DECIPHER GENETIC BASIS OF FEED EFFICIENCY IN DAIRY CATTLE

James, Caelinn[1]; Fang, Lingzhao[2]; Wall, Eileen[1]; Coffey, Mike[2]; Li, Bingjie[2]

[1]Animal and Veterinary Sciences, Scotland s Rural College (SRUC), Roslin Institute Building, Easter Bush, Midlothian, EH25 9RG, United Kingdom; [2]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark.

Feed efficiency (FE) is a complex trait in dairy cattle and is highly linked to milk production and agricultural Net Zero. This study aims to unravel the genetic basis of dry matter intake (DMI) using 40-year longitudinal data on dairy FE in the UK. Previous research has revealed that the underlying genetic architecture of DMI varies across days in milk (DIM) during lactation; heritability estimates vary across DIM, and DMI in early lactation is in low or negative genetic correlation with DMI in mid and late lactation. Despite this, there has been little work to identify the genetic basis underlying DMI across lactation stages. A total of 701,457 daily DMI records were available for ~2,100 Holstein cows

(longitudinal records across multiple lactations per animal), together with animals' pedigree information, 80K SNP genotypes, and imputed sequence-level genotypes. We performed genome-wide association analyses (GWAS) and regional heritability mapping (RHM) for DMI across multiple lactations and lactation stages (early, mid, late lactation) using 80K SNP genotypes. Fine-mapping analyses were conducted afterwards using sequence-level genotypes to narrow down candidate regions associated with DMI at each lactation/stage. In the results, candidate QTLs in genetic control of FE were identified through systematic GWAS, RHM, and fine mapping analyses. The results showed significant differences between early and mid/late lactation in the genetic basis for feed intake, and differences between primiparous and multiparous cows. Through integrating deep phenotyping of FE and high-throughput genomic information, our findings offer deep insights into the dynamics of genetic basis of dairy FE across lactations, identifying candidate QTLs underlying FE that informs dairy breeding and genotyping chip design.

CROSSOVER: THE POTENTIAL OF SELECTION ON CROSSOVERS IN ANIMAL BREEDING

Jansen, A.C.M.; Wientjes, Y.C.J.; Katz, O.; Calus, M.P.L.

Animal Breeding and Genomics, Wageningen University & Research, The Netherland

Crossing-over of homologous chromosomes during meiosis generates variation in gametes, and thereby results in more variable offspring. Crossovers have the ability to break down linkage between alleles, resulting in new possible allele combinations in gametes. This can result in an increase of the additive genetic variance. It is therefore possible to achieve a higher response to selection with a higher CC. The number of crossovers is found to be variable between individuals, sexes, and species, and this variation is at least partly heritable. However, crossover count (CC) is currently not selected for in animal breeding programs. It is hypothesized that artificial selection in animal breeding programs has resulted in an increase in CC, because of its potential to increase the additive genetic variance. Besides that, the formation of crossovers is a necessary step for successful completion of meiosis. Therefore, the variation in CC is expected to be directly associated with variation in fertility. In addition, variation in CC may be indirectly correlated to other breeding goal traits. The aim of our CrossOver project is to investigate the potential and consequences of targeting CC in animal breeding programs. The potential of selecting on CC will be studied in a dataset from a pig breeding program, consisting of >10 generations of a dam line, and a sire line, with for both lines >100,000 animals genotyped for ~20,000 markers. Phenotypic records are available on breeding goal traits, including male fertility, female fertility, growth/feed efficiency traits, and the number teats (a high heritability trait). Firstly, the location and number of crossovers in the genotype data will be determined. Using that output, the heritability of CC and genetic correlations of CC with breeding goal traits will be determined. Secondly, the selection history of CC will be determined to detect if there has been indirect selection in the breeding program. Models that are available to study selection history will first be validated in a livestock population

using simulated data. Thirdly, the expected correlated response to selection will be determined for breeding goal traits if there would be direct selection on CC. This will be done using the genetic correlations between the breeding goal traits and CC. This will give insight in either positive or negative consequences of selection on CC regarding the overall breeding goal traits. We expect that the results will give insight in the potential value of CC in animal breeding programs, and thereby reveal possible consequences of selecting on CC directly.

A COMPREHENSIVE OVERVIEW AND BENCHMARKING ANALYSIS OF FAST ALGORITHMS FOR GENOME-WIDE ASSOCIATION STUDIES

Liu, Fang; Zhang, Jie; Zhao, Yusheng; H. Schmidt, Renate; Marscher, Martin; C. Reif, Jochen; Jiang, Yong

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben

Genome-wide association studies (GWAS) are a ubiquitous tool for identifying genetic variants associated with complex traits in structured populations. During the past 15 years, many fast GWAS algorithms based on a state-of-the-art model, namely the linear mixed model, have been published to cope with the rapidly growing data size. In this study, we provide a comprehensive overview and benchmarking analysis of 33 commonly used GWAS algorithms. Key mathematical techniques implemented in different algorithms were summarized. Empirical data analysis with 12 selected algorithms showed differences regarding the identification of quantitative trait loci (QTL) in several plant species. The performance of these algorithms evaluated in 10,800 simulated data sets with distinct population size, heritability and genetic architecture revealed the impact of these parameters on the power of QTL identification and false positive rate. Based on these results, a general guide on the choice of algorithms for the research community is proposed.

CHARACTERIZATION OF GENETIC REGULATORY VARIANTS ASSOCIATED WITH ENERGY BALANCE IN HOLSTEIN DAIRY COWS VIA RNA-SEQ

Improving the feeding efficiency of dairy animals is a key component of optimizing land resource utilization and meeting the demand for high-quality protein. However, a thorough understanding of the genetic variants regulating gene expression is essential for interpreting molecular mechanisms underlying key traits and their breeding value and for increasing genetic gain in genomic selection. Energy balance (EB) is an important complex trait at the postpartum stage, and as the intake of nutrients cannot meet the demand of rapidly increased milk production, dairy cows usually suffer from a negative energy balance post calving and are at increased risk of developing clinical or subclinical ketosis. This study seeks to characterize genetic regulatory variants associated with EB in Holstein dairy cows during the early lactation period. Data source came from Northern Ireland Farm Animal Biobank (N.I.FAB). The genetic basis of EB was examined by utilizing high-throughput RNA-sequencing and establishing the bulk blood transcriptomes of dairy Holstein during the early lactation period, differentiating the gene expression profile with phenotype records, and conducting gene ontology (GO) annotation and Kyoto Encyclopedia

of Genes and Genomes (KEGG) pathway analysis. RNA sequence data and phenotype information were acquired from 77 Holstein-Friesian cows from a single herd, fed with diets comprised of 4 varying protein levels. 8 differentially expressed genes (DEGs) across the herd were identified through differential gene expression tools of STAR and DESeq2 (P value < 0.05 and Log Fold Change > 0.584). These genes are associated with lipid modification, adipose tissue structure and function, and calcium activated chloride channels, reflecting early lactation energy transformation difference. DEGs (n=24392) were examined by GO annotation and KEGG pathway analysis. Enriched GO:BP terms highlight important of lipid metabolism including prostanoid, prostaglandin, icosanoid, unsaturated fatty acid and fatty acid biosynthetic. The result of KEGG pathway enrichment revealed most DEGs to be enriched in pathways related to Herpes simplex virus 1 infection, Natural killer cell mediated cytotoxicity and Bile secretion. These findings provide a novel insight into the genetic basis of EB and identify candidate genes and biological mechanisms associated with energy utilization, metabolic processes, and negative energy balance in early lactation period of dairy cows.

PREDICTION OF NON-NEUTRAL VARIANTS WITH TEMPORAL ALLELE FREQUENCY INFORMATION TO IMPROVE GENOMIC PREDICTION MODELS

Johansen, Natasha[1]; Sarup, Pernille[2]; Ramstein, Guillaume[2]

[1]1,3Center For Quantitative Genomics and Genetics, Aarhus University, 8000, Denmark; [2]2Nordic Seed A/S, 8300, Denmark

The aim of breeding is to accumulate beneficial variants. Detection of beneficial variants is a challenge for breeders as variant effects are dependent on both the genetic context (e.g. epistasis) and dependent on the environment (e.g. genotype x environment interactions). Previous studies have although documented improved model performance, when incorporating fitness information in genomic prediction models as utilization of fitness information have been shown to facilitate the detection of non-neutral variants. In this study we infer the relative fitness effect of observed variants, in a winter wheat (Triticum aestivum L.) breeding population, by tracking the temporal tendencies in allele frequency over a nine-year time-period. This allows for the identification of variants that have been under consistent selection across the breeding period. We hypothesize that utilization of fitness estimates in genomic prediction models can enhance model performance, as we predict that variants with non-neutral effects on fitness will belong to a different variance distribution than variants with a neutral effect on fitness. Based on the estimates of the relative fitness of observed variants we classified the SNPs into different 'fitness' categories i.e. deleterious, neutral, and beneficial, which allows for the construction of a multi-kernel genomic prediction models that assumes differential effect distributions for variants with non-neutral effects on fitness. We compare an extended genomic prediction model that incorporates fitness information to a baseline approach. The performance of these two approaches is compared and validated on recent breeding lines, and the genome-wide effect of the (putative) non-neutral variants are inferred for the traits: yield, protein quality, plant height

and heading date in different environments. Additionally, we investigate the genetic correlation, between traits, for each of the three SNP classes, i.e. (deleterious, beneficial, and neutral), to investigate for possible signals of genetic constraints between classes of SNPs. In our study we detected significant signals of selection. Additionally, we observed a significantly improved model fit for the fitness-informed model, which assumes different variance distributions for neutral and non-neutral SNP-classes, compared to the baseline model for the traits: yield ($p<0.0001$), heading date ($p<0.0001$) and protein quality ($p < 0.01$). However, this improvement in model-fit was environment-dependent, and further investigation is necessary to understand this pattern. Additionally, we documented that putative 'beneficial' variants significantly increased 'mean' protein quality in certain environments ($p<0.01$), while putative deleterious variants where not shown to have any significant effect on mean phenotypic performance for any of the traits, not in any of the environments. Consequently, in this study we found some improvement when utilizing fitness-information, however, depending on the focal species and the available data, different methods for inference of fitness estimates might be more appropriate e.g. comparative genomics methods that infers evolutionary constraints across species, or other annotation-based methods.

EFFECT OF USING PRESELECTED MARKERS FROM IMPUTED WHOLE-GENOME SEQUENCE IN GENOMIC PREDICTION IN ANGUS CATTLE

Kamprasert, Nantapong[1]; Aliloo, Hassan[1]; J. van der Werf, Julius H.[1]; J.Duff, Christian[2]; A.Clark, Samuel[2]

[1]University of New England, Armidale, NSW, 235Australia 2 Angus Australia, Armidale, NSW, 2350 Australia; [2]Angus Australia, Armidale, NSW, 2350 Australia

The advent of next-generation sequencing offers to the potential to use whole-genome sequence (WGS) data for genomic prediction. Improvement in genomic prediction (GP) accuracy using WGS has been observed in simulation studies but the advantage of WGS in GP has not been consistently observed in real data. Some benefit of WGS for GP has been shown when selecting a subset of markers with significant associations with the trait of interest and using these explicitly in the model for evaluation. The main objective of this study was to use different statistical methods to investigate the predictive ability of adding preselected markers to the standard-industry 50k SNP array for economically important traits in Angus cattle. Phenotypes for economically important traits in beef production were obtained from Australian Angus cattle, these included: birth weight (BW, kg) and carcass Intramuscular Fat (CIMF, %). Animals were genotyped with low to medium density; then, the genotypes were imputed to WGS. The final genotypes comprised 44,827 and 7,899,466 SNPs for 50k and WGS, respectively. Informative markers associated with the desired traits were extracted from WGS and then added to the 50k genotype. Several approaches were used to select different sets of informative markers, namely LD-based pruning, genome-wide association study, functional annotation based on Gene Ontology, quantitative trait loci from the Animal QTL Database, and sequence

annotation. To avoid marker redundancy, each subset of preselected SNPs was locally pruned based on LD, then clumped with the 50k. Eight different sets of genotypes were investigated: (1) 50k panel as a control (50k), (2) pruned WGS (PR1), (3) 200k SNPs randomly picked from pruned WGS (PR2), (4) 50k with added the GWAS outputs (50kGWAS), (5) 50k with added markers filtered with GO terms (50kGO), (6) 50k with added markers from the cattle QTL database (50kQTL), (7) 50k with added SNPs located on the coding region (50kSC), and (8) 50k with markers located on the genic region (50kSG). We applied different statistical models for the prediction of genomic breeding values, including GBLUP, BayesR, and BayesRC. Moreover, the preselected genotype sets were fitted with two GRMs constructed separately from the 50k. The estimated heritabilities was similarly estimated across different set of genotypes for all traits. The log-likelihood ratio values revealed that two-GRMs GBLUP fit the data better than the single GRM GBLUP. The Bayesian methods indicated that BW was oligogenic, while CIMF were more polygenic. Prediction accuracy ranged from 0.619±0.007 to 0.645±0.008 for BW, and from 0.545±0.019 to 0.587±0.022 for CIMF. There was no significant difference in prediction accuracy and bias across different statistical methods. For BW, the Bayesian models slightly outperformed GBLUP. The GWAS, a method validated within the population, pointed out the potential benefit of preselected markers in GP for BW. In conclusion, the prediction accuracy using preselected variant sets is heavily influenced by the population structure, the method employed for variant selection, and the genetic architecture of traits.

TRAVERSING THE GENOME CONSTRAINT: THE ROLE OF REGULATORY PLASTICITY IN WHEAT ADAPTATION

Kang, Lipeng[1]; Zhang, Zhiliang[2]; Dong, Jiayu[3]; Guo, Yafei[1]; Xu, Jun[2]; Xu, Song[3]; Zhang, Jijin[1]; Niu, Zelin[2]; Bi, Aoyue[3]; Xu, Daxing[1]; Qiu, Xuebing[2]; Jiang, Liping[3]; Song, Xinyue[1]; Niu, Beirui[2]; Zhu, Bingjie[3]; Li, Yiwen[1]; Wang, Jing[2]; Yin, Changbin[3]; Lu, Fei[3]

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.; [2]University of Chinese Academy of Sciences, Beijing, China.; [3]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. #These authors contributed equally to this work.

Understanding the mechanisms underlying adaptive innovation is a key to crop breeding. In this study, we sequenced 18 Triticeae species to construct a high-resolution genome constraint map of wheat, which captures evolutionary footprints of 74 Poaceae species covering ~101 million years (MYA) evolutionary history. An atlas of conserved non-coding sequence (CNS) was built, with 2.8 million CNSs identified and proved to be potential regulatory elements based on multi lines of evidence. Our findings revealed that CNS offers greater environmental adaptability compared to coding sequence (CDS) by swiftly replacing deleterious mutations, thus facilitating wheat's rapid adaptation to diverse global environments. At the species level, 2,355 accelerated CNSs were

found to converge on photosynthesis in Triticeae crops through polygenic selection. Our results showed that regulatory elements play a pivotal role in the adaptation of both rapid environmental response and barely unchanged photosynthetic efficiency, underscoring their critical importance in future breeding for enhancing both adaptability and productivity in changing environments.

Genetic parameters for genetic variance uniformity in Swiss pigs' birth weight

Kasper, C.[1]; Lepori, A.[2]; Gutiérrez, J.P.[3]; Formoso-Rafferty, N.[1]; Khayatzadeh, N.[2]; Sell-Kubiak, E.[3]; Cervantes, I.[3]

[1]Agroscope, Animal GenoPhenomics, Switzerland, SUISAG, Switzerland, ABS Global, Inc., USA,; [2]Department of Animal Production, Veterinary Faculty, UCM, Spain.; [3]Department of Agricultural Production, ETSIAAB, UPM Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Poland

Selection for increased litter size in pigs led to greater within-litter birth weight variation, necessitating cross-fostering to regulate litter size and homogenize the weights of piglets reared together. This practice increases labour costs and raises immunity concerns, but increases preweaning survival and therefore farm productivity. Selection for uniformity in birth weights (BW) could lead to a more ethical and efficient livestock production, as reducing the sensitivity to microenvironmental changes is expected to increase robustness, feed efficiency, survival and welfare, and reduce intrauterine growth retardation, thereby improving piglet health and carcass value. BW variability is a maternal genetic trait, whereas individual BW may also be partially influenced by direct genetic effects. This study aimed to estimate the genetic component of the residual variance for BW using two datasets, one from an experimental farm with 43,135 BW records from 986 sows and pedigree data for 45,737 individuals, and another one of similar size from two commercial farms. To estimate the genetic component of the residual variance for BW, models assuming heterogeneity of the residual variance (heteroscedastic model solved via double hierarchical animal models), were fitted using ASReml. The full model included the additive genetic effect for BW and the maternal genetic effect for BW and its environmental variance. The litter effect as well as several systematic effects, such as parity, age, sex, litter size, farrowing batch, and sire breed were included. In addition to the maternal genetic covariance between BW and its environmental variance, covariances between direct genetic and maternal effect and between direct genetic and maternal genetic effect for the environmental variance were analysed. The fit of models containing different sets of random effects was compared. The full model for BW had the best fit and yielded an additive genetic variance close to zero, a maternal genetic variance of 0.03, and a litter variance of 0.01. The genetic coefficient of variation for the residual variance was 0.27 and the litter variance for environmental variance was 0.10. The genetic correlation between the mean BW and its variability at the sow level was positive (0.31 ± 0.07), as was the genetic correlation between direct genetic effect and maternal effect (0.04 ± 0.30). The genetic correlation between direct genetic effect and maternal genetic effect for the environmental variance was

negative (-0.43 ± 0.36). These results were confirmed in the commercial farm data set, except for the litter, the environmental litter effect and the genetic correlation between direct genetic effect and maternal effect, which were considerably lower in the commercial farm data set. These preliminary results showed that modelling BW and its environmental variance, including only the maternal genetic effect for both, could be used to provide accurate estimates of BW variability for breeding programs, as the additive genetic effect for BW is negligible. The results are consistent with previous genetic parameter estimates for BW uniformity in pigs. Even though the mean BW might decrease, selection for uniformity is expected to be beneficial since selection for uniformity has been shown to increase the animal's robustness.

PARENT-OF-ORIGIN EFFECTS IN BIRTH WEIGHT IN LARGE WHITE PIGLETS: DISENTANGLING GENOMIC IMPRINTING AND MATERNAL EFFECTS

Kasper, Claudia[1], Principal Investigator/Group Leader; Jahnel, R. E.[2]; Reinsch, N.[2]; Lepori, A.[3]; Khayatzadeh, N.[4]

[1]; [2]Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl- Allee 2, 18196 Dummerstorf, Germany; [3]Suisag AG, Allmend 10, 6204 Sempach, Switzerland; [4]ABS Global, Inc., USA

Piglets' birth weights are key to the survival and homogenous growth of litters and are, therefore, an important trait for an ethical, sustainable and cost-effective pork industry. This trait is influenced by parent-of–origin effects, such as maternal genetic effects and genomic imprinting. Maternal effects are already implemented into models used for genetic evaluations of birth weights in many different breeding programs. Moreover, previous studies have found significant imprinted loci associated with birth weights in pigs. However, none of these studies have accounted for maternal effects and genomic imprinting using only one model. Neglecting one epigenetic effect could result in an overestimation of this effect, where the effects are mistakenly confounded. Thus, the objective of this study was to estimate genomic imprinting effects and maternal effects in birth weights simultaneously. Birth weights of 42,367 Swiss Large White piglets with 49,734 individuals in the pedigree were collected between 2004 and 2022. Animal models including gametic effects as sire and as dam, as well as maternal genetic effects were implemented into Echidna MMS. A direct heritability of 0.13 and a maternal heritability of 0.03 was estimated. We found a negative genetic correlation between the gametic variance as sire and as dam of -0.61 and a small positive genetic correlation between gametic variance as dam and maternal genetic effects of 0.22. The share of the imprinting variance of the direct variance was 29.92%. Based on the assumption of previous studies that both effects exist in this polygenic trait, we separated genomic imprinting effects from maternal effects. This resulted in a shift in the values of the variance components. For the scenario when imprinting effects were not accounted for in the model, the maternal genetic variance was higher than the direct variance. When estimating a gametic variance as dam and as sire instead of one direct variance, the gametic variance as dam had the highest values, indicating that some amount of the variance accounted for genomic imprinting might be

confounded in the maternal genetic effects if not accounted for in the statistical model. Here, we present preliminary evidence that genomic imprinting as well as maternal effects contribute to the birth weight in pigs. This is a first step towards better understanding epigenetic effects that influence birth weight expression for future selection decisions.

INVESTIGATING THE POTENTIAL FOR SELECTING PIGS ON RECOMBINATION RATE

Katz, Olivia; C.J.Wientjes, Yvonne; C.M.Jansen, Anne; P.L.Calus, Mario

Animal Breeding and Genomics, Wageningen University & Research, The Netherlands

In the genomic era, the possibility to use genomic information for selection enables greater genetic gain due to higher accuracies in the estimation of breeding values and shorter generation intervals. However, genetic gain can also increase with higher additive genetic variance in the breeding population. Additive genetic variance can be released in a population when crossing over events occur during meiosis, where chromosomes exchange DNA, creating new combinations of alleles in the population. The number of these events -or recombination rate- is a complex trait which is at least partially heritable. Recombination rate shows variation between the individuals within a population, and even between different sexes and species. Considering the goal of animal breeding, this means that individuals with higher recombination rate are expected to have more genetic variation within their offspring and also have a higher chance of giving birth to animals of high genetic merit. Additionally, certain threshold of recombination rate appears to be important for correct meiotic segregation, and lack of recombinations negatively affects fertility in multiple species. To our knowledge, recombination rate hasn't been included as a breeding goal in breeding schemes. Hence, the consequences of directly selecting for a higher recombination rate is yet unknown. By using genetic analysis and a method to detect recombination events, we'll study if recombination rate correlates with fertility traits and other traits of importance in pigs, like average daily gain and number of teats. Thus, this project aims at studying the impact of selecting animals for recombination rate on additive genetic variance for important breeding goal traits. Moreover, it aims to see whether recombination rate can predict fertility performance, by investigating the correlation with traits like litter size, number of born alive piglets, and stillborn piglets. For the analysis, we will use a pig dataset composed of a dam and a boar line, with for both lines >100,000 animals genotyped for ~ 20,000 SNPs. If recombination rate enhances additive genetic variance for important traits, including it in breeding schemes will contribute to genetic improvement of the population. Moreover, when recombination rate can predict fertility performance, early animal selection will be possible for fertility traits.

IDENTIFYING THE IMPORTANCE OF NON-ADDITIVE EFFECTS WITH EXPERIMENTAL EVOLUTION STUDY.

Kelkar, Vidyadheesh; Goel, Prerna; Nolte, Viola; Schlötterer, Christian

Institut für Populationsgenetik, Vetmeduni, Vienna

Polygenic adaptation is typically studied under the assumption of an additive genetic architecture. Also many GWAS studies are relying on the assumption of purely additive effects. On the other hand, non-additive effects are ubiquitous in phenotypic analyses of individual genotypes. In order to resolve this apparent discrepancy, we designed an experimental evolution study to address the role of non-additive effects during adaptation. Experimental evolution starting from a founder population containing only two genotypes provides a very powerful appraoch to detect selection signatures even within a small number of generations (Burny et al., 2021). Here, we modify this approach by using two inbred founder stains (Samarkand and Oregon-R), but rather than performing reciproal crosses between the founder strains, we generated two types of founder populations, one with Oregon females and Samarkand males (OS), and the second one with Samarkand females and Oregon males (SO). This resulted in founder populations with the balanced (50:50) allele frequencies on the autosomes, but different frequencies on the X-chromosome (33:67 vs. 67:33). Since the same genetic variants are present in the two founder populations, with additive effects the same selection strengths should be observed for all chromosomes. Contrary to these expectations, we observed strikingly different selection response in the two founder populations, suggesting non-additive effects. We performed computer simulations to explore the impact of dominance and epistatis. While dominance could not explain the different selection response, epistatic interactions could explain the pattern. Interestingly, it did not matter whether epistatically interacting loci were in close proximity or only loosely linked, suggesting that the interacting loci may be rather unlinked. In the light of the pronounced epistatic effects on the X-chromosome, it was remarkable that the direction of the cross did not matter for the selection response on autosomes. The highly parallel selection response on the autosomes suggests that epistatic interactions between chromosomes are sufficiently weak to be ignored. Our study demonstrated that epistatic interactions are a major factor determining the selection response within chromosomes, but not between chromosomes. We anticipate that this result will have a major impact not only for the interpreation of polygenic adaptation signatures, but also for GWAS signals.

FAMILYWISE HERITABILITY ESTIMATES OF THE BRAIN

Koten, Jan Willem; Teeuw, Teeuw; Hulshoff, Hulshoff; Pol, Pol; Boomsma, Boomsma

University of Graz, University Medical Center Utrecht, University Medical Center Utrecht, Vrije Universiteit Amsterdam

So far heritability is estimated on the level of entire populations. Here we present a novel quantitative genetics method that allows to estimate the heritability of brain shape and cortical thickness per extended twin family. Twin families consisted of a monozygotic twin pair with a sibling. The curvature and cortical thinks of the brain were obtained using FreeSurfer. Cortical thickness and curvature data were analyzed with a novel sliding window technique that allows

to estimate the heritability of brain shape and cortical thickness per single family. Here we show that brain regions that are under genetic control vary substantially from family to family. Brain regions which high heritability did not overlap between families. The brains of every single family are characterized by a highly specific gene expression pattern.

RANK AGGREGATION-BASED FEATURE SELECTION FOR COMPREHENSIVE MULTI-BREED CATTLE CLASSIFICATION

Kotlarz, Krzysztof[1]; Slomian, Dawid[2]; Szyda, Joanna[3]

[1]Department of Genetics, Biostatistics Group, Wrocław University of Environmental and Life Sciences, Wrocław, Poland; [2]Department of Genetics, Biostatistics Group, Wrocław University of Environmental and Life Sciences, Wrocław, Poland; [3]Department of Genetics, Biostatistics Group, Wrocław University of Environmental and Life Sciences, Wrocław, Poland

The rapid advancement of high-throughput sequencing technologies has revolutionised genomic research by providing enormous access to large amounts of genomic data. However, the most serious disadvantage associated with the use of Whole Genome Sequencing (WGS) data is its statistical nature, the so-called p>>n problem, where the number of predictors i.e. variants (p) is much larger than the number of available phenotypic records (n). In this context, the 1000 Bull Genomes Project has generated an extensive dataset comprising Single Nucleotide Polymorphisms (SNPs) from WGS representing diverse bovine breeds. The primary aim of this study was to present a methodology aimed to circumvent the p>>n problem demonstrated through the practical and biological application of classifying 5 breeds in the context of 1945 individuals and 11,897,040 SNPs. To address this challenge, we propose a novel approach that combines feature selection and deep learning techniques. In the first step, we employ the rank aggregation algorithm with linear logistic regression to prioritize informative features from the WGS data aggregated by a linear mixed model. Next, we introduce a Deep Learning (DL) algorithm adapted for multibreed classification tasks. The DL algorithm leverages the subset of informative features identified through rank aggregation to train a robust breed classification model. Our approach aims to capture complex patterns and interactions present in the genomic data, which allows for precise and effective breed classification across diverse bovine populations. Our work addresses the urgent need for computational techniques that are both effective and efficient in the analysis and interpretation of large-scale genomic datasets. We offer an extensive model for utilizing the predictive abilities of WGS data in multibreed classification tasks by fusing feature selection and deep learning techniques. The suggested method has the potential to improve breeding projects, make genomic selection easier, and expand our knowledge of the biological basis of complex traits in cattle populations.

A LARGE META-ANALYSIS OF HEALTH TRAITS IN HUNDREDS OF THOUSANDS OF GERMAN HOLSTEIN COWS

Križanac, Ana-Marija[1]; Reimer, Christian[2]; Heise, Johannes[3]; Liu, Zengting[4]; Pryce, Jennie[5]; Bennewitz, Jörn[6]; Thaller, Georg[7]; Falker-Gieske, Clemens[6]; Tetens, Jens[6]

[1]Department of Animal Sciences, University of Goettingen, Burckhardtweg 2, 37077 Göttingen, Germany; [2]Center for Integrated Breeding Research, Department of Animal Sciences, University of Goettingen, Albrecht-Thaer-Weg 3, 37075 Göttingen; [3]Center for Integrated Breeding Research, Department of Animal Sciences, University of Goettingen, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany; [4]Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut, 155 Neustadt, Germany 4Vereinigte Informationssysteme Tierhaltung w.V. (VIT), 2728 Verden, Germany; [5]Vereinigte Informationssysteme Tierhaltung w.V. (VIT), 27283 Verden, Germany; [6]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia; [7]Institute of Animal Science, University of Hohenheim, 70599 Stuttgart, Germany 8Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, 24118 Kiel, Germany

Health traits such as mastitis negatively impact milk performance and animal health and can lead to premature culling. Intensive selection for milk yield additionally contributed to a decrease in the health status of dairy cattle, mainly due to unfavorable genetic correlations between health traits and milk production. For these reasons, more knowledge is needed about the genetic architecture of health traits, in order to prevent economic losses and improve animal welfare. Genome-wide association studies (GWAS) made it possible to infer the relationships between genome regions and traits of interest and revealed many quantitative trait loci (QTL). However, a large number of individuals combined with whole-genome sequence (WGS) data is usually necessary to obtain reliable associations. In this study, our aim was to identify novel trait-specific and trait-shared QTLs in 11 common health traits that include: mastitis and somatic cell score, metabolic diseases (ketosis and displaced abomasum), and claw diseases (digital dermatitis, claw ulcers, interdigital hyperplasia, laminitis, interdigital phlegmon, white line disease). For this purpose, a large number of German Holstein cattle (100,809-180,217) were first imputed to the WGS level in a two-step imputation approach. Then, GWAS for 11 health traits was carried out using deregressed proofs as phenotypes. Meta-analysis was subsequently applied to individual GWAS summary statistics and functional and evolutionary trait heritability scores were assigned to candidate variants as established by Xiang and colleagues. Our analyses confirmed many previously reported and revealed novel QTLs.

A LARGE META-ANALYSIS OF HEALTH TRAITS IN HUNDREDS OF THOUSANDS OF GERMAN HOLSTEIN COWS

Križanac, Ana-Marija[1]; Reimer, Christian[2]; Heise, Johannes[3]; Liu, Zengting[4]; Pryce, Jennie[5]; Bennewitz, Jörn[6]; Thaller, Georg[7]; Falker-Gieske, Clemens[8]; Tetens, Jens[8]

[1]Department of Animal Sciences, University of Goettingen, Burckhardtweg 2, 37077 Göttingen, Germany; [2]Center for Integrated Breeding Research, Department of Animal Sciences, University of Goettingen, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany; [3]Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut, 31535 Neustadt, Germany; [4]Vereinigte Informationssysteme Tierhaltung w.V. (VIT), 27283 Verden, Germany; [5]Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia; [6]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia; [7]Institute of Animal Science, University of Hohenheim, 70599 Stuttgart, Germany; [8]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, 24118 Kiel, Germany

Health traits such as mastitis negatively impact milk performance and animal health and can lead to premature culling. Intensive selection for milk yield additionally contributed to a decrease in the health status of dairy cattle, mainly due to unfavorable genetic correlations between health traits and milk production. For these reasons, more knowledge is needed about the genetic architecture of health traits, in order to prevent economic losses and improve animal welfare. Genome-wide association studies (GWAS) made it possible to infer the relationships between genome regions and traits of interest and revealed many quantitative trait loci (QTL). However, a large number of individuals combined with whole-genome sequence (WGS) data is usually necessary to obtain reliable associations. In this study, our aim was to identify novel trait-specific and trait-shared QTLs in 11 common health traits that include: mastitis and somatic cell score, metabolic diseases (ketosis and displaced abomasum), and claw diseases (digital dermatitis, claw ulcers, interdigital hyperplasia, laminitis, interdigital phlegmon, white line disease). For this purpose, a large number of German Holstein cattle (100,809-180,217) were first imputed to the WGS level in a two-step imputation approach. Then, GWAS for 11 health traits was carried out using deregressed proofs as phenotypes. Meta-analysis was subsequently applied to individual GWAS summary statistics and functional and evolutionary trait heritability scores were assigned to candidate variants as established by Xiang and colleagues. Our analyses confirmed many previously reported and revealed novel QTLs.

EVOLUTIONARY MINING OF COLD TOLERANCE ALLELES IN MAIZE FOR FUTURE FARMING SUSTAINABILITY

Lai, Wei-Yun[1]; V. Franco, Jose A.[2]; Johnson, Lynn[3]; Costa Neto, Germano[1]; Romay, M. Cinta[2]; Stitzer, Michelle C.[3]; Buckler, Edward S.[3]

[1]Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA 14853;; [2]Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY USA 14853;; [3]USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA 14853

Shfting the maize growth cycle by planting earlier in the season can lead to better utilization of sunlight and prolonged land coverage, which in turn helps

reduce fertilizer inputs and prevent greenhouse gas emissions. Consequently, increasing seedling cold tolerance is a crucial breeding target for maize cultivation. One strategy to help achieve this goal is the exploration of the existing allelic diversity in the global maize population. We hypothesizes that maize lines developed in northern regions and high-altitudes have adaptive alleles for improved fitness in cold conditions, as they have undergone selection for cold tolerance. We make use of 75 high-quality, long-read genome assemblies from diverse global maize accessions, and perform a genome-wide scan to identify signatures of diversifying selection among these accessions. Further, we will explore the functional consequences of identified variants using an evolution-informed machine learning model. Loci identified through these methods are unbiased with respect to the targets of selection. To identify causal variants for cold adaptation, we will conduct an environment-genotype association analysis using measurements of the environment where the lines were developed. Focusing on loci identified as selected and functionally important will reduce false-positives, improving our power to detect directly actionable alleles. This study aims to provide valuable insights for maize crop improvement and global agricultural sustainability.

Machine Learning Predictions of Obesity: Analysis of Gene–Diet Interactions Using Genome and Epigenome Data

Lee, Yu-Chi[1]; Zheng, Shuai[2]; D. Parnell, Laurence[3]; L. Miller, Eric[4]; M. Ordovas, Jose[3]; Xu, Dong[4]; Lai, Chao-Qiang[2]

[1]USDA ARS, Nutrition and Genomics Laboratory, JM-USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA; 2-Department of Electrical Engineering and Computer Science, Bond Life Sciences Center, University of Missouri, Columbia, M; [2]Department of Electrical and Computer Engineering, Tufts Institute for Artificial Intelligence at Tufts University, Medford, MA, USA;; [3]Nutrition and Genomics Laboratory, JM-USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA;; [4]IMDEA Food Institute, CEI UAM + CSIC, Madrid, Spain; 6-Consortium CIBERObn, Instituto de Salud Carlos III (ISCIII), Madrid, Spain.

Obesity poses a significant risk for numerous chronic conditions. This study explores the roles of genetic, epigenetic, and environmental influences on obesity, and seeks to: 1) Identify which genetic, epigenetic, and dietary factors interact to increase the risk of obesity; 2) Use machine learning to create models that predict obesity risk, leading to personalized prevention and treatment plans. Within the Framingham Heart Study Offspring Cohort, we employed Generalized Multifactor Dimensionality Reduction (GMDR) to examine interactions across 422,793 Single Nucleotide Polymorphisms (SNPs), 415,202 DNA methylation sites, and 397 dietary and lifestyle variables. We developed optimal predictive models that correlate well with actual data, achieving a Pearson correlation coefficient (PCC) of up to 0.65. The most accurate model, with a PCC of 0.65, pinpointed 224 SNPs, 530 epigenetic markers, and 46 dietary factors as significant to obesity risk. These approaches have broader applications for

simulating outcomes of personalized dietary and lifestyle changes in complex diseases, assuming genomic and epigenomic data, along with dietary habits, are available at the outset.

EARLY VIGOR IN RAPESEED DEPENDS OF THE GERMPLASM TYPE AND THE GENETIC DETERMINANTS WERE PARTLY SELECTED DURING GROWTH HABIT SELECTION.

Laurençon, Marianne[1]; Alix, Elise[1]; Baron, Cécile[1]; Guichard, Solenn[1]; Jumel, Stéphane[2]; Moulin, Bernard[2]; Gauthier, Marion[2]; Richard-Molard, Céline[2]; Nesi, Nathalie[2]; Laperche, Anne[2]

[1]IGEPP, INRAE – Institut Agro – Univ Rennes, 35653, Le Rheu, France; [2]Ecosys, INRAE – AgroParisTech – Université Paris Saclay, 9110 Palaiseau, France

Rapeseed cultivated surfaces recently decreased worldwide, mainly due to poor plant establishment conditions. Indeed, seedlings are affected during early growth by various abiotic and biotic stresses, thus impacting yield potential. Improving plant establishment using vigorous varieties is a promising way to overcome these stresses, especially under low-input and agroecological systems. To achieve this, it is necessary to target early vigor, i.e. the ability of the young plant to develop rapidly and uniformly, from emergence to the 5-6 leaf stage under a wide range of environmental conditions. Vigor is an integrative and a complex trait, often approximated by leaf area or total biomass but its underlying processes remain largely unraveled. Moreover, rapeseed has a low genetic diversity, due to its allotetraploid nature, its recent selection about 400 years ago resulting in a differentiation between winter (WOSR) and spring (SOSR) oilseed rape; and strong selection pressure from the 80s that led to genetic bottlenecks. Therefore, the present study aims to decipher the genetic determinism of early vigor in rapeseed. A population of 233 genetically diverse rapeseed lines (B. napus) was phenotyped for early vigor and potentially related traits. It included 141 WOSR, 65 SOSR and 27 Asian genotypes, with high ("++") and low ("00") erucic acid and glucosinolates seed content. Experiments were conducted in a tunnel in Le Rheu (France) in 2022, from sowing to the 5-6 leaf developmental stage. Traits related to the following biological processes were assessed: morphological traits (shoot and root biomass, leaf area) and functional traits (related to carbon and nitrogen production and allocation). Our results showed that WOSR, SOSR and Asian genotypes were clearly discriminated based on their phenotypic profile. Indeed, SOSR and Asian genotypes showed a higher nitrogen quantity in their aerial parts. SOSR genotypes also produced a higher root biomass compared to their WOSR counterparts, enhancing nitrogen absorption, as shown by higher total plant nitrogen quantity. Fifty-four QTL were detected, mainly on the A subgenome (40 out of 54). Some of these QTLs are involved in both early vigor and efficiency traits. An FST scan was also carried out to compare the genetic differentiation between each WOSR and SOSR along the genome. This scan highlighted 15 highly differentiated regions that overlapped with previously identified QTLs. Interestingly, when comparing "00" and "++" genotypes, we observed no impact of modern selection on early vigor. These results provide more knowledge on the phenotypic characterization of the early vigor of the various rapeseed germplasms. In addition, we provide more

information on the genetic control of this highly complex trait, which is controlled by a set of low effects loci and which was partly selected during rapeseed growth habit selection. This study also demonstrates the value of utilizing functional traits, in particular carbon and nitrogen fluxes, for more detailed characterization of the genetic control of complex traits. The next step will be to evaluate the potential of genomic prediction to improve rapeseed early vigor.

GENOME-WIDE ASSOCIATION STUDY AND ITS IMPACT ON THE ACCURACY OF GENOMIC PREDICTION FOR LIVE BODY WEIGHTS IN AUSTRALIAN MERINO SHEEP

Van Le, Sang[1]; Moghaddar, Nasir.[2]; van der Werf, Julius H.[3]
[1]School of Rural & Environmental Science, University of New England, 235, Armidale, Australia

Body weight traits such as birth weight (BW), post-weaning weight (PWW), and adult weight (AW) are economically important traits in sheep. While numerous quantitative trait loci (QTL) have been identified for production traits in other animals over the past decades, few QTL studies have been reported for sheep. This study aimed to elucidate the genetic architecture underlying body weight traits in Australian Merino sheep using medium (50K) and high-density (600K) SNP arrays as well as assessing the accuracy of predicting genetic merit for those traits in both density panels and the potential impact of explicitly incorporating QTL information on prediction accuracy. By conducting a Genome-Wide Association Study with 6890, 4169, and 2963 Merino sheep for BW, PWW and AW, respectively, we identified significant SNP associated with these traits. The single regression analysis revealed 15, 4, and 8 significant SNPs for BW, PWW, and AW, respectively, using the medium-density Chip, and 207, 53, and 53 significant SNPs using the high-density chip. We identified 27 genes located on OAR6 and OAR11 that were associated with body weight of sheep. Genomic predictions based on genomic best linear unbiased prediction demonstrated moderate to high accuracy (ranging from 0.46 to 0.63) for predicting genetic merit in body weight traits. Notably, we observed a 0.01 to 0.05 increase in prediction accuracy when utilizing the high-density panel compared to the medium-density panel. Moreover, incorporating the top two most significant SNPs as separate random effects into the prediction model led to an increase from 0.006 to 0.045 in the prediction accuracy of those traits. However, fitting a larger number of significant SNPs gave generally a decrease in prediction accuracy.

SELECTION SIGNATURE ANALYSES IDENTIFY GENOMIC FOOTPRINTS IN LAO NATIVE GOATS

Van Le, Sang[1]; de las Heras-Saldana, Sara[2]; Alexandri, Panoraia[3]; Olmo, Luisa[1]; Walkden-Brown, Stephen W.[2]; J. van der Werf, Julius H.[3]

[1]School of Rural & Environmental Science, University of New England, 235, Armidale, Australia; [2]AGBU, a joint venture of NSW Department of Primary Industries and University of New England, 351, Armidale, Australia; [3]Extensive

Livestock unit of NSW Department of Primary Industries, 2568, Menangle Australia

Lao native goats (Kangbing-Katjang) are the main goat breeds raised throughout Laos' tropical regions. This goat serves as a crucial component of the smallholder production system due to their economic, nutritional and cultural contributions. Most goats in Laos are managed by smallholder farmers with small herd sizes, typically comprising 2 to 12 goats, resulting from random mating. This breed has a small body size, taking 2 to 2.5 years to reach a weight of 20kg. As of 2023, the Lao goat population exceeds 700 thousand heads, primarily bred for meat consumption and export to Vietnam and China. Goat breed differentiation mainly attributed to adaptation to different breeding objectives and adaptation to different environments. Since there is no established breeding program for Lao goats, this study focuses on identifying selection signatures related to adaptation to local environmental conditions. This study aimed to identify signatures of selection for Lao goats using the fixation index (FST) and runs of homozygosity (ROH). A total of 420 Lao goats from the Phin, Songkhone and Sepon districts in the Savannakhet province were genotyped with Illumina's Goat SNP50 BeadChip comprising 59,727 single nucleotide polymorphisms (SNPs). Principal component and admixture analyses revealed the existence of genetic structure in Lao goats and showed that goats from Sepon belong to a different subpopulation compared to those from Phin and Songkhone. Sepon, being closer to the Vietnam border, facilitates goat trading with Vietnam, and previous studies have reported low genetic differentiation between Lao and Vietnamese goats. Therefore, a pairwise FST comparison was conducted between Sepon and the combined data from Phin and Songkhone goats, alongside separate ROH analyses for each of these two subpopulations. Results from the FST analysis identified 21 candidate regions containing 488 significant SNPs and 199 genes across 12 chromosomes. In the ROH analyses for the group of Phin and Songkhone, we found eleven candidate regions in eight chromosomes containing 268 genes. For Sepon, nine candidate regions in six chromosomes containing 257 genes were identified. In total, 248 genes in seven candidate regions under selection on chromosomes 9, 13, 14, 18 and 19 were jointly identified by both statistical methods. We observed one common genomic region for both ROH in two subpopulations in Laos on chromosome 13 (45.84 - 46.81 Mb). This region harboured seven genes associated with milk protein (LARP4B, PRNP; ZMYND11); sperm traits and tail type in sheep (DIP2C); thermotolerance traits in cattle (SLC23A2); scrapie disease (PRNP); immune response (RASSF2); and meat quality (ZMYND11) in livestock. Furthermore, a Gene Ontology (GO) and KEGG pathway enrichment analyses revealed 9 and 1 significant terms (FDR <0.05), respectively. Those GOs are related to metabolic process of linoleic and arachidonic acids (GO:0043651 and 0019369) as well as lipid oxidation (GO:0034440), which plays a central role in regulating the interaction between innate and adaptive immunity. Overall, our results provide new insight into regions of the genome targeted by selection in Lao native goats, llustrating the relevance of using multiple complementary approaches to identify genomic regions putatively under selection in livestock.

ABSOLUTE MEASURES OF GENETICS SIMILARITY OF POPULATIONS USING SNP MARKERS AND CONSIDERING COMPLEX PEDIGREES

Legarra, Andres[1], Principal Investigator/Group Leader; Bermann, Matias[2]; Mei, Quanshun[3]; Christensen, Ole F[4]

[1]CDCB, 4201 Northview Drive, Bowie, MD 20716, USA; [2]Animal and Dairy Science, University of Georgia, 425 River Rd, Athens, GA 30602, USA; [3]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; [4]Center for Quantitative Genetics and Genomics, Aarhus University, C. F. Møllers Allé 3, Bld. 1130, 8000 Aarhus C, Denmark

There is not a single consensus measure of genetic similarity among populations although frequently used ones are Principal Components, Nei's distances and Fst indices. However, these are not invariant to the set of populations being analyzed. We want to construct, based on commercial SNP chips, non-ambiguous (meaning: not depending on the adding or removing one population to the analysis) measurements of similarity (relationships) across populations and of diversity ("inbreeding") within population. The "genotype" of a population is twice its allele frequency minus one. The relationship of a population to itself (gamma_i,i) is twice the cross-product of these "genotypes" of a population divided by the number of markers, and the relationship between the two populations (gamma_i,j) is the cross-product across the two populations. From here, gamma_i,i minus one describes the Inbreeding (with -1 meaning maximum heterozygosity and 1 meaning minimum heterozygosity), Nei's distances are a simple function (1/8)*(gamma_i,i + gamma_j,j -2*gamma(i,j)) and Fst indices are Nei's distances over (1/2)-(gamma_i,j)/4 . Gamma coefficients can also be expressed as a function of evolutionary branch lengths. In Livestock Genetics individuals trace back to "breeds" defined *circa* 1900-1950, breeds that are (or were) "quite" closed populations mating "quite" at random. Pedigrees are long with 10, 20 generations and individuals are genotyped at recent generations, therefore allele frequencies at the original populations are not available. The estimation of these population relationships when populations is then hard. For a single population, a single equation describes the maximum likelihood estimate. For several populations maybe with crosses we use an approximated pseudo-Expectation Maximization as follows. Include breed origins as individuals ("metafounders") using an initial value of their relationships Gamma. Then create pedigree-based relationships. Then, using these, propagate marker-based relationships backwards to metafounders, so-called H-matrix. The updated relationship across population contains the new estimates of Gamma.

BENCHMARK AND IMPROVEMENT OF GENETIC CELL TYPE MAPPING

Li, Ang[1]; Wray, Naomi R.[2]; Zeng, Jian[1]

[1] University of Queensland, Institute for Molecular Biosciences, Brisbane, Australia; [2]University of Oxford, Department of Psychiatry, Oxford, UK

Genome-Wide Association Studies (GWAS) has achieved great success in the past decade. To understand how genetics influence complex traits or diseases, it is crucial to identify the biological context, such as cell types, in which the genetic variants have an impact. Statistical methods have been proposed to integrate single cell RNA-seq data with the GWAS data to prioritise cell types relevant to the trait. These methods can be broadly classified into computational frameworks that test for SNP-heritability enrichment (e.g., sLDSC), gene-set enrichment (e.g., MAGMA-gene-set), and cell-score enrichment (e.g., scDRS). However, a comprehensive evaluation between these methods has not been done. To achieve a fair comparison, while no gold standard for true/false positives, we set 72 cell type and trait/disease pairs, with one putatively critical and one control cell type per trait, covering major tissues and organs (brain, lung, heart, and blood etc.), as the ground truth, based on general knowledge and empirical evidence from prior studies. We analysed these pairs with 36 publicly available well powered GWAS data and 6 human/murine sc(n)RNA-seq data sets, three of which are atlas-level data sets, consisting of the same cell types with different number of cells across datasets. In this study, we systematically evaluated and compared current methods for identifying critical cell types relevant to a disease/trait inferred by GWAS risk variants. We investigated the impact of both cell and gene level statistics, including 7 different cell type specificity metrics, variable numbers of cells within the cell type, and methods for detecting disease-associated genes. We also assessed the influence of using atlas datasets involving multi-tissue versus single tissue (background of genes and cells), species (e.g. mouse vs. human), biotypes of genes and number of genes considered, and the number/type of other functional annotations (e.g., tissue-specific enhancers annotations) used in the analysis. The results showed that scDRS with disease genes identified from mBAT-combo had the highest power among all methods in comparison for identifying the putatively critical cell types for traits and controlled false positive rate. In sLDSC, for a given cell type specificity statistic, using the continuous value of the statistic as the annotation performed better than using the threshold-based binary value. Expression proportion, differential expression t-statistics and Cepo showed superior performance to other statistics we tested. Incorporating tissue-specific enhancers annotation in sLDSC improved the power of identifying trait-associated cell types within the tissue, but decreased the power when the enhancers annotation was not derived from the target tissue. MAGMA-gene-set showed inflation in false positives, however, decreasing the number of cell type specific genes could help control false positives. To maximise the power, we further introduced a novel Cauchy combination strategy to combine results of sLDSC using different cell type specificity statistics and scDRS using mBAT-combo. This study provides a systematic evaluation of state-of-the-art methods for identifying cell types in which the trait-associated genetic variations have impact. Our findings help enhance our understanding of genetic influences on health and disease and are useful for the development of targeted therapies.

CONSISTENT ESTIMATION OF LOCAL HERITABILITY AND GENETIC COVARIANCE USING HIGH-DEFINITION LIKELIHOOD

Li, Yuying; Pawitan, Yudi; Shen, Xia

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Genetic correlation is a key parameter in the joint genetic model of complex traits, but currently it is usually estimated on a global genomic scale. Information on the local genetic correlation allows for a more detailed understanding of the shared genetic architecture. We have extended the high-definition likelihood (HDL) method to a local version HDL-L and showed that it offers a more consistent and efficient estimation of local genetic variances and covariances compared to the state-of-the-art tool called LAVA. Using HDL-L and LAVA for 30 phenotypes in the UK Biobank, HDL-L discovered 854 significant local genetic correlation estimates spanning 179 loci, compared to 696 significant estimates over 160 loci identified by LAVA. Furthermore, HDL-L's computational speed significantly outperformed LAVA, being at least 20 times faster in simulations. Thus, HDL-L is useful for revealing the detailed genetic landscape that underlies complex human traits.

UNDERSTANDING CELL-TYPE SPECIFIC GENETIC REGULATION OF GENE EXPRESSION IN CATTLE

Li, Houcheng[1]; Zhang, Huicong[2]; Liu, George[3]; Cai, Zexi[4]; Sahana, Goutam[5]; Sun, Huizeng[1]; Jiang, Yu[2]; Sun, Dongxiao[3]; Fang, Lingzhao[4]*

[1]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark;; [2]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service (ARS), U.S;; [3]Institute of Dairy Science, College of Animal Sciences, Zhejiang University, Hangzhou, 310058 China;; [4]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, 712100, China;; [5]Department of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, Key Laboratory of Animal Genetics, Breeding and Reproduction of Ministry of Agriculture and Rural Affairs, National Engineering Laboratory for Animal Breed

Genome-wide association studies (GWAS) have discovered thousands of non-coding genomic loci associated with complex traits of economic and ecological value in farm animals. However, dissecting the molecular mechanisms underlying such trait-associated variants is currently challenging as most of them reside in non-coding region of the genome. Although substantial strides have been made by the Farm animal Genotype-Tissue Expression (FarmGTEx) toward understanding association between traits, genes and bulk tissues, our understanding of the cellular mechanisms underlying GWAS loci is still severely limited, because tissues are mixture of many cell types and states. By leveraging the newly built Cattle Cell Atlas (CattleCA, http://cattlecellatlas.farmgtex.org), representing 1,793,854 single-cell/nuclei from 130 major cell types of 59 tissues/organs of 15 animals, we conducted cell type deconvolution for 11,219

public bulk RNA-seq samples of 33 bovine tissues from the CattleGTEx dataset using DWLS, after carefully benchmarking benchmarked seven popular deconvolution methods, including Cibersort, DWLS, MuSic, SCDC, BisqueRNA, DeconRNASeq and CDSeq. The cell components of common cell types and rare cell types in each tissue were estimated, revealing that the dynamic landscape of cell types and states across bulk RNA-seq samples. For example, intermediate monocytes and club cells were predominant cell types in blood and lung samples respectively. In summary, cell components of the predominant cell type within each tissue generally exhibited greater variability in standard deviation across samples relative to other cell types, suggesting a correlation between cell components and biological conditions in most common cell types. These deconvolution results provide a foundational dataset for identifying cell type interaction variants that influence gene expression and alternative splicing, which will be useful for understanding cellular and molecular mechanisms underlying complex traits in cattle.

GENETIC VARIATION IN A POLYGENIC TRAIT UNDER STABILIZING SELECTION AND POPULATION STRUCTURE

Li, Juan; Hermisson, Joachim; Sachdeva, Himani

Faculty of Mathematics, University of Vienna.

We aim to understand polygenic variation across a subdivided population. We assume that the population consists of many demes connected by migration. In each deme, individuals exhibit a polygenic trait subject to stabilizing selection towards the same optimum. We investigate how genetic variation within demes and across the whole population depends on the rate of mutation, effect sizes, the number of loci, and migration. Through diffusion approximation and simulations, we seek to identify the conditions under which a polygenic trait uses either the same or different genetic variants across different demes.

ONE-DIMENSIONAL ASSOCIATION MAPPING SCAN FOR HETEROTIC QTL LEVERAGING WHOLE-GENOME RESEQUENCING DATA

Li, Guoliang; Zhao, Yusheng; H. Schmidt, Renate; C. Reif, Jochen; Jiang, Yong

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

Abstract: Heterosis, the superior performance of hybrids over their parents, has been exploited systematically in plant breeding and is considered a major asset in meeting the world's food needs. However, its genetic and molecular basis is complex and has been extensively studied. Inspired by a pioneering study on bi-parental populations (Melchinger et al. 2007), a tailored quantitative genetic framework to study the genetic basis of heterosis in diverse hybrid populations has been developed and applied to elucidate the genetic architecture of heterosis for grain yield in wheat (Jiang et al. 2017). In this framework, the heterotic effect of a given locus is defined as a linear combination of its dominance effect and its digenic interaction effects with the entire genetic background. A multi-step approach was designed to test the heterotic effect (Hqtl_MSS). Despite its

successful applications, a clear disadvantage is that it involved a two-dimensional genome-wide association scan for interaction effects between each pair of markers. Because of the large number of pairwise interactions, the two-dimensional scan faces enormous computational challenges and often suffers from low statistical power due to multiple test correction. In particular, the time required for conducting Hqtl_MSS increases quadratically with increasing marker number. Thus, Hqtl_MSS is not feasible for high-density marker panels such as whole genome resequencing (WGS) data. To address this issue, we have developed a novel efficient strategy for mapping heterotic QTL. Instead of a multi-step procedure involving a two-dimensional scan for pairwise interactions, we directly test the heterotic effects in a one-dimensional scan (Hqtl_ODS). It is much more efficient than Hqtl_MSS as the time for computation increases only linearly with increasing marker number. Applying both approaches to a data set consisting of 1,557 hybrids with 57,846 SNPs, we illustrate that Hqtl_ODS is 20 times faster than Hqtl_MSS. Further, Hqtl_ODS is also capable of handling the WGS data with millions of markers.

EXPRESSION AND ALTERNATIVE SPLICING QTL MAPPING REVEALS NOVEL LOCI AND GENES ASSOCIATED WITH DOWNY MILDEW RESISTANCE IN SPINACH

Zhang, Ruifeng; Zhang, Hui; Li, Zhengcao

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, 510006, Guangzhou, China

Downy mildew, caused by the oomycete Peronospora effusa, is the most economically devastating disease on spinach. To better elucidate the genetic basis of complex traits in spinach, such as downy mildew resistance, we perform expression and alternative splicing quantitative trait locus (eQTL and sQTL) mapping analysis based on transcriptome data of leaf tissue from 111 cultivated and wild spinach accessions.We identify 378 novel loci associated with 19 agronomic traits by GWAS, and 29,676 cis- and 407,267 trans-eQTLs, as well as 5,757 cis- and 367,553 trans-sQTLs modulating the expression of 28,721 genes by e/sQTL mapping, among which 203 cis-eQTLs and 44 cis-sQTLs are validated by allele-specific expression (ASE) analysis. Bayesian colocalization analysis suggests that 13 cis-eQTL-GWAS colocalized loci and 9 candidate genes associated with downy mildew resistance in a hotspot region on chromosome 3. Our study constructs the first e/sQTL map for spinach leaf tissue, providing novel insights into the genetic basis of complex traits and genetic materials for breeding downy mildew-resistant cultivars in spinach.

GENETIC ATLAS OF HUMAN MEMBRANE PROTEIN COMPLEXES.

Li, Ting[1]; Klaric, Lucija[2]; F. Wilson, James[3]; Wu, Di[1]; Shen, Xia[2]

[1]State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai, China; [2]Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China; [3]Centre for Global Health

Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

Proteins serve as the building blocks of biological functionality and are encoded in the human genome. Unraveling the genetic underpinnings of the human proteome is critical in gaining insights into the regulation of human complex traits and diseases. Integrating this information with existing genomic data from genome-wide association studies (GWAS) is especially beneficial. While various assays have been developed to measure protein abundance in human plasma samples, they lack the capability to assess the functional interactions between these protein molecules. In this study, utilizing a novel technology, we analyzed plasma from 1,000 individuals and detected and quantified millions of membrane protein complexes (MPCs) in each individual. Our genetic atlas of these MPCs identified 377 MPCs with significant cis-regulatory loci and specific protein-protein combinations. The mapped trans-regulatory loci provide insights into the genetic mechanisms of extracellular vesicles in biological processes, such as cell signalling and neurodegeneration. By integrating with established genome-wide association study (GWAS) findings, we were able to draw inferences regarding the potential causal links between specific membrane protein complexes (MPCs) and complex diseases, such as lung cancer, ALZasthma, ulcerative colitis, inflammatory bowel disease, among others. This discovery positions MPCs as a highly promising area of focus for future medical research.

MASSIVELY PARALLEL REPORTER ASSAYS ENABLE THE IDENTIFICATION OF GENES ASSOCIATED WITH COMPLEX TRAITS

Chen, Andy; Yu, Xuhong; Chu, Xiaona; Lai, Dongbing; Wang, Yue; Edenberg, Howard; Liu, Yunlong

Indiana University School of Medicine

Non-coding regulatory elements such as in the 3’ untranslated regions (3’-UTR) and enhancer regions can regulate gene expression. Variants in these regions are implicated in many diseases, but their mechanisms are less straightforward than those in the coding sequence, which may directly alter structures of proteins. Using a massively parallel reporter assay, we can evaluate the functional effect of thousands of such variants by inserting their sequences into a reporter plasmid and observing the resulting gene expression changes in a transfected cell line. However, despite the high-throughput nature of these assays, the number of possible variants to be evaluated is much larger than could feasibly be performed. To address this, we built a machine learning model to predict the potential outcomes of the MPRAs. Our multi-task model consisted of a convolutional neural network layer (to model motif-like sequences) and a long short-term memory layer (to model interactions between regulatory elements) trained to use the reference and alternative sequences evaluated by the MPRA to predict both sequence activity and variant impact. Using the results of MPRA experiments testing 9,550 3’-UTR variants (918 significant) and 23,122 enhancer variants (4,456 significant), we trained models to predict the impact of novel variants. Using these predictions, we leverage a larger pool of regulatory variants to integrate impact with genome-wide association to identify genes that

contribute to substance use disorders. This approach has uncovered potential new molecular mechanisms of addiction and potential therapeutic targets, showcasing the power of integrating diverse genetic and computational approaches.

EVALUATING METHODS FOR TRACING BREED ORIGIN ALLELES IN CROSSBRED DAIRY CATTLE

Recently, crossbred animals have become integral as parents in subsequent generations of dairy and beef cattle systems in Nordic countries, which has raised interest in routine genomic evaluation of these animals. Given that the effects of marker alleles in crossbred animals can vary based on the breed origin of the alleles (BOA), achieving accurate genomic prediction requires a reliable and efficient method for detecting BOA from reference purebred populations, especially in rotational crossbreeding setups. Therefore, this study aimed to evaluate different software and methodologies for their accuracy and efficiency in tracing the BOA using reference purebred populations. We conducted simulations and evaluated BOA methods with two populations exhibiting different levels of admixture: Nordic Red Cattle (RDC), an admixed breed, and a dairy cross population consisting of Holstein, Jersey, and RDC, similar to those in official Nordic Cattle Genetic Evaluation (NAV). In Nordic countries, RDC population consists of animals with varying proportions of genetic materials from Danish Red (RDM), Swedish Red (SRB), Finnish Ayrshire (FAY). We identified a total of 304 reference purebred animals by selecting those with a pedigree-based breed proportion of RDM, SRB or FAY greater than 0.9 and conducted PCA analysis on these animals, resulting in three distinct clusters. The number of animals was adjusted to match the real RDC breed proportions in the NAV RDC population: 8 RDM (8.5%), 29 SRB (39.6%) and 38 FAY (51.9%). We phased these animals' genotypes using Beagle software and used them as the animals in base population and purebred reference populations. We used ADAM simulation software to randomly mate animals for 10 generations and gradually expand the population to approximately 4000 RDC animals by allowing donor scheme with Multiple Ovulation and Embryo Transfer (MOET), matching the real RDC population's breed proportions and LD structure. Then, we simulated Holstein (HOL), Jersey (JER) and RDC crosses using real HOL and JER haplotypes from DairyCross project, along with simulated RDC haplotypes in the final generation to create crossbred population with the breed proportions similar to NAV crossbred population 0.52 HOL, 0.15 JER and 0.33 RDC. ADAM was updated to enable BOA detections for markers and QTL. Next, our plan is to trace BOA for these two admixed populations using different software such as AllOr and ChromPainter with varying parameter setups and compare them with real BOA from ADAM simulations. Ongoing analyses will contribute to improvement of genomic evaluation in DairyCross NAV routine genomic evaluation, and the development and implementation of a single-step model combined with BOA.

EXPLORING IMPRINTING PHENOMENA: ALPHASIMR'S NEW FUNCTIONALITIES

López-Carbonell, David; Gaynor, R.Chris; Varona, Luis; Gorjanc, Gregor

Facultad de Veterinaria, Instituto Agroalimentario de Aragón (IA2), Universidad de Zaragoza, 50013, Zaragoza, Spain

Imprinting is an epigenetic effect that causes partial or complete allele silencing, depending on the parent of origin of the allele. This epigenetic effect is caused by methylation and histone modifications, and it has been associated with phenotypic variation of several complex traits. For instance, the IGF2 (Insulin Growth Factor 2) in pigs or the Callipyge in sheep have been widely studied. Quantitative genetic model with imprinting effects enables modelling this source of variation and predictions. In animal and plant breeding research, stochastic simulations play a crucial role in comparing breeding strategies and in validating new methods. However, imprinting effects are generally not considered in such simulations, limiting the study of this source of variation. AlphaSimR, is an R package for general stochastic simulations of breeding programmes. Here, we report on the recent addition of imprinting effects to AlphaSimR. This was done following the orthogonal model for diploids that assigns the opposite imprinting genotype effect to maternal and paternal heterozygotes in addition to the standard additive and dominance genotype effects as required. Owing to the flexibility of AlphaSimR, imprinting genotype effects can be parametrised under multiple scenarios and hypotheses. A user can obtain key quantitative genetic statistics, such as imprinting deviations or breeding values for males and females and associated variances. Starting haplotypes and all genotype effects can be supplied by users to kick-start a simulation. These features make AlphaSimR a useful tool for testing imprinting predictive models and breeding hypotheses. Users can now study the impact and implications of imprinting on breeding strategies. Currently, this new functionality is incorporated for diploid genomes, but polyploid extensions are possible. To illustrate the potential application of this software, we have simulated a simple breeding program to evaluate the dynamics of selection and response to selection for an imprinted trait.

MULTIMODAL MACHINE LEARNING-BASED INTEGRATION OF GENOMICS, RADIOMICS AND PATHOMICS PROGNOSTIC SCORES IMPROVES PANCREATIC DUCTAL ADENOCARCINOMA CLINICAL-BASED PATIENT STRATIFICATION

López de Maturana, Evangelina[1]; Sabroso-Lasa, Sergio[2]; Alonso, Lola[1]; Malats, Núria[2]

*[1]Genetic & Molecular Epidemiology Group (GMEG), Spanish National Cancer Research Centre (CNIO) and CIBERONC, Madrid, Spain n; [2]2PANCAIM Project (https://pancaim.eu/) * Equal contributio*

With the development of high-throughput technologies, large amounts of omics data are available, engendering the investigation of cancer biology and patient stratification according to their prognosis. At the same time, the rapid evolution of artificial intelligence (AI) has also supported an increase of its applications in the medical setting. To date, multiple research has been done using molecular/imaging unimodal data, and in a lower extent, multimodal data. However, the integration of molecular omics data with macro/microscopic morphological information, such as radiomics and pathomics data, has not been explored yet. This approach, which potentially can stratify patients according to

their prognosis in a more useful manner, would be very useful for cancers such as pancreatic ductal adenocarcinoma (PDAC), whose patients suffer poor prognosis. Therefore, the aim of this study was to develop, for the first time, a PDAC classification machine-learning-based model integrating multimodal omics data including genomics, pathomics, and radiomics, along with clinical and other tumour variables, exploiting the complementary prognostic information of those layers. The study population comprised 253 PDAC cases belonging to the PanGenEU study with clinical information and at least one modality of omics data. The extraction of genomic features of the tumour was performed using HTSeq, varca and Kraken2 for RNAseq, DNA mutations and tumour microbiome data, respectively. PyRadiomics and the Deep Learning CLAM model were applied for radiomics and pathomics data. Since not every individual had information for every layer, an imputation step was performed with missforest after feature normalization in each modality. Then, omics data were integrated as predictors of overall survival (OS) using a Bayesian multikernel-based regression accounting for censoring, which also involved the computation of different similarity matrices: (co)variance structure matrices for RNAseq, radiomics of the normal pancreas/tumour and pathomics, and cosine and linear kernels for DNA mutations and tumour microbiome. ior means of the unknown parameters were obtained using the BGLR R package, considering a McMC chain of 500,000 iterations, discarding the first 100,000 as burn-in. The application of a backsolving strategy to the prognostic score for each layer allowed obtaining a multimodal ranking of features. All omics layers explained ~56% of the total OS variance, being pathomics the most relevant one (26.1%). The multimodal score outperformed the clinical based-score in the classification of PDAC patients based on their OS. Ranking of features revealed that eleven pathomics features were in the top 30 positions. These features were correlated with the expression of genes previously related to pancreatic cancer and metastasis. The most important feature was the expression of AHNAK2 gene, followed by the KRASG12D mutation. Interestingly, a bacterial species (Pastereulla Multocida) also ranked among the top positions showing a potential prognostic value for OS. Remarkably, our multimodal profiling approach offers an improved PDAC patient stratification based on their prognosis compared to clinical-based one. Additionally, we derived a singular ranking, independent of the omics layer, in which 25/30 top features belonged to radiomics and pathomics modalities. Interestingly, we validated previously reported features as KRASG12D, and some novel ones (P. Multocida).

POPULATION GENOMICS UNRAVELS THE HOLOCENE HISTORY OF BREAD WHEAT AND ITS RELATIVES

Zhao, Xuebo; Guo, Yafei; Kang, Lipeng; Lu, Fei

Institute of Genetics and Developmental Biology, Chinese Academy of Sciences

Deep knowledge of crop biodiversity is essential to improving global food security. Despite bread wheat serving as a keystone crop worldwide, the population history of bread wheat and its relatives, both cultivated and wild, remains elusive. By analyzing whole-genome sequences of 795 wheat

accessions, we found that bread wheat originated from the southwest coast of the Caspian Sea and underwent a slow speciation process, lasting ~3,300 years owing to persistent gene flow from its relatives. Soon after, bread wheat spread across Eurasia and reached Europe, South Asia, and East Asia ~7,000 to ~5,000 years ago, shaping a diversified but occasionally convergent adaptive landscape in novel environments. By contrast, the cultivated relatives of bread wheat experienced a population decline by ~82% over the past ~2,000 years due to the food choice shift of humans. Further biogeographical modeling predicted a continued population shrinking of many bread wheat's relatives in the coming decades because of their vulnerability to the changing climate. These findings will guide future efforts in protecting and utilizing wheat biodiversity to enhance global wheat production.

MAPPING THE RELATIVE ACCURACY OF CROSS-ANCESTRY PREDICTION

Lupi, Alexa S[1]; Vazquez, Ana I[2]; de los Campos, Gustavo[2]

[1]Department of Epidemiology and Biostatistics, Michigan State University (MSU), East Lansing, Michigan 48824, United States.; [2]Institute for Quantitative Health Science and Engineering, Systems Biology, MSU. Department of Statistics and Probability, MSU.

The overwhelming majority of participants in genome-wide association studies (GWAS) have European (EU) ancestry, and polygenic scores (PGS) derived from EUs often have poor predictive performance in other ancestry groups. Previous studies suggest that genome differentiation (i.e., between-ancestry differences in allele frequencies and linkage disequilibrium patterns) is a significant factor contributing to the poor portability of PGS in cross-ancestry prediction. We hypothesize that the portability of (local) PGS varies significantly over the genome due to varying levels of genome differentiation between ancestries. Therefore, we developed a method, MC-ANOVA, to estimate the accuracy loss in cross-ancestry prediction attributable to genome differentiation for short chromosome segments. We applied MC-ANOVA to data from the UK Biobank to develop PGS relative accuracy (RA) maps of EU-derived SNP effects in non-EU ancestries. We report substantial variability in RA along the genome, suggesting that even in ancestries with low overall RA of EU-derived effects (e.g., African), there are regions with high RA. We substantiated our findings using six complex traits, which show that EU-derived effects from regions where MC-ANOVA predicts high portability also have high empirical RA in real PGS. We provide software for MC-ANOVA and maps of the RA for several non-EU ancestries. These maps can be used to interpret similarities and differences in GWAS results between groups and to prioritize variants to improve cross-ancestry prediction.

ESTIMATING THE GENETIC BASIS OF QUANTITATIVE TRAITS ACROSS POPULATIONS

Maksimova, Ekaterina S.; Krätschmer, Ilse; Tkacik, Gasper; Robinson, Matthew R.

Institute of Science and Technology Austria

Genome-wide association studies are radically changing our understanding of the genetic architecture of human health and disease. Still, current research is being disproportionately skewed towards analyses in the European population, which induces health disparities for underrepresented groups, limits drug discovery, and hinders understanding of the underlying genetic mechanisms. However, these issues are not resolved by simply increasing the diversity of participants, while applying conventional existing statistical methodology that is not designed to account for differences in allele frequencies, linkage disequilibrium (LD) patterns, and environmental exposures across populations. Here, we propose a multivariate Bayesian approach for when phenotypes are measured across independent genomic data sets, which models effect sizes jointly both across genome and across populations and explicitly utilizes population-specific allele frequency and LD. Our method is developed to work efficiently on marginal summary statistics estimated from large-scale genotype data for millions of markers, as well as on individual-level genomic data for hundreds of thousands of individuals and markers. The proposed joint modeling of effect sizes allows for precise estimation of the covariance of effect sizes across the genome and testing for whether effects are shared, or specific to a set of populations. In simulation analyses, we demonstrate how modeling the effects allowing for genetic covariance leads to improved fine-mapping and better cross-group polygenic prediction accuracy and enables comparison of the genetic architectures of traits measured across populations. Moreover, we analyze how sensitive our method is to misspecification of LD matrices by comparing its performance across a range of LD matrix approximations. We demonstrate that using an extremely sparse precision matrix inferred using linkage disequilibrium graphical models does not only provide an accurate fine-mapping of causal variants and prediction, but allows for fast multi-ancestry analysis across millions of markers. Overall, we provide empirical evidence that discovery, genomic prediction, and understanding of genetic architecture of traits are greatly improved by analyzing effect sizes across the genome and across populations jointly. Thus, multivariate analysis incorporating multi-ancestry data should be commonplace, improving our ability to infer the shared and population-specific genetic architecture patterns of complex traits in the current human population.

THE EVOLUTIONARY LANDSCAPE OF COMPLEX TRAITS IN EAST ASIA USING ANCIENT DNA

Li, Jielin; Sun, Ningbo; Mao, Xiaowei

Health Management Center, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 611731,China

The changes in environment, lifestyle, and other factors during human evolution led to the emergence of specific phenotypes in modern humans. For modern East Asians, unique environmental and historical factors have shaped distinct traits. The origins of distinct traits in East Asian populations, with the help of time-series genomes (ancient genomes at different historical time points), can be explored from an evolutionary perspective. Utilizing 447 published ancient

genomes of East Asia over the past 10,000 years, we firstly estimated the effective population size throughout the population history. Then we explored the evolutionary characteristics of genes related to typical traits through selection signature detections. After detection, 13 genome-wide significant regions were identified, including immune-related gene FADS2, alcohol dependency-related gene ADH1B, skin color related genes, and genes related to human diseases such as type 2 diabetes. For Tibetan populations specifically, due to limited sample size and temporal distribution, we constructed an admixture model utilizing ancient south and north Chinese population to identify the selection signals. The results showed three significantly selected genomic regions, including the typical high-altitude adaptation gene EPAS1 and genes associated with high-altitude related diseases. Our results reveal the evolutionary characteristics of genes related to some distinct traits in East Asia, providing new insights into the origin of traits and diseases.

SIMILARITY MATRIX FROM MICROBIOTA DATA: IMPACT OF METHODS ON 1-WEEK STABILITY

MARIE-ETANCELIN, Christel; DAVID, Ingrid

INRAE, UMR Genphyse, Université Toulouse, ENVT

High-throughput 'omics data is now available on animals, which allows us to include this new non-genomic information in mixed models to estimate more precisely variance components or to improve genetic predictions. The integration of this omics information requires the construction of similarity matrices between individuals. A large number of metrics are available without knowing which ones are the more suitable. Regardless of the trait of interest, we propose to use the Compsim App (David et al, ICQG 2024) to evaluate the stability of these different similarity matrices build from microbiota data collected one week apart on the same individuals. The ruminal microbiota, obtained by 16S sequencing, from 118 dairy ewes sampled twice, one week apart, formed the database. The ewes were adult animals at a lactation stage of 3-month and fed the same mixed ration. Sequences were analysed using the FROGS pipeline and 3 microbial datasets were produced for the 2 x 118 ewes: Amplicon Sequence variant (ASV) abundances (n=2079), genus abundances (n=117) after taxonomy assignment using SILVA database, and microbial functions abundances (n=251) after functional inference using PiCRUSt2 bioinformatic approach. For each of the 2 weeks, 10 similarity matrices were calculated: 3 distance methods (Bray-Curtis, Jaccard, Euclidean), 4 kernel methods (linear, polynomial, gaussian and arc cosine), 2 ordination methods (Multi-Dimensional Scaling, Detrented Correspondence Analysis) and Poisson log-normal model. The compositional microbiota data were transformed beforehand to apply part of the methods: zero value imputation using Geometric Bayesian Method (GBM) for Bray-Curtis and the 2 ordination methods, centred log-ratio transformation after imputing zero (GBM) for Euclidean distance and the 4 kernel methods. The matrices were compared between the 2 weeks according to the type of similarity matrix and by type of data, using graphical approaches, ranges of variation according to the week and correlations between the off-diagonal elements of the 2 weeks. To

provide a basis for comparison, the milk MIR spectra (1060 wavelengths) recorded a week apart were analysed in the same way. Whatever the method used to build the similarity matrix, the correlations between the 2 weeks varied between -0.02 and +0.47 for the 3 types of microbial data, while for the MIR spectra, the correlations ranged from -0.06 to +0.40. Among the microbiota data, stability between the 2 weeks was much better with ASV (correlations from + 0.12 to +0.47) than with genus or bacterial functions (correlations under +0.09). In detail for ASV, linear and polynomial kernels, Poisson log-normal and Detrented Correspondence Analysis performed better (correlations > 0.31) compare to gaussian Kernel, Bray-Curtis and Jaccard (correlations < 0.19). In contrast, for MIR spectra, Detrented Correspondence Analysis and Bray-Curtis gave the best stability results (correlation > 0.31) while all the 4 kernel methods gave null correlations. These results show that the choice of method and type of data used to construct similarity matrices has a major impact on the analyses that can be carried out with them. This work was carried out as part of the INRAE-INRIA PEPR "Holobiont" program.

UNDERSTANDING THE OPTIMAL GENOTYPING STRATEGY TO ASSESS THE INDIRECT GENETIC EFFECT IN DAIRY CATTLE

Marina, Hector[1]; Hansson, Ida[2]; Fikse, Freddy[3]; Robinson, Matthew[4]; Rönnegård, Lars[5]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, SE-750 07 Uppsala, Sweden.; [2]2 Växa, Swedish University of Agricultural Sciences, Ulls väg 26, SE-756 51 Uppsala, Sweden.; [3]3 Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria.; [4]School of Technology and Business Studies, Dalarna University, SE-791 88 Falun, Sweden.; [5]The Beijer Laboratory for Animal Science, Swedish University of Agricultural Sciences, Box 7024, SE-750 07 Uppsala, Sweden.

Genetic selection to increase the milk yield of dairy cows has been very successful over the last decades. However, the genetics of the animal itself is not the only factor that can influence the phenotype; the effect of the individual genome on the trait values of other individuals, known as the indirect genetic effects (IGEs), can also influence the phenotype of animals. IGEs contribute to heritable variation in livestock species but have been scarcely studied in dairy cattle compared to other livestock species. Understanding IGEs is relevant for predicting response to selection and improving dairy cattle breeding programmes. Social interactions between dairy cattle are an integral part of their daily behaviour, and their disruption can impact their phenotypes and welfare. However, there has been little research into the genetics of social interaction in this species, in part due to the reliance on labour-intensive methods of behavioural observation. The study of IGEs would require a continuously monitored genotyped population using precision livestock farming technologies (PLF) installed on dairy farms to optimise production. However, the lack of resources and the lack of knowledge of the appropriate method to genotype and monitor social interactions are the main constraints behind the scarcity of studies

on dairy cattle. This study aimed to perform a simulation study to assess the optimal genotyping strategy to underlying the indirect genetic effect in dairy cattle. We studied how the number of herds, availability of genomic information and accuracy of monitoring systems could affect the outcomes of studies focused on IGEs. Phenotypes for a conceptional trait (e.g. MY), genomic information and social interactions were simulated for a population of 20000 cows in 100 herds. All females were randomly assigned a sire from 200 unrelated sires. Spatial interactions were generated within each herd. For each female, we simulated a direct and social genetic effect, assuming to follow a normal distribution. The direct and social genetic effects variance was set at 2 and 1, respectively, considering different genetic correlations between both traits (-0.6, 0, and 0.6). The phenotype for each female was estimated as the sum of the direct genetic effect, the social genetic and environmental effect of the conspecifics the animal had contact with and a random residual. The estimated breeding values accuracy was evaluated by varying the number of genotyped individuals, number of herds and accuracy of the monitoring system. The simulated data provided a controlled scenario to understand the optimal approach to explore IGEs in dairy cattle. The precision of the IGE estimates increased with the number of individuals genotyped and the accuracy of the monitoring system, while the inclusion of information from different herds improved the results of the study. This simulation study provided guidance on the genotyping and monitoring strategies required for further implementation of this study on commercial dairy farms, paving the way for its evaluation in dairy breeding programmes.

UNDERSTANDING THE FORMATION OF STEM-CELL NICHES IN PLANTS BY MASSIVELY PARALLEL ANALYSIS OF SINGLE-ORGANISMS

Mechanisms regulating the maintenance and differentiation of the plant stem-cell niches (dubbed as “meristems”) have been defined in numerous experimental systems, such as the model plant Arabidopsis thaliana. However, little is known about how meristems are initially specified from non-meristematic tissues in any plant. This is due to the lack of a clear set of morphogenetic events preceding their appearance, and their inaccessible development in tiny embryos, deep inside layers of tissues. Moreover, it is unknown if a single developmental trajectory guides the formation of a meristem, and whether it involves the acquisition of one particular cell state, or the synchronous activity of multiple cell-types which together form a stable meristematic niche. To overcome these challenges, we have adopted modern microscopy and molecular profiling techniques to perform an unbiased characterization of the emergence of a meristem from the mass of cells developing from the single-celled, bare spore of Marchantia polymorpha. We utilized flow-cytometry and light-sheet imaging to monitor the development of hundreds of individual spores growing in parallel and in complete isolation. We observed that whereas most individuals develop a meristematic niche eventually, the timing of its initiation and the morphogenetic changes preceding it were highly variable. Thus, in order to reconstruct the single or multiple trajectories for meristem specification in a non-synchronous population, we have established a scalable experimental pipeline to derive live-imaging and low input transcriptome profiling of the same individuals Marchantia

plants at different stages of their early development (when each comprises of 1-100 cells). To this end, we have analyzed the transcriptomes of nearly 300 individuals, sampled 6 to 14 days after spore activation. The initial data revealed rich gene expression dynamics, and allowed start delineating alternative trajectories defining the formation of the meristematic niche. We could then reconstruct each trajectory in-silico based on hundreds of instantaneous snapshots derived from a non-synchronous population of developing spores. Our single-organism analysis, of a process that was thus far largely inaccessible, establishes an experimental and analytic framework for an unbiased, quantitative definition of the establishment of the first meristem. This could reflect more broadly on our understanding of how stochastic biological systems converge into the robust formation of stable niches with particular, multipotent characteristics.

EMPIRICAL INSIGHTS INTO TRAINING SET OPTIMIZATION FOR AAFC'S WHEAT BREEDING PROGRAM

We partnered with Agriculture and Agri-Food Canada's (AAFC) Swift Current Research and Development Centre (SCRDC) to implement genomic-assisted breeding tools in their wheat breeding program. This collaboration began with the establishment of a training population using grain yield and protein content data from SCRDC's ongoing breeding efforts. The primary goal was to compare state-of-the-art training set (TRS) optimization techniques, assess the impact of historical and parental data on the TRS, and evaluate the accuracy of both additive and non-additive genomic prediction models. For training set optimization, we used the Mean Coefficient of Determination (CDmean) and the Average Relationship (AvgRel) of the TRS and compared them with random sampling and a breeder-proposed TRS based on plant breeding pedigree information. Results indicated that both optimization methods surpassed random sampling, and the breeder-proposed TRS performed worse than random sampling for both traits. CDmean generally outperformed AvgRel, though there were specific cases where AvgRel showed better results. This variability supports the "no-free-lunch" theorem in statistics, which suggests that no single optimization method is universally best; the optimal method depends on the model and the trait being predicted. Additionally, incorporating parental and historical data had no significant impact on prediction accuracy. Finally, we found that non-additive models outperformed traditional parametric ones for both traits, with Reproducing Kernel Hilbert Space regression outperforming random forest among non-parametric methods. Further tests with more years and different TRS sizes are necessary to identify the most effective approach for AAFC's SCRDC breeding program.

GENOMIC REGIONS INFLUENCING MILKING SPEED IN FLECKVIEH CATTLE

Bucher, E.A.[1]; Me´sza´ros, G.[1]; Gebre, K.T.[2]; Emmerling, R.[3]; So¨lkner, J.[1]

[1]BOKU University, Austria; [2]Mekelle University, Ethiopia; [3]Bavarian State Research Center for Agriculture, Germany

Milking speed is crucial for maintaining the udder health of dairy cows and improving the efficiency of labor. However, research on the genomics of milking speed is limited. The main objective of this study was to determine genomic regions with a potential effect on milking speed in Fleckvieh (dual purpose Simmental) cattle. Genome-wide association studies with the GCTA software were conducted using de-regressed breeding values of bulls as phenotypes. The corresponding genotypes of 10,956 bulls were used, genotyped on the Illumina Bovine SNP50 Bead Chip (Illumina) platform. Fourteen SNP on seven autosomes were significantly associated with milking speed for additive effects and three for dominance effects. Significant regions on BTA4, BTA6 and BTA19 correspond with findings for other dairy cattle breeds. Based on the observation of Fleckvieh breed managers, variation of milking speed in batches of daughters of some bulls is much higher than in daughter groups of other bulls. This difference in within family variation may be caused by transmission of alternative alleles of bulls being heterozygous for a gene affecting milking speed. To check on this, we considered standard deviation of yield deviations in milking speed of half-sib daughters as a new trait and performed GWAS for dominance effects. No signal passed the genome wide Bonferroni threshold while two on BTA6 and BTA11 passed an indicative threshold of $-\log10(p) \geq 5$. These signals did not correspond with any of the significant signals from standard GWAS on de-regressed breeding values. A SNP-heritability of approximately 0.76 indicates a high proportion of the phenotypic variance explained by the studied markers. The key conclusion of this study is that several strong genomic signals were found for milking speed in Fleckvieh cattle and that the strongest of them are supported by similar findings in Brown Swiss and Holstein Friesian cattle. Milking speed is a complex trait whose sub-processes have not yet been elucidated in detail. Hence, it remains a challenge to link the associated regions on the genome with causal genes and their functions.

GENOME-WIDE ASSOCIATION STUDY ON WOOD QUALITY TRAITS IN TWO NORWAY SPRUCE POPULATIONS

Morales, Laura[1]; Nordström, Annica[2]; Hayatgheibi, Haleh[3]; Ranade, Sonali[4]; Niemi, Juha[5]; Holmgren, Johan[1]; Olofsson, Kenneth[2]; Lundqvist, Sven-Olof[3]; Scheepers, Gerhard[5]; Hall, David[5]; Karlsson, Bo[5]; García Gil, Rosario[5]

[1]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden; [2]Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden; [3]Research Institutes of Sweden, Stockholm, Sweden; [4]Forestry Research Institute of Sweden, Sävar, Sweden; [5]Forestry Research Institute of Sweden, Svalöv, Sweden

Norway spruce (Picea abies (L.) Karst.) is one of the most economically and ecologically significant tree species within Sweden and across Europe. The lumber industry depends on high and consistent wood quality from forest stands. Wood composition not only drives the quality and type of end-use products but can also influence tree health. Breeding can improve wood quality, and understanding the genetic mechanisms underlying wood quality traits can aid

selection efficiency in breeding programs. Here, we employed wood quality and genome-wide sequencing data from two Norway spruce populations established by the Forestry Research Institute of Sweden (Skogforsk). The first population is a clonal archive in Sävar, Sweden, comprising 1116 unrelated elite genotypes with provenance from across Fennoscandia. The second population is a progeny trial in Höreda, Sweden, which includes 1373 half-sib families with genetic variation spanning southern and central Sweden and central Europe, from which we have sampled one tree per family. The wood quality traits included wood physical (wood density, radial and tangential fiber diameter, fiber wall thickness and coarseness, microfibril angle, wood stiffness) and compositional (lignin, cellulose, hemicellulose content) properties. The wood quality traits were measured on annual rings across 22 and 29 years in Sävar and Höreda, respectively, with data from the early, transition, and late wood tissues from each year. The diameter at breast height of each tree in the Sävar trial was predicted from remote sensing images. We used exome capture sequencing to genotype the material in this study, resulting in approximately 670K single nucleotide polymorphisms. We will present results from (1) phenotypic analyses to estimate heritability, genetic variation, and trait correlations within and across years within each population, (2) population structure analysis within and across populations, and (3) genome-wide association analysis to characterize the genetic architecture of and loci associated with wood quality within and across populations. These results could aid breeding decisions for improved wood quality in Norway spruce.

EXPLORING TECHNIQUES FOR VARIANT EFFECT PREDICTION

M., Camous; R., Guillaume; A., Torben; Y., Xiaqing

Aarhus University

Background: Variant effect prediction (VEP) can help improve yields in staple crops, but can be quite challenging, especially in regards to non-coding variants. Two limitations worth mentioning are: 1. Statistical associations can be used for variant effect prediction, but they lack spatial resolution. 2. Methods based on comparative genomics or molecular measurements like chromatin accessibility tend to lack interpretability. Motivation: We hope to overcome the above limitation by developing an effective process for predicting the effect of variants induced via mutagen exposure in a study cohort of inbred Brachypodium distachyon plants. If successful we aim to extend and apply this workflow to other grasses, such as rice. Methods: Genetically homozygous BD plants have been exposed to a mutagen which induces G:C>A:T mutations. The exposed plants have been sequenced to identify mutagen induced variants, and phenotyped (seed weight, height, heading days, germinations) across several generations. Different methods of identifying coding and non-coding variant effects have and are being explored. - Protein coding variants ESM: Transformer protein language model that scores likelihood of variant occurrence based on protein sequence context in training data (UniRef90). SIFT: Scale-invariant feature transform is a tool that scores variant likelihood based on observed occurrence in multiple sequence alignment (UniRef90). Depletion Analysis: Our

mutated BD population has been sequenced at two different generations (M2 and M5) after self-crossing. Heterozygous CDS mutations at M2 are inspected at M5 generation to see whether they reverted to wildtype or were maintained. The variants are grouped according to the pathway of their genes. - All variants Plant Deep Sea: CNNmodel for predicting the effect of variants on chromatin region openness. GERP: Genomic Evolutionary Rate Profiling is a tool for scoring variants according to expected number of substitutions based on natural rate of evolution. The prediction power of each scoring method is examined using permutations tests and linear models with the VEP scores as explanatory variables and phenotypes as the response. In the future we hope to develop our own VEP large language models (LLM). Results: ESM, SIFT, PDS and GERP scores demonstrate a clear correlation between the measured phenotypes and scores. In all cases the phenotypes are impacted negatively by the induced variants. A stronger correlation is observed between certain phenotypes, such as plant height and seed weight and VEP scores, such as ESM, and the correlation is modulated by score threshold. Variants in genes belonging to pathways with essential functions show a higher rate of depletion from M2 to M5 generation. Work on making VEP is ongoing and upon successful completion will be ported to a dataset of rice plants to identify potentially beneficial/harmful naturally occurring variants within natural populations.

INCREASED ACCURACY OF GENOMIC PREDICTIONS FOR FRUIT QUALITY TRAITS AND TREE VIGOUR IN MANGO (MANGIFERA INDICA. L) USING PRESELECTED VARIANTS FROM GENOME-WIDE ASSOCIATION STUDIES (GWAS) WITH WHOLE-GENOME SEQUENCE (WGS) DATA.

Munyengwa, Norman[1]; Ortiz-Barrientos, Daniel[2]; Dillon, Natalie[3]; Bally, Ian[4]; Wilkinso, Melanie[1]; Al, Asjad[2]; A. Myburg, Alexander[3]; M. Hardner, Craig[4]

[1]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD, Australia,; [2]Queensland Department of Agriculture and Fisheries, Mareeba, QLD, 4880, Australia; [3]School of Biological Sciences, University of Queensland, Brisbane, Qld, Australia,; [4]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa

Genomic selection (GS) using whole-genome sequence (WGS) data is expected to enhance genetic gains in fruit tree breeding programmes as variation at all causative loci are expected to be included in the training data. However, until now WGS data has not increased GS accuracies compared to high-density marker arrays. Prior research suggests that accuracy of GS using WGS data is enhanced when causative mutations are preselected based on GWAS information. Nevertheless, results to date have been ambiguous, and there is a lack of studies evaluating the impact of different GWAS methods on the accuracy of GS when utilizing GWAS-preselected variants. The aim of this study was to evaluate the hypothesis that, utilizing GWAS to preselect trait-associated markers with WGS loci would improve prediction accuracy of GBLUP models in mango (Mangifera indica. L). WGS and phenotypic data were available for four traits [fruit blush colour (FBC), average fruit weight (AFW), fruit firmness (FF)

and trunk circumference (TC)] from 222 individuals in a gene-pool collection. To evaluate our hypothesis, predictive ability (PA) was evaluated using variants preselected based on multi-locus GWAS (ML-GWAS) and single-locus GWAS (SL-GWAS), and these were compared with the PAs derived from WGS data. We performed GWAS using two ML-GWAS methods [Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) and Fixed and random model Circulating Probability Unification (FarmCPU)] and one SL-GWAS approach [a general linear model (GLM)]. The different marker sets, consisting of variants preselected from GWAS, were utilized for GP with the genomic best linear unbiased prediction. Additionally, we assessed the impact of incorporating significant SNPs, identified by GWAS, as a fixed effect in the GS model on PA. Our analysis revealed a marked increase in PA across the four traits when employing variants preselected from GWAS compared to using all WGS data. Specifically, when all the WGS markers were used for GP, genomic PA for AFW, FBC, TC and FF was 0.67, 0.66, 0.51 and 0.41, respectively. By utilizing preselected variants from GWAS, PAs for FW, FBC, TC and FF reached 0.77, 0.71, 0.56 and 0.45, respectively, depending on the GWAS method used for variant preselection. Notably, markers preselected using ML-GWAS with BLINK achieved a genomic PA of 0.77 and 0.71 for FW and FBC, which was higher than the PA of ML-GWAS FarmCPU (0.68 and 0.67 for FW and FBC) and the GLM (0.67 and 0.62 for FW and FBC). Our results demonstrate the advantages of using the ML-GWAS approach based on BLINK when preselecting markers for GS based on prior biological information. Fitting the most significant marker from GWAS increased genomic PA for FBC by up to 8%, further demonstrating the potential of integrating GWAS to optimize GP, particularly in fruit tree crops such as mango where long breeding cycles hamper genetic gains. Key words: GWAS, genomic selection, tree vigour, mango, preselected variants

OPTIMIZING MULTI-TRAIT IMPROVEMENTS THROUGH STRATEGIC PARENTAL SELECTION IN GENOMIC BREEDING

Musa, A.A.; Reinsch, N.

Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf

Background: Genetic improvement of traits with negative genetic correlations—where enhancing one trait can adversely affect another—presents substantial challenges in breeding programs. Traditional methods like genomic estimated breeding values (GEBV) often lead to rapid allele fixation, reducing genetic variance and compromising long-term sustainability. This study introduces "Index_AG," a refined genomic selection strategy integrating Mendelian covariance to better manage genetic trade-offs under complex genetic architectures such as pleiotropy and linkage disequilibrium (LD). Methods: We simulated a cattle breeding program evaluating three selection strategies: traditional GEBV; an Index that combines GEBV with Mendelian sampling variance; and "Index_AG," which enhances the Index by further incorporating Mendelian covariance into variance calculations for aggregate genotypes. We assumed equal index weights of 1 in deriving the aggregate breeding values and

variance for both traits. Our simulations, assuming known marker effects and genetic maps, covered scenarios affected by pleiotropy and LD to assess genetic gain and variance over 20 generations. Results: Under pleiotropy, GEBV exhibited variable trait gains, with rapid increases followed by declines due to negative correlations. Conversely, both the Index and Index_AG demonstrated more consistent and sustainable improvements across traits. The Index was particularly effective, preserving up to 78% of the initial genetic variance, significantly outperforming Index_AG (which preserved 15%) and GEBV (which preserved none). Under the LD scenario, while GEBV showed the highest initial gains for one trait, it underperformed for the second trait. Index_AG maintained a balanced performance, achieving the best genetic gains for the second trait and the highest aggregate breeding value. Conclusion: Integrating Mendelian covariances through Index_AG significantly advances genomic selection strategies, offering potent solutions for multi-trait genomic selection that enhance genetic improvements and preserve genetic diversity. This study advocates for a shift toward more sophisticated genomic selection methods that incorporate comprehensive genetic information, particularly in contexts involving complex genetic architectures. Future research should focus on empirical validation and evaluating the practical feasibility of these strategies in large-scale breeding programs, considering their computational demands.

ADDITIVE AND DOMINANCE GENETIC VARIANCE ARE SIGNIFICANT PREDICTORS OF EARLY SURVIVAL IN ARTIFICIALLY REARED OSTRICHES

The heritability of survival traits has been characterised as low to very low in most farmed animal species. For fitness traits, non-additive dominance effects can be important, meaning that an individuals' performance can depend on the particular gene combination value of their parents rather than the parent average alone. Early survival could be considered a good indicator of fitness but estimates of dominance variance for survival traits are rare. The Oudtshoorn ostrich research farm in South Africa maintains the only pedigreed ostrich breeding population of its kind worldwide. The objective of this study was to estimate the additive and dominance genetic effects of early mortality using the data recorded in this population. The dataset consisted of 16 239 records for early survival (0 to 44 days of age; 0 = mortality; 1 = survived) recorded on South African Black ostriches from 2001 to 2022. The management of this population includes a pair-wise breeding program which leads to more full sib relationships (mean ~ 15) compared to that expected from most other livestock species, which is expected to be favourable for the estimation of dominance variance. The dominance relationship matrix D was built from the information in the pedigree according to Henderson (1985). Prior to the derivation of D-1, values in D smaller than 1/32 were set to zero to ease computation. The data was analysed in a linear mixed model structure using ASREML V4.2 software. Fixed effects were the year by hatch batch contemporary group ($P < 0.01$) and sex ($P < 0.01$), with hatch weight as a squared covariate ($P < 0.01$). The effect of individual inbreeding was not significant ($P > 0.05$) but retained as a linear covariate. Random effects considered were the additive (σ_A^2) and dominance (σ_D^2) genetic variance components and their corresponding variance ratios,

after first determining that maternal effects were not significant and could be excluded from the analysis. On the fixed level, hatch weight was an important predictor of early survival outcomes. The non-linear trend showed that both light (< 0.8 Kg) and heavy (> 1.2 Kg) chicks were close to 8% more likely to succumb. The additive genetic effect was small, but significant ($\sigma_A^2 = 0.0037 \pm 0.0012$), while the dominance effect was slightly larger, but estimated with a lower level of confidence ($\sigma_D^2 = 0.0060 \pm 0.0024$). The corresponding variance ratios for heritability (h2) was 0.029 ± 0.010 and dominance (d2) was 0.048 ± 0.020. When d2was excluded from the model, the h2 of early survival was estimated at 0.046 ± 0.01, which shows some margin of overestimation of direct genetic effects when non-additive effects are ignored. These results point to relative importance of non-additive effects for early survival in this species, which is important if genetic selection for higher survival rates were to be considered. A more articulate phenotyping protocol and continuous data collection should allow for a more detailed analysis in future.

IMPUTATO: FAST AND ACCURATE REFERENCE-BASED IMPUTATION OF DIPLOIDS AND AUTOPOLYPLOIDS BASED ON SEPARATE HAPLOID MODELS

Nettelblad, Carl; Thor, Filip; Kovalenko, Max

Division of Scientific Computing, Department of Information Technology, Science for Life Laboratory, Uppsala University

Genotype imputation is a cornerstone for turning lower-density genotypes into high quality data, as a step for pre-processing for downstream analysis. Imputation can be applied to SNP array data as well as, more recently, low-pass sequencing data. In earlier work, we've demonstrated the ability to use imputation to reconstruct complete genotypes from pooled testing, with overlapping pools. Specifically, we could demonstrate how a feedback methodology adjusting per-variant, per-sample genotype priors to ensure consistency between imputed genotypes and pooled observations could bring up overall concordance from 94.4 % to 98.4 % in a diploidized wheat dataset (Clouard, Nettelblad, biorxiv doi 10.1101/2023.12.12.571203). While these improvements were impressive, explicit pooling is a bit of a niche case. However, any genotyping experiment of a diploid or autopolyploid sample can be interpreted as a set of pooled observations of the constituent haploids. We are therefore developing a tool, Imputato, that performs reference-based imputation on haploids with allele probabilities, but enforces consistency with the pool (organism) level genotype observations by gradually shifting haploid prior probabilities for variants where the sum of the imputed genotypes is inconsistent with observations. This makes imputation of autopolyploids just as tractable as diploids. We demonstrate our preliminary results on 1000 Genomes (human) data and a public dataset for auto-tetraploid potatoes. The former is to establish a baseline against other imputation algorithms in common use for diploid data, when parts of the dataset are used as study individuals, and parts as a pre-phased reference dataset. Comparing accuracy as well as computational time, we find that Imputato is competitive in some scenarios, even for diploid data. For the potato data, we consider within-population imputation, using only

tetraploid SNP observations and no external pre-phased reference, a scenario for which there are few existing approaches that work well. Finally, we discuss how this approach lends itself to GPU implementation for efficient scaling, and the putative inclusion of other data sources, such as explicitly including short or long reads for linkage data, rather than a purely SNP-centric approach, and how it would be possible to handle allopolyploid species in this type of approach.

GENETICS OF WHEAT CULTIVAR MIXTURES

A new project has been started aiming to estimate the genetic correlation between yield of wheat cultivars grown as pure cultivar crop and in cultivar mixtures. We are working with a data set of 35012 yield plots, consisting of 919 different pure wheat cultivars and 31 different mixtures of wheat cultivars, organised in an alpha-design with 4 replicates across 3 to 6 locations per year over the last 20 years. Each year, two identical cultivar mixture plots were included as references in the experiment to test the reliability of the yearly results. In total, the dataset included 943 mixture plots of 3 to 4 wheat cultivars. All data was provided by TystofteFonden, which is the organisation responsible for testing and approval of new cultivars for the Danish market. Since 1994, different combinations of wheat cultivar mixtures have been used as references in the test trials. We estimated a combination of the additive and non-additive genetic effect along with individual line effect from data using a linear mixed regression model. This model included two random yield levels of the individual cultivars, i.e. yield level for growth as a pure cultivar and yield level for growth in cultivar mixtures. Preliminary analysis of yield data shows a phenotypic correlation of 0.98 between pure cultivar growth and cultivar mixtures. For individual wheat cultivars, the additive mixture effect ranged from -0.05 up to 0.13 ton/ha (-0.5 to 1.3 %). On average, the yield level increased by 0.025 ton/ha (0.25%) when cultivars were grown in mixture plots compared to pure cultivar plots. From the data, we estimate an indication of the genetic gain of wheat cultivars to be 0.075 ton/ha/year for the twenty-year time period. In the future, we will include genotyping data of individual wheat cultivars. By combining genotyping data with yield data, we aim to study the genetic architecture for the mixture effects in wheat. Firstly, we will estimate the genetic correlations between yield of cultivars grown in pure cultivar plots and in cultivar mixtures plots. We hypothesize that the additive genetic yield potential for the individual wheat cultivars differs depending on whether they are grown as cultivar mixtures or as pure cultivar crop. Secondly, we will study the non-additive genetic interactions between wheat cultivars of cultivar mixtures plots. Genetics of yield stability and resilience to diseases will also be considered in future analysis from the project.

NATURE MEETS SCHOOLING: WHY BOYS FALL BEHIND IN GPA DESPITE EQUAL GENETIC POTENTIAL

Berg Ofstad, Sverre; Demange, Perline; Eilertsen, Espen; Eftedal, Nikolai; Gillespie Cheesman, Rosa; Ystrøm, Eivind

Institute of Psychology, University of Oslo, Oslo, Norway

Girls are now outperforming boys in school. In Norway, this is most noticeable in the middle school grade gap: out of a maximum of 6 points, the average GPA is 4.5 for girls, but only 4.1 for boys. As we found that middle school GPA in Norway is highly heritable (h2~.77), we conducted a genetically-informed study of whether the gender gap could be explained by differential gene-environment interaction (G x E) across genders. We linked data on the GPA of Norwegian 10th-grade students and population-wide administrative data with molecular genetic data from the Norwegian Mother, Father and Child (MoBa) study which consists of > 27,000 genotyped parent-child trios. G x E was investigated by regressing GPA on polygenic indexes (PGI) for education. We found that within-family PGI for education predicted GPA similarly for both genders (girls: β = .22, SE = .01, boys: β = .24, SE = .01), but that this association varied more for girls across schools (SD = .06) compared to boys (SD = .03). To address limitations in the PGI approach, we will also extend our study to incorporate an extended family model which includes relations as distant as second cousins, who attend both same and different schools. This approach allows us to expand our sample size to 1.2 million students and analyze over 6 million genetic relationships. Combining PGIs and the extended family by school design will allow us to paint an unprecedented detailed picture of how gender-specific gene-environment interactions produce gender differences in GPA.

Utilizing machine learning in genomic selection of Holstein dairy cattle's gross feed efficiency

Jun Kiat, Edwin Ong[1]; Mooney, Mark[2]; Ferris, Conrad[3]; Rezwan, Faisal[4]; Wang, Hui[4]; Shirali, Masoud[3]

[1]Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, Northern Ireland, UK; [2]Agri-Food and Bioscience Institute, Northern Ireland, UK; [3]Aberystwyth University, Wales, UK; [4]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Northern Ireland, UK

The economic sustainability and viability of most livestock production systems have been limited by feed costs, which represent more than half of the production cost in dairy production systems. Herds that have highly feed-efficient (FE) cattle that can reduce feed costs and maintain dairy output would experience greater profitability. Using genomic selection with precision livestock farming (PLF) can equip producers with the tools to identify cattle possessing the genetic traits associated with high FE and aid construction of a breeding program to improve FE across all cattle within herds. However, cattle FE traits are a complex trait and challenging to measure in genomic selection programs being constrained by the limited number of observations that can be recorded. Utilizing machine learning (ML)-based approaches would be a practical approach to better integrate these large and complex genomic datasets, aiding the identification of genetic markers associated with FE and increasing the accuracy of genomic selection models. This study has applied machine learning algorithms to the genomic selection of Holstein dairy cattle based on gross feed efficiency (GFE) estimated through milk production and feed consumption. SNPs from a

50K Illumina Bovine panel were used as genotype data to predict the GFE in dairy cattle using a mixed model employed in Genomic Best Linear Unbiased Prediction (GBLUP) and various fixed models using ML algorithms, including linear models, support vector machine (SVM), Stochastic Gradient Boosting (SGB) and random forest (RF) methods. 70% of data were randomly chosen for training, with the remaining 30% used for validation and evaluate performance of models. Performance metrics used included the root mean square error (RMSE), mean absolute error (MAE), and correlation of determination value (R2). Fisher's Least Significant difference (LSD) test of accuracy was performed to determine the difference in performance between models. Conducting genomic selection using SVM methods were found to provide the highest R2 value with the lowest MAE and RMSE values, significantly different from other algorithms such as the linear models. The present study demonstrated that genomic data can predict complex traits such as GFE and that random forest outperforms linear models. Subsequent analysis will apply the genetic markers identified to be highly associated with GFE to gene ontology to develop an understanding of associated gene pathways and further explain the genetic effects of identified markers. In conclusion, this study has demonstrated how a data-driven framework to detect, evaluate and select highly feed-efficient Holstein cattle based on genomic profiling can be improved using machine learning approaches.

MACHINE LEARNING COMBINED WITH LOCUS-SPECIFIC DEGREE OF DOMINANCE TRANSFORMATION FOR GENOMIC PREDICTION IN HYBRID MAIZE

Enogieru Osatohanmwen, Bright[1]; Cunha Vieira Júnior, Indalécio[2]; Gholami, Mahmood[3]; Sharifi, Reza[4]; Beissinger, Timothy[4]

[1]Department of Crop Science, Division of Plant Breeding Methodology, University of Goettingen, Von-Siebold-Straße 8, Goettingen, 37075, Germany; [2]KWS SAAT SE & Co. KGaA, Einbeck, Germany; [3]Animal Breeding and Genetics Group, Department of Animal Sciences, University of Gottingen, 37075 Gottingen, Germany; [4]Google X, Mountain View, California, United State

ABSTRACT Introduction: Research on trait genetics significantly impacts genomic model predictability. While classical models like genomic best linear unbiased prediction (GBLUP) dominate plant breeding, Machine Learning (ML) methods are gaining traction for their superior handling of non-linear effects. Despite ML's reputation, there are ongoing efforts aimed at boosting its performance. Strategies include combining parametric and non-parametric techniques, explored by Budhlakoti et al. (2022) and Zhao et al. (2021), and innovative data transformations like those studied by Mathew et al. (2022). Though outcomes vary, these approaches aim for more robust predictive frameworks, with some studies showing notable improvements while others face challenges. Objective: This study assessed two models integrating ML and classical statistical methods, incorporating a novel locus-specific weighted dominance effect transformation matrix proposed by Liu et al. (2022) for genomic prediction in hybrid maize. We compared their performance with three baseline models (ML and classical). Materials and Methods: We employed two

simulated maize populations (one with polygenic traits and the other with oligogenic traits) to evaluate these models across traits exhibiting diverse proportions of dominance variance (0% to 40%) and three levels of heritability (Low = 0.3, Medium = 0.6, and High = 0.8). We utilised the Extreme Gradient Boosting (XGBoost) method for ML and implemented GBLUP for classical methods. Models: In this study, we employed five models: (1) XGBoost only (ML), (2) XGBoost combined with locus-specific weighted dominance transformation (ML_Transformed), (3) GBLUP model with additive effects only (GBLUP_Additive), (4) GBLUP model including both additive and dominance effects (GBLUP_Add_dom), and (5) GBLUP combined with locus-specific degree of dominance transformation (GBLUP_Transformed). For the ML_Transformed model, the genomic Single Nucleotide polymorphism (SNP) marker matrix is converted to a transformed matrix using a formula first proposed by Liu et al. (2022), and the new matrix is used in the ML pipeline that involves Bayesian hyperparameter tuning on the training set to get best parameters. Finally, the best hyperparameter is used to train a model of all data. Results: The results show that the ML Transformed model did not perform better than the regular ML model, showing that the transformation with the dominance effects did not improve ML methods. While ML models performed better than base classical models, the GBLUP Transformed model surpassed all others, including ML models across the three levels of heritability, demonstrating the superiority of genomic prediction with transformed GBLUP. Conclusions: The GBLUP_Transformed model consistently outperforms all others for traits with a high dominance variance, whereas the ML_Transformed model did not exhibit the same level of performance, suggesting that the transformed matrix may not be well-suited for ML. These findings highlight the potential of the classical transformed model for hybrid prediction and improvement in traits characterised by a high dominance variance.

MACHINE LEARNING COMBINED WITH LOCUS-SPECIFIC DEGREE OF DOMINANCE TRANSFORMATION FOR GENOMIC PREDICTION IN HYBRID MAIZE

Osatohanmwen, Bright Enogieru[1]; Vieira Júnior, Indalécio Cunha[2]; Gholami, Mahmood[3]; Sharifi, Reza[4]; Beissinger, Timothy[4]

[1]Department of Crop Science, Division of Plant Breeding Methodology, University of Goettingen, Von-Siebold-Straße 8, Goettingen, 37075, Germany; [2]KWS SAAT SE & Co. KGaA, Einbeck, Germany; [3]Animal Breeding and Genetics Group, Department of Animal Sciences, University of Gottingen, 37075 Gottingen, Germany; [4]Google X, Mountain View, California, United States

ABSTRACT Introduction: Research on trait genetics significantly impacts genomic model predictability. While classical models like genomic best linear unbiased prediction (GBLUP) dominate plant breeding, Machine Learning (ML) methods are gaining traction for their superior handling of non-linear effects. Despite ML's reputation, there are ongoing efforts aimed at boosting its performance. Strategies include combining parametric and non-parametric techniques, explored by Budhlakoti et al. (2022) and Zhao et al. (2021), and innovative data transformations like those studied by Mathew et al. (2022).

Though outcomes vary, these approaches aim for more robust predictive frameworks, with some studies showing notable improvements while others face challenges. Objective: This study assessed two models integrating ML and classical statistical methods, incorporating a novel locus-specific weighted dominance effect transformation matrix proposed by Liu et al. (2022) for genomic prediction in hybrid maize. We compared their performance with three baseline models (ML and classical). Materials and Methods: We employed two simulated maize populations (one with polygenic traits and the other with oligogenic traits) to evaluate these models across traits exhibiting diverse proportions of dominance variance (0% to 40%) and three levels of heritability (Low = 0.3, Medium = 0.6, and High = 0.8). We utilised the Extreme Gradient Boosting (XGBoost) method for ML and implemented GBLUP for classical methods. Models: In this study, we employed five models: (1) XGBoost only (ML), (2) XGBoost combined with locus-specific weighted dominance transformation (ML_Transformed), (3) GBLUP model with additive effects only (GBLUP_Additive), (4) GBLUP model including both additive and dominance effects (GBLUP_Add_dom), and (5) GBLUP combined with locus-specific degree of dominance transformation (GBLUP_Transformed). For the ML_Transformed model, the genomic Single Nucleotide polymorphism (SNP) marker matrix is converted to a transformed matrix using a formula first proposed by Liu et al. (2022), and the new matrix is used in the ML pipeline that involves Bayesian hyperparameter tuning on the training set to get best parameters. Finally, the best hyperparameter is used to train a model of all data. Results: The results show that the ML Transformed model did not perform better than the regular ML model, showing that the transformation with the dominance effects did not improve ML methods. While ML models performed better than base classical models, the GBLUP Transformed model surpassed all others, including ML models across the three levels of heritability, demonstrating the superiority of genomic prediction with transformed GBLUP. Conclusions: The GBLUP_Transformed model consistently outperforms all others for traits with a high dominance variance, whereas the ML_Transformed model did not exhibit the same level of performance, suggesting that the transformed matrix may not be well-suited for ML. These findings highlight the potential of the classical transformed model for hybrid prediction and improvement in traits characterised by a high dominance variance.

GENETIC ANALYSIS OF LAYING HEN MOVEMENT BASED ON COMPUTER VISION DATA

Osorio-Gallardo, T.[1]; van Putten, A.[2]; Janssen, D.[3]; Giersberg, M.F.[1]; Rodenburg, B.[2]; Bijma, P.[3]

[1]Animal breeding and Genomics, Wageningen University and Research, Wageningen 6708 PB, The Netherlands;; [2]Population Health Sciences, Veterinary Medicine, Utrecht University, Utrecht 3584 CM, The Netherlands;; [3]Hendrix Genetics Research, Technology and Services B.V, Boxmeer 5831 CK, The Netherlands

The genetic analysis of animal behaviour can be challenging because human observation is required to obtain phenotypic records. Video recordings can be

used for constant observation of a population, and computer vision algorithms can be used to track individuals and identify their behaviours. This research is part of the IMAGEN project, which aims to study the genetic background of behaviour of laying hens in semi-industrial housing. The present study evaluates the suitability of using ArUco- markers for tacking identified individuals to generate data for genetic analysis. For this, we analysed the phenotypic variance of individuals' distance moved, movement speed, and time detected. Videos of seven days of eight hours each, from 12 pens with 133 Dekalb White hens each were used. All hens in each pen were identified by an unique ArUco tag. Spatial coordinates were obtained for every ArUco code detected in every frame of each video, and Euclidean distances between consecutive frames were calculated. The total distance moved per hour of every individual was calculated by aggregating the distances between frames for every hour. In addition, the total minutes that each individual was detected in each hour was also recorded. By using distance moved and total minutes detected, the moving speed per hour per individual was calculated. Univariate linear mixed models were fitted for each trait, using pen-day-hour as a fixed effect, and a random permanent animal effect. For "distance moved", 30% of variance was attributable to the individual, for "movement speed", 25% of the variance was attributable to the individual, and for "minutes detected", 27% of the variance was attributable to the individual. These results show clear inter-individual differences in movement traits among individuals, and the fractions of variance found present an upper bound for heritability. These results also confirm that ArUco-markers are suitable for computer-vision based tracking of identified individuals in field conditions, and allow to identify individual effects. Genotypes will be available around mid-May of this year, and we will show the results of the genetic analyses of the traits at the conference.

GENOME-WIDE INTERACTION STUDY BETWEEN PM10 AND THYROID STIMULATING HORMONE AMONG KOREANS REVEALS FUNCTIONAL POLYMORPHISMS

Park, Young Jun[1]; Kim, Juhyun[2]; Son, Ho-Young[3]; Kim, Hyun-Jin[4]

[1]Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Republic of Korea,; [2]Department of Translational Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea,; [3]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea,; [4]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Republic of Korea, 5.National Cancer Control Institute, National Cancer Center, Goyang-Si, Gyeonggi-Do, Republic of Korea

Background and Aim: Although associations regarding thyroid function and air pollution have been presented utilizing real-world data, a genome-wide interaction study(GWIS) linking thyroid stimulating hormone and particulate matter has not been conducted on a large-scale to this point. Methods: A cross-sectional investigation aimed at assessing the link between genetic variations and various health indicators was performed based on participants sourced from

the Health Promotion Center and Healthcare System Gangnam Center at Seoul National University Hospital, spanning from December 2009 to December 2013. As a result, a total of 1,863 adult males were included in the discovery genetic analysis and 1,490 individuals additionally recruited between 2014 and 2015 were enrolled for the replication analysis. PM10 was either perceived as a continuous variable or a binary variable with a 75-percentile cut-off value of 53.5 μg/m3. Results: For PM10 as a continuous variable, rs6998646 near the PCAT1 gene and two SNPs with high linkage disequilibrium with each other(rs79013868 and rs9597585, r2 = 0.75), in the intron of STARD13 were statistically significant in both the discovery and replication set. For PM10 as a binary variable, a total of seven SNPs, of which each is a peak SNP in the Manhattan plot locus, were identified in the discovery cohort but not replicated in the replication cohort. Among the seven SNPs, two SNPs(rs11781213 and rs7837316) were located near and in the MSRA gene. rs73563822 located near GOT2, rs10109092 near PRAG1, rs7169081 located near GCOM1 were also derived from statistical analyses. Interestingly, rs7169081 located within the enhancer region of GCOM1 acted as its possible regulator, as shown from the Activity-By-Contact model and sc-ATAC-seq mainly in atrial cardiomyocytes. Further, increased GCOM1 expression was mainly found in cardiomyocytes according to the Gene Tissue Expression and Human Protein Atlas database. Conclusion: GWIS of PM10 and TSH revealed various single nucleotide polymorphisms that were revealed to be functionally related to (MSRA) or near a functional gene(GCOM1) with potential interactions.

QUANTITATIVE ANALYSIS OF GENETIC DOMINANCE IN THE PURA RAZA ESPAñOLA HORSE

Perdomo-González, D.I.[1]; Sánchez, J.P.[2]; Molina, A.[2]; Valera, M.[2]

[1] Universidad de Sevilla, Dpto. Agronomía. ETSIA, Ctra Utrera Km, 41005, Sevilla, España (Corresponding Author); [2]IRTA Torre Marimon, C-59 Km 12, 08140, Barcelona, España; Universidad de Córdoba, Dpto. Genética. Ctra Madrid-Córdoba Km 396, 14071, Córdoba, España

Genetic dominance is a mode of gene action involving interaction between alleles of the same locus and it has been largely ignored in animal breeding due to its complexity and lack of accuracy of pedigree-based models. In this work, variance components due to dominance effects were analyzed for two reproductive traits, age at first foaling (AFF) and reproductive efficiency (RE) and one morphological trait, scapula-ischial length (SIL) in Pura Raza Española (PRE) mares. The PRE is an autochthonous Spanish horse with a census of more than 250,000 individuals and a deep and large pedigree since its stud book creation in 1912 with a high pedigree completeness and more than 40 years of proven parental information. Mares with full sisters and at least one foaling were identified from the complete PRE pedigree. Additional records of other females foaling in the same stud and year were added. The number of animals with phenotypic records was 3,967 (AFF and RE) and 3,092 (SIL), and the relationship matrix was formed by 11,331 individuals with a mean inbreeding coefficient of 4.5% and an average relatedness coefficient of 5.1%. First,

condensed probabilities of identity based on pedigree data were estimated using a recursive algorithm. Then an equivalent linear model in which variance components (dominance, additive and their covariance) can be estimated, from pedigree and phenotypic data, using closed-form algorithms was applied. The model included the country, birth years grouped in five-year intervals, the size of the herd, and the inbreeding as systematic factors, with year and herd considered as random effects. The dominant variance was partitioned into a term linked to relationships between non-inbred individuals and another linked to relationships between inbred individuals, with the latter assumed to be correlated with the additive effect. The broad-sense and narrow-sense heritability estimates, referring to the base population, were 0.29 and 0.16, 0.19 and 0.12, and 0.73 and 0.27 for AFF, RE, and SIL, respectively. A comparison among them highlights the relevance of dominance effects in PRE. In a theoretical population with an average inbreeding of 0.2, the estimates of heritabilities in the narrow-sense increase slightly, as a result of both the variance of dominant deviations due to relationships between inbreds and the additive variance increasing; the total genetic variance does not increase since the estimates of covariance between additive effects and dominant deviations are negative. Furthermore, estimates of the additive-dominance correlation are -0.81, -0.77 and -0.55 for AFF, RE, and SIL, respectively. Dominance deviations are modes of inheritance that could be relevant in the PRE breeding program and can potentially lead to a bias in the estimation of the total genetic variance and an overestimation of the estimate of the response to selection if they are not included in the models. In addition, the consideration of these type of dominance deviation models could be a tool for mating design aiming to maximize the overall genetic value of newborn animals.

DYNAMIC INCLUSION OF FUNCTIONAL GENOME ANNOTATIONS TO IMPROVE ACCURACY OF GENOMIC PREDICTION IN PIGS

Genomic prediction is integral to contemporary livestock breeding, particularly employing the Genomic Best Linear Unbiased Prediction (GBLUP) model. However, this model assumes equal importance for all Single Nucleotide Polymorphisms (SNPs), irrespective of their genomic positions and functions. In this study, we aimed to augment prediction accuracies by introducing high-quality functional genome annotation maps. Our customized pipeline, based on Functional-And-Evolutionary Trait Heritability (FAETH) scores, dynamically assessed the relevance of variants in the whole-genome sequence (WGS) of Sus Scrofa. Out of 15 million SNPs in the WGS dataset, 6.2 million had non-neutral FAETH scores. Thirty-two functional annotation maps were incorporated, encompassing various aspects such as expression associations, chromatin accessibility, differential methylation/expression, selection signatures, and variant annotations in the pig genome. Six traits related to reproduction, production, and health were considered, of which heritability varied from 0.10 to 0.38. A forward validation approach, utilizing animals born after July 2020, was employed. The reference datasets ranged from 7,000 to 22,000 observations, with approximately 1,000 animals in the validation set per trait. The results suggest the relevance of specific annotation maps, with lncQTL,

sQTL, and eQTL demonstrating high informativeness across the traits. Our study suggests that FAETH scores effectively rank SNP variants informative across traits, however, the impact varied across traits analyzed. Our approach is dynamic on the process of updating FAETH scores when new annotation maps emerge as well as accumulation of data on the reference population. Incorporating high FAETH ranking variants into commercial SNP chips holds potential for enhancing predictive accuracy in genomic selection for pig breeding.

ESTIMATION OF COVARIANCE MATRICES WITH COMPLETELY MISSING INFORMATION FOR ONE COMPONENT

Poulsen, Bjarke G.

Quantitative Genetics and Genomic, Aarhus University

The estimation of covariance matrices is a key step in many genomic analyses. However, some environmental covariances cannot be estimated between traits such as sex-specific traits where no individual can have both male-specific measurements and female-specific measurements. Traditionally, these un-estimable covariances are assumed to be zero. However, this assumption can cause the covariance matrix to not be positive definite even though the underlying covariance matrix is. Therefore, the aim of this analysis was to describe the viability of an alternative approach, where the un-estimable covariance is updated depending on the other covariances rather than assumed to be zero. The new approach was validated based on 100 simulated replicates with 500 males, 500 females, and three traits each. To create a replicate, I first sampled a positive-definite 3x3 correlation matrix which would not be positive-definite if the covariance between traits 1 and 2 was assumed to be zero. Then I sampled 3 residuals for each individual (one for each trait) while assuming that residuals across individuals were independent and means for all traits were zero. Next, I defined three models: 1) STANDARD where missing information on covariance between residuals is handled as standard in the DMU software, 2) TRUE where the missing covariance was fixed at its true/simulated value, and UPDATE where the covariance with missing information initially was fixed at zero, but also updated to the value that gives the largest determinant for the residual covariance matrix when DMU stopped iterating. Afterwards, I defined two datasets. In the first dataset (FULL), all individuals had measurements on all three traits, while, in the second dataset (MISSING), only males had measurements for the first trait, only females had measurements for the second trait, and all individuals had measurements for the third trait. Lastly, I estimated the residual covariance matrix for all combinations of models and datasets using the standard convergence parameters and method in DMUAI. The estimations of the residual covariance matrix converged for all models and replicates when the dataset was FULL. However, when the dataset was MISSING, STANDARD converged in 5 of the replicates, TRUE converged in 96 of the replicates, and UPDATE converged in all 100 of the replicates. For the FULL dataset, the models had the same mean square error for the estimable residual covariances and variances (0.001-0.002); however, for the MISSING dataset and only for converged combinations of models and replicates, the mean square error was

0.008-1.030 for STANDARD, 0.002-0.003 for TRUE and 0.002-0.003 for UPDATE. Thereby, this analysis shows that the UPDATE approach is stable in the general scenario where all individuals have all measurements, while providing benefit to the scenario where some covariances are un-estimable. Furthermore, the use of the UPDATE approach may enable estimation of genomic models with missing information for some covariances. Lastly, the UPDATE is easily extendable to a scenario with multiple un-estimable covariances.

EXPLORING THE APPLICATION OF PHENOMIC SELECTION IN CORN BREEDING

Phenomic Selection (PS) has recently been proposed as a cost-effective method for predicting complex traits and enhancing genetic gain in breeding programs. This new technique maintains the statistical procedure used in GS-based prediction models but replaces the molecular markers data (e.g., SNP data) with variables obtained from a multi-variate phenotyping method (e.g. , near-infrared spectroscopy (NIRS) data). This study aimed to explore the application of PS using single kernel NIR in a sweet corn breeding program. Here, we focused on predicting field-based traits of economic importance, including ear traits and plant traits. First, on a diversity panel, three models were employed: G-BLUP and NIRS BLUP models, which utilized relationship matrices based on SNP and NIRS data, respectively, and a third model that uses both matrices as independent terms. The genomic relationship matrices were evaluated when the number of SNPs used to build the matrix varied from 500 to 200,000 SNPs. In a second approach, we utilized the NIRS BLUP model trained on the diversity panel to select doubled haploid (DH) lines for germination before planting. Our findings reveal that PS generated good predictive ability (e.g., 0.46 for plant height.). Also, it effectively distinguished between high and low germination rates in DH lines. This highlights the potential of NIRS to enable the selection of DH candidates before planting. Although GS generally outperformed PS, the model combining both information (PS+GS) yielded the highest predictive ability. Furthermore, accuracies of the PS+GS model were considerably higher than GS when low marker densities were used. This indicates NIRS's potential to maintain/improve accuracy together with SNP-based information while reducing marker density, which could decrease genotyping costs in the breeding program. In conclusion, PS is a promising low-cost tool that could help to optimize breeding programs.

PHENOTYPIC PATTERNS OF POLYGENIC ADAPTATION IN LARGE AND SMALL POPULATIONS OF DROSOPHILA

Ramirez-Lanzas, Claudia[1]; Bargh, Neda[2]

[1](1)Institut Für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria.; [2](2)Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria.

Upon environmental changes, complex traits undergo adaptive changes with the contribution of many loci until the new optimal phenotype is reached. However, the patterns of genomic and phenotypic changes in evolving populations are not

fully understood. Experimental evolution (EE) was conducted to study the phenotypic changes of replicated populations undergoing polygenic adaptation in response to an environmental change. We aimed to empirically determine when populations reached equilibrium by quantifying the phenotypic plateau, and to compare the effect of census size on population dynamics. We investigated the phenotypic changes in experimental Drosophila simulans populations of 100,000 and 800 individuals, hereafter referred to as large and small populations, adapting to a new protein-rich diet. Time-series transcriptomic analysis (with three time points spanning 31 generations) showed a fast and highly parallel response in replicate large populations. Interestingly, most differentially expressed genes in the early stage of adaptation plateaued later on, suggesting that populations approached optimal phenotype after a few generations. In small populations, however, the patterns of gene expression exhibited a slower and more heterogeneous response with the majority of genes responding at later generations. Additionally, time-series analysis of fecundity as a fitness component showed similar dynamics to that of the gene expression. These findings corroborate theoretical predictions that adaptation in large populations is faster. The empirical patterns of phenotypic changes as populations reach the new trait optimum after an environmental change could facilitate the identification of polygenic adaptation in two main phenotypic phases, as predicted in theoretical studies: an early phase of directional selection, followed by stabilizing selection phase where the optimal phenotypes plateau.

DISSECTING THE GENETIC AND PROTEOMIC RISK FACTORS FOR DELIRIUM

Introduction: Delirium is a complex neurocognitive condition, affecting nearly 25% of hospitalised older adults. It is characterised by an acute, but usually reversible, deterioration of the patient's cognitive ability, attention and awareness. Multiple adverse outcomes have been strongly associated with delirium, including increased mortality, prolonged hospitalisation and accelerated dementia onset. Despite its high healthcare burden, however, the current understanding of the genetic and biological mechanisms underlying delirium's pathophysiology is still limited, hindering efforts to effectively predict, prevent and treat the condition. In this work we aim to address this gap, by shedding light into the genetic and proteomic factors shaping delirium pathophysiology. Results: Here, we present preliminary results from the largest to-date genome wide association meta-analysis conducted on delirium risk (n=751,972; 10,215 cases). Contributing genome-wide association studies (GWAS) were sourced from European (UK Biobank, n=392,273; 7,176 cases; Michigan Genomic Initiative: n=44,654; 160 cases) and Finnish (FinnGen: n=388,560; 3,371 cases) populations. The ε4 haplotype in the Apolipoprotein E protein (APOE gene) stood out as a major risk factor for delirium, significantly replicating in the "All of Us" Research Programme cohort (European sub-cohort: n= 120,476; 691 cases). However, the association did not significantly replicate in the African and native American populations in "All of Us". Furthermore, a proteome-wide association analysis on 2,914 Olink plasma proteins in UK Biobank (n=32,652; 541 cases) uncovered 161 immune response enriched

proteins significantly associated with incident delirium up to a 15-year follow-up. A large overlap with proteins previously implicated in dementias was also observed. Implications: Taken together, genomic and proteomic results suggest a shared aetiology between delirium and dementias, contributing to the efforts to better understand delirium's complex biological origin. Additionally, identified "omics" risk factors could pave the wave to advancing future endeavours for the identification of effective therapeutic targets for delirium, such as through drug repurposing. Finally, the lack of transferability of the APOE ε4 risk haplotype in non-European populations could suggest ancestry-specific genetic effects, highlighting the importance of including ancestrally diverse populations in genomic studies of complex diseases.

HETEROGENEOUS GENETIC COVARIANCES IN GENOMEWIDE PREDICTION IN MAIZE

Rebollo, Inés[1]; Bernardo, Rex[2]

[1]Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 41Borlaug Hall, 199Upper Buford Cir., St. Paul, MN, 55108; [2]Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Cir., St. Paul, MN, 55108

Models commonly used for genomewide prediction assume a single genetic variance for all individuals, even when the individuals belong to different populations that may vary in their genetic covariances. Our objective was to determine if predictive ability in structured populations is higher with heterogeneous genetic covariances than with a homogenous genetic covariance. We analyzed data from four biparental maize populations, each with 134 to 242 individuals testcrossed to one or two testers, that were phenotyped in 18 environments and genotyped with 1260 single nucleotide polymorphism markers. We studied the traits grain yield, grain moisture, test weight, plant height, and ear height. By genomic best linear unbiased prediction, we assessed predictive ability for each trait in each biparental population with all the remaining biparental populations serving as the training population. We calculated genomewide marker effects in each biparental population by ridge regression-best linear unbiased prediction, then estimated the genetic covariance between each pair of populations from the correlation between their genomewide marker effects. We also estimated the heterogeneous covariances, simulating prediction prior to phenotyping any of the individuals in the test population. Compared with the homogeneous-variance model, the predictive ability with a heterogeneous covariance structure across eight population x tester combinations was higher for grain moisture, test weight, and plant height. We will confirm these results among virtual populations simulated from genomewide marker effects.

ON THE RELATIONSHIP BETWEEN GENOMIC KINSHIP AND OPPOSING HOMOZYGOTES IN A SELECTED POPULATION OF AUSTRALIAN ANGUS CATTLE

Reverter, Antonio; Samaraweera, Malshani; Alexandre, Pâmela A.; Duff, Christian; Porto-Neto, Laercio

CSIRO Agriculture and Food, 306 Carmody Rd., St. Lucia, Qld 4067, Australia. 2 Angus Australia, 86 Glen Innes Road, Armidale, NSW 2350, Australia.

Using SNP genotypes to check for Mendelian inconsistencies allows the identification of animals for which pedigree and genotype information are not in agreement. We sourced from the Angus Australia database a selected population of 10,399 genotyped progeny born from 2013 to 2023 from genotyped sires with at least 100 progeny and genotyped dams with at least 10 progeny. Imputed genotypes were available for 61,105 autosomal SNPs. The genomic relationship (GR) between parent-offspring (N = 21,307 pairs), full-sibs (N = 35,486), half-sibs (N = 677,421), grandparent-grandoffspring (N = 16,308) and unrelated (N = 62,232,954 pairs) was compared against the number of opposing homozygotes (OpH). Theoretical expectations for means and variances were compared against empirical observations. Consistent with expectations, the variance of GR among full-sib pairs was higher than the variance among half-sibs, and the number of OpH among full-sibs was half the number of OpH among half-sibs. Expected to be 0.5, the observed GR among full-sib pairs and parent-offspring pairs was 0.483 (SD = 0.054) and 0.488 (SD = 0.037), respectively. Expected to be zero, the number of OpH among parent-offspring averaged 11.6 and showed an exponential decay with 77.5% of all parent-offspring pairs having an OpH ≤ 12. Among full-sib pairs, the observed OpH averaged 1,162.45 (expected = 1,150.17) and this average was surpassed by only 14 parent-offspring pairs (or 0.07%) and attributed to pedigree errors. Crucially, the anticipated negative correlation between GR and OpH was clearly apparent. However, this correlation was affected by the degree of pedigree relationship being strongest negative among unrelated pairs (r = -0.762), followed by grandparent-grandoffspring (r = -0.740), half-sibs (r = -0.721), full-sibs (r = -0.665), and parent-offspring (r = -0.299). For 3,540 (or ~10%) full-sib pairs with a GR > 0.55, the correlation between GR and OpH was closer to zero at r = -0.276; while for 17,586 (or ~82%) parent-offspring pairs with 0.45 < GR ≤ 0.55 this correlation was merely r = -0.028 anticipating the independence between parent-offspring GR and the number of OpH. We conclude that, in our selected population of Australian Angus cattle, both the mean and variance of GR are close to expectations, while Mendelian inconsistencies are rare and likely attributed to errors in pedigree recording, genotypes and genotype imputation.

INTEGRATION OF GENETIC INFORMATION FROM EQTL MAPPING FOR GENE REGULATORY NETWORK RECONSTRUCTION

Riccucci, Ettore[1]; Mbebi, Alain[2]; Razaghi-Moghadam, Zahra[3]; Tong, Hao[4]; Caproni, Leonardo[2]; Royles, Jessica[3]; Burnett, Angie[4]; Aguilera Miranda, Mariela Paz[2]; Kromdijk, Johannes[3]; Dell'Acqua, Matteo[4]; Nikoloski, Zoran[4]

[1]Institute of Plant Sciences, Scuola Superiore Sant'Anna, 56127 Pisa, Italy;; [2]Systems Biology and Mathematical Modeling Group, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany;; [3]Bioinformatics Department, Institute of Biochemistry and Biology, University of Potsdam, Potsdam,

Germany;; [4]Department of Plant Sciences, University of Cambridge, Cambridge, UK

It is well known that genetic variants, such as single nucleotide polymorphisms (SNPs), are often associated with variation in gene expression. The expression of all genes in an organism is regulated by gene regulatory networks (GRNs). These networks are commonly reconstructed using gene expression profiles from multiple conditions and/or genotypes. SNPs significantly associated with gene expression may therefore be informative to infer the network of interactions between genes that code for transcription factors (TFs), acting as regulators, and their targets. Unraveling the relationships between regulators and targets and the extent to which these are affected by genetic variants is, therefore, critical to accurately reconstruct the GRN of the species of interest at a population level. In this study, we developed a regularized regression approach for GRN reconstruction, which accounts for the genetic makeup of the samples under investigation by combining SNP and RNA-seq data from a collection of maize (Zea mays) genotypes. Data was generated on 315 Recombinant Inbred Lines (RILs) belonging to the Multi-parent Advanced Generation InterCrosses (MAGIC) maize population, characterized by large diversity and dense recombination. Leaf samples were collected in the field with two replicates for each RIL and used for RNA-seq to estimate the transcription levels of 39,092 genes. Genotyping data were developed using Single Primer Enrichment Technology (SPET), targeting highly diverse genomic regions within the population and leading to the identification of 70,617 SNPs. First, we performed expression Quantitative Trait Locus (eQTL) mapping to pinpoint the genetic variants associated with differences in expression profiles. Since our analyses are performed on RILs from a multi-parent mapping population, we can exploit its wide genetic diversity and the extensive gene expression dataset produced to accurately map eQTLs and be confident in the application of the output to reconstruct the regulatory network. TFs found in the proximity of eQTLs were considered regulators of those genes (target genes, TGs) whose expression was tested for association with SNPs. Our GRN is, therefore, defined by the identified TF -TG interactions. The p-value eQTL metrics were subsequently used as genetics-informed weight to penalize the strength of association between TFs and TGs. To evaluate the performance of our model, we compared the constructed network with results from other approaches, and with experimentally verified interactions included in a gold-standard. The usage of SPET SNPs allows us to exploit verified genome-wide TFs binding sites to further confirm the truthfulness of the identified interactions, and to accurately define the set of proven negative interactions of the gold-standard. The proposed approach can be applied to any organism with a mapping population for which gene expression and genome-wide SNPs are available. The reconstructed network is then specific for the genetic background of the population, and can help us improve our understanding of the mechanisms behind the manifestation of complex traits.

GENETIC MAPPING AND MEDIATION ANALYSIS REVEALS IMMUNE PHENOTYPES UNDERLYING GENETIC SUSCEPTIBILITY TO SEVERE CORONAVIRUS DISEASE IN MICE

Risemberg, Ellen L[1]; Schäfer R Leist, Alexandra[2]; Kamat, Kalika[3]; A, Timothy[4]; Hock Bell, Pablo[5]; L, Colton[1]; R Linnertz, Darla[2]; D Miller, Ginger[3]; Pardo Shaw, Fernando[4]; de Villena, Martin T Manuel[5]; Valdar Ferris, William[4]; Baric, Ralph[5]

[1]Curriculum in Bioinformatics and Computational Biology, UNC Chapel Hill; [2]Department of Genetics, School of Medicine, UNC Chapel Hill; [3]Department of Epidemiology, School of Public Health, UNC Chapel Hill; [4]Lineberger Comprehensive Cancer Center, UNC Chapel Hill; [5]Department of Microbiology and Immunology, School of Medicine, UNC Chapel Hill

Zoonotic coronaviruses have caused three severe epidemics in the 21st century, including SARS and COVID-19, while climate change and increasing human-animal interaction raises the likelihood of future outbreaks. This motivates improved understanding of viral pathogenesis and mechanisms of susceptibility to severe disease. Studies utilizing the substantial number of COVID-19 patients worldwide have identified genomic regions associated with disease severity in humans; however, the specific genes and mechanisms underlying these associations are unclear. To identify disease-associated loci and study underlying mechanisms in more depth, we created a genetic mapping population from an F2 cross between coronavirus-susceptible and coronavirus-resistant Collaborative Cross mouse strains (CC006/TauUnc and CC044/UncJ, respectively). Approximately 1200 F2 mice were infected with mouse-adapted SARS-CoV, SARS-CoV-2, HKU3-CoV or saline. Weight loss, lung congestion score, viral titer, and immune profiles were measured. We recently reported several loci associated with disease outcome (i.e., weight loss and congestion score) following infection with all three coronaviruses, one of which is homologous with a COVID-19-associated locus in humans. That initial analysis suggested that some mechanisms of susceptibility are 1) conserved from mice to humans and 2) shared across multiple coronaviruses, including one (HKU3-CoV) not present in humans. Here, we extend that study to dissect specific mechanisms of genetic susceptibility. We perform an integrative analysis of disease severity phenotypes and immune phenotypes (i.e., viral titer and immune cell concentrations) in both control and infection groups. We apply gene-by-treatment and multi-trait quantitative trait loci (QTL) mapping to identify several loci associated with immune composition at baseline (in control mice) and following infection. We use a Bayesian model selection approach for mediation analysis to identify causal relationships underlying associated loci. Our results suggest that infiltration of certain immune cells, immune cell status, and failure to control viral replication are mediators of genetically-driven disease risk. For example, we find evidence for a mediating role of genetically driven variation in proportion of MHCII positive antigen-presenting cells. This work takes advantage of invasive phenotyping and environmental control not possible in humans to improve our understanding of coronavirus disease susceptibility, which may improve our ability to treat and control current and future outbreaks.

GENOMIC ESTIMATION OF DOMINANCE VARIANCE AND INBREEDING DEPRESSION IN A LOCAL SHEEP BREED

Rochus, Christina Marie[1]; Špehar, Marija[2]; Kasap, Ante[3]; Barac, Zdravko[4]; Ramljak, Jelena[4]; Pocrnic, Ivan[4]

[1]University of Edinburgh, Roslin Institute, Easter Bush Campus, Midlothian, EH25 9RG, UK;; [2]Croatian Agency for Agriculture and Food, Svetošimunska 25, 10000, Zagreb, Croatia; [3]Faculty of Agriculture, University of Zagreb, Svetošimunska 25, 10000, Zagreb, Croatia;; [4]Ministry of Agriculture, Grada Vukovara 78, 10000 Zagreb, Croatia

Dominance variance can account for a substantial proportion of total genetic variance depending on the species and trait. Its assessment was historically cumbersome primarily due to the size and structure of available pedigree data and corresponding computational complexities. Estimations were simplified with the availability of genome-wide SNP markers, opening the potential to study non-additive genetic variation even in small populations such as local sheep breeds. Further, it has been argued that when estimating dominance, it is essential to account for genomic inbreeding because a decrease in heterozygosity can lead to an overall decrease in the mean of a trait of interest due to directional dominance. Local livestock breeds are susceptible to inbreeding and inbreeding depression due to small population sizes and the resulting increased relatedness between mates. Local breed conservation should be prioritised because they are reservoirs of genetic diversity and have characteristics that could be essential resources for adapting to future challenges. Therefore, when designing selection programmes for small livestock populations, the effects of dominance and inbreeding need to be considered. We aimed to estimate additive and dominance genetic variances and genomic inbreeding for milk traits in Pag sheep. This local Croatian breed is adapted to a Mediterranean island with a marginal grazing system and is raised for milk (mainly used for cheese) and lamb production. We had 50K SNP array genotype data for 2134 animals, of which 1744 were ewes with milk records. After quality control of genomic data, we imputed sporadic missingness and corrected genotype errors using AlphaPeel. We detected runs of homozygosity (ROH) with Plink 1.9 to estimate genomic inbreeding (FROH). For each recorded milk trait (milk, fat and protein yields in kg, and somatic cell score), we compared four single-trait models with additive and dominance effects, and FROH as a covariate to account for directional dominance, fitted with both Bayesian and REML methods as implemented in BLUPF90. We found additive variance was stable across the different models for each trait, while dominance variance varied and was impacted by the inclusion of FROH covariate in the model. Dominance variance accounted for 10-30% of genetic variance across models and traits. Using a genome-wide association analysis approach, we will also detect regions of the genome with additive genetic, dominance genetic or ROH effects on milk traits. Finding the best approach to utilise available non-additive genetic variation in the local (small) livestock breeding programmes via either optimal contribution selection and mate allocation schemes or leveraging SNP effects associated with the non-additive genetic effects is an important open question. Genomic data has allowed us to estimate additive and dominance variance and

inbreeding in Pag sheep, which will contribute to developing sustainable genomic selection programmes for this and other small livestock populations.

COPULA MULTI-TRAIT ANIMAL MODEL TO IMPROVE THE GENETIC SELECTION

Tom, Rohmer; Brüning, Victoria; Kuhn, Estelle

INRAE

In quantitative genetics, linear mixed models (G+E) are used to deal with genetic and environmental effects. Variance components are frequently estimated using the restricted maximum likelihood (REML) estimator, based on assumptions of normality. Concerning multiple traits, the hypothesis of multi-normality of phenotypes can be violated in particular by the non-normality of the dependence structure between phenotypes, notably by a non-Gaussian copula structure on the residual part. In a recent article, it is shown that the use of a multi-trait Gaussian animal model, for traits sampled via a non-Gaussian copula for the residual part, could lead to biases on the estimated genetic parameters, such as heritability or genetic correlations, for animal populations subject to non-random selection of breeding stock. We propose here a non-Gaussian model to take into account the genetic and environmental part while considering a non-Gaussian copula structure on residual part. Stochastic gradient strategies are developed to carry out the maximum of the considered likelihood to jointly estimate genetic variance components, residual variances and copula parameters. Genetic and residual parameter estimates are compared to those obtained in the Gaussian multi-trait model estimated by Average Information-REML in simulation studies for simulated data generated from different copulas, including the Gaussian one, and also non-Gaussian copulas notably distributions with tail dependence as the Clayton copula. Illustrations on real data, from breeding traits for which multivariate Gaussian models seem irrelevant are presented. Finally, the choice of the parametric copula in the animal copula model and the robustness of the estimators to copula misspecification are discussed.

COPULA MULTI-TRAIT ANIMAL MODEL TO IMPROVE THE GENETIC SELECTION

Tom, Rohmer; Brüning, Victoria; Kuhn, Estelle

INRAE

In quantitative genetics, linear mixed models (G+E) are used to deal with genetic and environmental effects. Variance components are frequently estimated using the restricted maximum likelihood (REML) estimator, based on assumptions of normality. Concerning multiple traits, the hypothesis of multi-normality of phenotypes can be violated in particular by the non-normality of the dependence structure between phenotypes, notably by a non-Gaussian copula structure on the residual part. In a recent article, it is shown that the use of a multi-trait Gaussian animal model, for traits sampled via a non-Gaussian copula for the residual part, could lead to biases on the estimated genetic parameters, such as heritability or genetic correlations, for animal populations

subject to non-random selection of breeding stock. We propose here a non-Gaussian model to take into account the genetic and environmental part while considering a non-Gaussian copula structure on residual part. Stochastic gradient strategies are developed to carry out the maximum of the considered likelihood to jointly estimate genetic variance components, residual variances and copula parameters. Genetic and residual parameter estimates are compared to those obtained in the Gaussian multi-trait model estimated by Average Information-REML in simulation studies for simulated data generated from different copulas, including the Gaussian one, and also non-Gaussian copulas notably distributions with tail dependence as the Clayton copula. Illustrations on real data, from breeding traits for which multivariate Gaussian models seem irrelevant are presented. Finally, the choice of the parametric copula in the animal copula model and the robustness of the estimators to copula misspecification are discussed.

USING NON-ADDITIVE EFFECTS IN GENOME-WIDE ASSOCIATION STUDIES AND GENOMIC PREDICTIONS TO IMPROVE BIOTIC STRESS TOLERANCE IN PEACH

Roth, Morgane[1]; Serrie, Marie[2]; Segura, Vincent[1]; Dlalah, Naïma[2]; Cabel, Octave[1]; Malbot-Calonnec, Lucas[2]; Quilot, Bénédicte[2]

[1]UR 1052 GAFL, INRAE, Montfavet, France; [2]UMR 1334 AGAP, INRAE, Montpellier, France

Accounting for genetic architecture is crucial to breed for sustainable disease resistances and tolerances in plants. Indeed, (i) harnessing together minor and major effects genes allows to design a more durable plant immunity with large spectrum (ii) access to non-additive variance allows for a better exploitation of the total genetic variance when it comes to breeding, which is particularly relevant for clonally propagated crops. Our study system, Prunus persica (peach tree), is a major temperate fruit crop characterized by an overall high susceptibility to several pests and diseases, illustrated by a frequency treatment index around five times higher than in cereals. In this work, we phenotyped symptoms of two pests (leafhopper and twig moth) and four diseases (rust, leaf curl, mildew and shot hole) under low pesticide cover over three years in a peach core-collection replicated at three sites. This population consists in 192 unique accessions representing peach worldwide diversity and has been genotyped with the IRSC 16K SNP array. We used linear mixed models and the natural orthogonal interactions approach (Vitezica et al. 2017) to explicitly decompose genetic variance into additive, dominant and epistatic effects, and genotype x environment interactions. Genome-wide associations studies (GWAS) were performed with single-locus mixed models including kinships accounting for different dominance inheritance patterns. Genomic predictions consisted in a comparison of five GBLUP models incorporating different combinations of non-additive and inbreeding effects. After describing significant non-additive genetic variance and inbreeding effects across traits, we show that in addition to additive quantitative trait loci (QTLs), three to eight additional QTLs have been detected when accounting for dominant architecture. We were also able to improve genomic predictions by up to +0.05 in predictive ability with models

incorporating non-additive and inbreeding terms in comparison to the additive baseline GBLUP model. Our results indicate the presence of very contrasted genetic architectures within the six biotic stress responses studied, traits being more strongly influenced either by dominance, epistasis or inbreeding effects. We also present the exploitation of GxE variance to find robust QTLs and improve environment-specific predictions. Finally, we introduce a multitrait approach to exploit jointly the complementary between resistance and tolerance to several biotic stresses. Our results could translate into high genetic gain in peach given its long juvenile phase, and could contribute to a long-term reduction of pesticide reliance in fruit production. Reference Vitezica, Z.G., Legarra, A., Toro, M.A., and Varona, L. (2017). Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. Genetics 206, 1297–1307.

PLEIOTROPY AND SELECTION ON MULTIPLE TRAITS: CONNECTIONS WITH GWAS

Sachdeva, Himani; Hermisson, Joachim

Faculty of Mathematics, University of Vienna

We consider how real or apparent stabilizing selection on multiple quantitative traits shapes the distribution of trait variances and co-variances across loci in a finite population. We characterise the loci that make significant contributions to quantitative genetic variation across multiple traits, and explore how the number, effect sizes and allele frequencies of such loci depend on mutational pleiotropy and the strength of selection. Our analysis sheds light on the observation that a sizable fraction of variants are found to be significantly associated with multiple traits in genome-wide association studies (GWAS).

CROSSING STRATEGY CONSIDERING MULTIPLE TRAITS BASED ON THE ABILITY OF FUTURE INBRED LINES IN PLANT BREEDING PROGRAMS

Sakurai, Kengo[1]; Moreau, Laurence[2]; Mary-Huard, Tristan[3]; Iwata, Hiroyoshi[1]; Charcosset, Alain[2]

[1]Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo, Tokyo 113-8657, Japan; [2]Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Evolution (GQE) - Le Moulon, 91190, Gif-Sur-Yvette, France; [3]Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, 91120, Palaiseau, France

In plant breeding programs, we have to consider multiple traits, such as yields, protein contents, flowering time, and environmental tolerance. There are many cases in which we want to improve yields and maintain other traits in the desirable ranges. There are several selection strategies for multiple traits, such as index selection, tandem selection, and independent culling selection. However, there are few crossing strategies for multiple traits. Because crossing pairs largely affect the results of the breeding program, it is necessary to invent a new crossing strategy for multiple traits. Nowadays, we can estimate the progeny distribution of each crossing pair using genome-wide marker data. Utilizing the estimated progeny distribution, we can evaluate the potential of

each cross based on the genotypic value of the progeny. We invented a new crossing strategy named Cross Potential Selection for Multiple Traits (CPS-MT) to produce desirable varieties. In our research, we compared an individual selection strategy and a new crossing strategy (CPS-MT) under the breeding simulations and evaluated CPS-MT. In this breeding simulation, we conducted a 10-year breeding program. The objective of this breeding program is to improve the genotypic value of the target trait while maintaining the other trait in the desirable range. We simulated four types of relationships (no relation, positive correlation, negative correlation, and non-linear relation) between these two traits. Also, we assumed three types of causes for these relationships (pleiotropy, tight linkage, and loose linkage). We simulated 12 types of situations and evaluated CPS-MT in each situation. The evaluation of each strategy was based on the highest genotypic value of the target trait among genotypes whose genotypic value of the other trait was in the desirable range. In all situations, CPS-MT outperformed individual selection in the final year of the breeding program. Especially in the situation of the combination of non-linear relation and tight linkage, the genetic improvement of CPS-MT was 20% higher than that of individual selection. In these breeding simulation experiments, the usefulness of considering the potential of each cross was indicated.

LESSONS LEARNED FROM MODERNIZING A BREEDING PROGRAM WITH HIGH-DIMENSIONAL GENOMICS AND HIGH-THROUGHPUT PHENOTYPING

Loeb, Santantonio N.; Frias S, A.; Sabadin, J.F.G.

Virginia Tech

The small grains breeding program at Virginia Tech has made significant efforts to incorporate high-dimensional genotyping and high-throughput phenotyping into routine breeding decision-making processes. Genomic information has been used for parent selection, mate pair selection, early generation advancement and early environmental targeting of inbred lines for three years. Use of genomic information has led to increases in population level inbreeding, emphasizing the need for strategies that limit inbreeding for prolonged genetic gain, such as optimal contributions. Prediction of family merit of mate-pairs ranges from 0.64 to 0.8 for agronomic traits including grain yield and heading date. Sparse testing approaches have shown to be highly predictive of unobserved individual-environment combinations within a given year, ranging from 0.76-0.86 across 3 locations, but are highly influenced by genotype by year interactions, leading to poor prediction across years (0.0-0.4). Aerial imaging conducted throughout the growing season is being used to model genotype specific response to the environment through time using random regressions with Legendre polynomials. Vegetative indices are shown to detect genetic differences in the field, but these can be misleading when other latent factors are present, such as disease pressure. Modeling plasticity of growth curves across environments can result in reliable estimation of biomass accumulation, but weaker relationships with end-use traits such as grain yield highlight the need to account for differential carbon translocation. Reaction norms applied to time-series data have also provided

insight at which stages of growth more prominently contribute to genotype by environment interactions. However, applying these growth and development models to routine decision-making for genetic improvement is much less clear. Successes and lessons learned from routine implementation of these technologies will be discussed and a recommendation of efficacy for various efforts will be given.

ASSESSING MYBAITS TARGET CAPTURE SEQUENCING METHODOLOGY USING SHORT READ SEQUENCING FOR VARIANTS DETECTION IN OAT GENOMICS AND BREEDING

Mahmood, Khalid; Sarup, Pernille; Oertelt, Lukas; Jahoor, Ahmed; Orabi, Jihad

Targeted sequence capture systems, coupled with next-generation sequencing, have emerged as efficient tools for exploring specific genetic regions with high resolution, facilitating the rapid discovery of numerous genetic polymorphisms. Despite these advancements, the application of targeted sequencing methodologies, such as the myBaits technology, in polyploid oat species remains relatively unexplored. In this study, we utilized the myBait target capture method offered by Daicel Arbor Biosciences to detect variants and assess their reliability for variant detection in oat genomics and breeding. Ten oat genotypes were carefully chosen for targeted sequencing, focusing on specific regions on chromosome 2A to detect variants. The selected region harbors 98 genes. Precisely designed baits targeting genes within these regions were employed for target capture sequencing. We have employed various mapper and variant callers to identify the variants. After identification of variants, we focused on the variants identified by all variants callers to assess applicability of myBait sequencing methodology in oat breeding. In our efforts to validate the identified variants, we focused on two SNPs, one deletion, and one insertion identified by all variant callers in genotypes KF-318 and NOS 819111-70 but absent in the remaining eight genotypes. However, sanger sequencing of targeted SNPs failed to reproduce target capture data obtained through myBaits technology. Similarly, validation of deletion and insertion variants via high resolution melting (HRM) curve analysis also failed to reproduce target capture data, again suggesting limitations in the reliability of myBaits target capture sequencing using short read sequencing for variant detection in the oat genome. This study shed light on the importance of exercising caution when employing the myBaits target capture strategy for variant detection in oats. This study provides valuable insights for breeders seeking to advance oat breeding efforts and marker development using myBaits target capture sequencing, emphasizing the significance of methodological sequencing considerations in oat genomics research. Key words: Oat genome, myBaits technology, targeted sequencing, variant calling, genetic variants, genomic regions

HYBRID FITNESS AND THE QUANTITATIVE GENETICS OF AUTOPOLYPLOIDS: A FITNESS LANDSCAPE PERSPECTIVE

Schneemann, Hilde; Welch, John

Institute of Science and Technology Austria, University of Cambridge

When genetically differentiated populations or species come into contact and interbreed, their hybrid offspring will contain a mosaic of the genetic variants characterizing the parental lineages, re-arranged into novel combinations. The fitness of these hybrids is central to the evolution of reproductive isolation, but also plays an important role in crop and animal breeding. Hybridization is sometimes associated with polyploidization, and may involve more than two lineages. Yet how these two factors affect fitness outcomes remains unclear, partially because they greatly increase the number of parameters required in standard models. Here, I use a fitness landscape to model epistatic interactions, and obtain simple predictions for the fitness of hybrids of any ploidy and between any number of parental lineages. Re-analyzing published data on diploid and tetraploid hybrid crosses in rye (Secale cereale) and maize (Zea mays), I demonstrate that this model effectively captures dosage and interaction effects. While results resemble classical quantitative genetics models (Gallais 2003), the fitness landscape provides both restrictions and a novel interpretation of their parameters.

ACROSS-SPECIES ASSOCIATION MAPPING TO IDENTIFY THE GENETICS OF PERENNIALITY

J Schulz, Aimee[1]; AuBuchon-Elder, Taylor[2]; Costa Neto, Germano[3]; O Hale, Charles[4]; S Seetharam, Arun[5]; C Stitzer, Michelle[3]; Cinta Romay, M[4]; A Kellogg, Elizabeth[5]; B Hufford, Matthew[3]; S Buckler, Edward[4]; Hsu, Sheng-Kai[5]

[1]Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY USA 14853;; [2]Donald Danforth Plant Science Center, St. Louis, MO USA 63132;; [3]Institute for Genomic Diversity, Cornell University, Ithaca, NY USA 14853;; [4]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA USA, 50011;; [5]USDA-ARS; Ithaca, NY, USA 14853

The Andropogoneae tribe contains some of the most productive C4 grasses relevant to agriculture, including maize, Miscanthus, sorghum, and sugar cane and has adapted to a wide host of environments. Notably, while most of these species are perennial, there have been dozens of transitions to an annual life history. Perennials have multiple traits that can be harnessed to confer an advantage to agricultural crops, reducing their environmental impact, such as nitrogen remobilization and freezing tolerance. Many of these favorable traits are hypothesized to have been lost during the transition to annuality. We have previously shown that within orthogroups, annual genes are enriched for reduced sequence conservation and premature stop codons, adding support for the loss-of-function hypothesis during perennial-annual transitions. In this study, we further test this hypothesis by leveraging an expanded set of 533 short read and 38 long read Andropogoneae genomes to investigate the genetic basis of perennial-to-annual transitions across the tribe. This set of genomes allows us to test a greater number of transitions and identify instances of repeated loss of function. Using phylogenetic mixed models, we identified genes associated with perenniality across species. Additionally, we identified orthogroups with differential selection constraint in annuals and perennials. Since it is hypothesized that changes in the environment lead to changes in life

history, our models were further expanded to include environmental data to determine environmental associations with the perennial growth habit. These models will provide insight into the ecological niches of perennials and annuals. The results from this research will provide a launching point for future work to understand the adaptive potential of perennials and develop maize varieties that are more perennial-like and better adapted for climate change.

SEX DIFFERENCES IN RECOMBINATION RATES ARE ASSOCIATED WITH HOTSPOT USAGE IN SHEEP

Servin, Bertrand; Faraut, Thomas; Hazard, Dominique; Johnston, Susan; Tortereau, Flavie

INRAE, University of Edinburgh

Recombination is a fundamental biological process for the reproduction and evolution of species. Recombination phenotypes have been shown to exhibit large inter-individual variation with a significant genetic determinism. Here we make use of large genotyping datasets in the Sheep to (i) study the distribution of recombinations rates along the genome (recombination maps) and (ii) evaluate their inter-individual variation using a phenotype termed hotspot usage (HSU). First, we combined data from two studies (Johnston et al. 2016, Petit et al. 2017) to estimate sex specific recombination maps using Poisson LogNormal models of crossover counts (Chiquet et al. 2021). We found that sex differences in recombination rates are concentrated in 16% of the genome, mostly at chromosome extremities. To understand factors driving these differences, 30 additional, connected nuclear families were genotyped with a High-Density genotyping array in order to localize the crossovers with increased precision in both sexes. With the method of Coop et al. (2008), these data allowed to estimate individual HSU which revealed striking differences between sexes: males are found to preferentially use LD-inferred recombination hotspots contrary to females. This difference is most pronounced in regions with large sex differences in recombination rate. This suggests that sex difference in recombination maps in Sheep could be due to different crossover determination processes in male and female meioses

INVESTIGATING THE HERITABILITY OF ATOPIC DERMATITIS

Shen, Silvia[1]; Navarro, Pau[2]; J Brown, Sara[3]; Knott, Sara[3]

[1]University of Edinburgh, Institute for Evolution and Ecology,; [2]University of Edinburgh, Roslin Institute,; [3]University of Edinburgh, Centre for Genomic and Experimental Medicine

Atopic dermatitis (AD) is an inflammatory skin condition that affects 25% of school-aged children and 10% of adults in the developed world. Its physical and psychological impact on patients makes it the leading cause of global burden due to skin disease. Twin studies have shown that the concordance of AD between monozygotic and dizygotic twins is between 75-95%, a result that is frequently cited as a statement about heritability in academic literature and

medical advice given to patients. Recent genome-wide association studies have identified at least 81 loci significantly associated with AD, indicating a highly polygenic genetic architecture. The genetic basis of AD has motivated dermatogenetic studies into disease mechanisms and proven successful in identifying a variety of novel disease pathways and potential targets for treatment. However, linkage disequilibrium score regression (LDSC) estimates of narrow-sense heritability have found only 5-7% of the population-wide variance in AD phenotype to be attributable to additive genetic factors in European ancestry groups. There are several possible reasons for the disparity between twin studies and summary-statistics heritability estimates. Apart from methodological considerations, AD may be mediated by rare variants, which have been shown to contribute 4.5% of the total trait variance and 23% of the heritability of AD in in-house dataset analysis. Another possible explanation is historically ill-defined AD phenotypes, which can lead to misclassification of cases and controls. In this analysis, we combine newly published atopic dermatitis trait definitions established primarily for clinical biomarker research with previous methodology developed to assess the effects of rare variants and familial effects. We estimate different components of heritability using full genomic relationship matrices (GRMs) and restricted GRMs in two population-based cohorts. The restricted GRM is an estimation of a pedigree relationship matrix, capturing both the effects of shared environment and rare variants. Prospective extensions include comparing such results to rare-variant GRMs, as well as investigating heritability contributions by chromosome or genomic annotation. Knowledge about the types of genetic variants and genomic annotations that contribute significantly to phenotype variance in the population can help prioritise genes for therapeutic targets and aid functional interpretation.

NORTHERN IRELAND FARM ANIMAL BIOBANK (N.I.FAB): THE FOUNDATION FOR PRECISION ANIMAL BREEDING IN SUSTAINABLE PRODUCTION

Shirali, Masoud; Razban, Vahid; Morrison, Steven; Magowan, Elizabeth

Agri-Food and Biosciences Institute, Large Park, Hillsborough, Northern Ireland, UK

In response to the pressing global challenges of climate change, population growth, and the imperative of food security, the Northern Ireland Farm Animal Biobank (N.I.FAB) was launched in April 2022. Its aim is to establish the foundation for developing data-driven solutions and knowledge-based innovations to enhance sustainability in the livestock sector in Northern Ireland and beyond. N.I.FAB is funded by the Department of Agriculture, Environment, and Rural Affairs (DAERA) as a strategic evidence and innovation project. It has developed a 'biobank' of data representing thousands of animals from across the Agri-Food and Biosciences Institute (AFBI) and College of Agriculture Food & Rural Enterprise (CAFRE) farms. This biobank serves as a centralized database of animal phenotypic records, animal precision records, feed lab analysis results, management practices information, environmental measurements, and animal multi-omics and meta-omics lab analysis. Furthermore, it includes sample management records and bio-banked biological samples for future lab analysis.

These data, originally gathered for specific research projects, routine phenotyping, or collected through precision recording systems, hold immense potential for future use in developing new breeding and livestock management tools for livestock production. N.I.FAB places significant emphasis on addressing the main critical areas that directly impact farmers and the livestock industry in Northern Ireland and beyond. These areas include the prediction and mitigation of methane and ammonia emissions, the enhancement of animal health and welfare, and the improvement of production quality, profitability, and sustainability. As of April 2024, the N.I.FAB database contains around 50,000 ruminant animals and has collected over 10 billion rows of phenotypes, including but not limited to animal characteristics, production yield, feed intake, and methane production. N.I.FAB also contains omics and meta-omics data including over 6,000 genotyped animals, over 200 meta-genomic sequencing from rumen samples, and over 200 blood transcriptomics data. N.I.FAB equips the Northern Ireland livestock production sector with advanced technologies and data-driven approaches, improving efficiency, profitability, and promoting sustainable farming practices. This empowers farmers to make informed decisions, increase productivity, reduce costs, and embrace innovation. The work of N.I.FAB also contributes to environmental stewardship by developing strategies to reduce emissions, improve animal health and welfare, while maintaining and improving the profitability of production. N.I.FAB ensures the long-term viability of the livestock industry in Northern Ireland and helps farmers thrive in an ever-evolving agricultural landscape. Further information about the N.I.FAB is available at https://www.afbini.gov.uk/articles/nifab

NORTHERN IRELAND FARM ANIMAL BIOBANK (N.I.FAB): FOUNDATION FOR PRECISION ANIMAL BREEDING IN SUSTAINABLE PRODUCTION

Masoud S., Dzinkevicius S.[1]; Shirali, Masoud[2]

[1]Agri Food and Bioscience Institute (Northern Ireland); [2]Agri-Food and Biosciences Institute, Hillsborough, UK

GBLUP AND DEEP LEARNING INTEGRATION: A NOVEL APPROACH TO EVALUATING NONLINEARLY RELATED GENETIC TRAITS

Shokor, Fatima[1]; Croiseau, Pascal[2]; Gangloff, Hugo[3]; Saintilan, Romain[4]; Tribout, Thierry[2]; Mary-Huard, Tristan[5]; Cuyabano, C.D.[4]

[1]Eliance, 49 Rue de Bercy, 7502 Paris, France; [2]Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; [3]Université Paris Saclay, INRAE, AgroParisTech, UMR MIA Paris-Saclay, 920 Palaiseau, France; [4]Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE – Le Moulon, 91190 Gif-sur-Yvette, France; [5]Université Paris Saclay, INRAE, AgroParisTech, UMR MIA Paris-Saclay, 9110 Palaiseau, France

Genetic evaluation aims to predict the genetic values and use them to select individuals in a breeding program. Breeders often focus on improving multiple traits, usually presenting a genetic correlation. While statistical methods have

been very successful in predicting genetic values, they are unable to model nonlinear genetic relationships between traits, if present. Due to its capacity in capturing complex and nonlinear patterns in large datasets, deep learning (DL) is a promising methodology to predict nonlinear genetic relationships between traits. We proposed a multi-trait DL model which obtains predicted genetic values (PGV) while accounting for nonlinear genetic relationships between traits. We extended this model to use the output of the traditional GBLUP as its input, and then enhance its PGV by using DL to model the nonlinear relationships between traits, a model that we refer to as DLGBLUP. Using simulated data, we compared the performance of GBLUP, DL, and DLGBLUP with respect to their PGV's accuracy and the genetic progress after performing selection over seven generations. Traits were simulated as a reference trait, and six traits with nonlinear relationship with the former. Both DL and DLGBLUP models either outperformed, or presented equally accurate PGV as GBLUP, with the best results with DL and DLGBLUP being for traits with a strongly characterized nonlinear genetic relationship (prediction accuracies for a trait with quadratic relationship with the reference trait were 0.84 for GBLUP, 0.91 for DL, and 0.94 for DLGBLUP). With respect to the the genetic progress, when selecting individuals based on the PGV obtained by DL, DLGBLUP, or GBLUP, for all traits with a nonlinear genetic relationship with the reference trait, the observed genetic gain after the fourth generation was superior when PGV were obtained with either DL or DLGBLUP, when compared to GBLUP. (genetic gains after seven generations for the quadratic trait were 1.04 for GBLUP, 1.11 for DL, 2.09 for DLGBLUP). Finally, DL has potential to bring advances in genomic prediction, when used to model nonlinear relationships between traits, by improving the prediction accuracy, when compared to GBLUP, thus allowing greater genetic progress. Moreover, our proposed DLGBLUP model shows that DL can also be used as a complement to statistical methods by enhancing their performance.

COMPARISON OF IMPUTATION METHODS FOR PROJECTION OF MARKERS FROM LOW DENSITY TO HIGH DENSITY FOR GENOMIC SELECTION IN SOYBEAN (GLYCINE MAX)

Singh, Lovepreet[1]; Ramasubramanian, Vishnu[2]; Harms, Benjamin[1]; Happ, Mary[2]; Graef, George[1]; Hyten, David[2]; Lorenz, Aaron[2]

[1]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, 55108; [2]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE, 68583

Despite significant reductions in genotyping costs during the previous two decades, genotyping many thousands of progenies for genomic selection at high marker density can still be cost prohibitive, especially for small- and medium-sized breeding programs. Leveraging imputation to project high-density (HD) markers onto breeding progenies genotyped with low-density (LD) marker panels is one way to further reduce genotyping costs while maintaining accuracy of genomic prediction. The aim of this study was to evaluate and compare imputation methods for projection of markers from low-density to high-density in terms of their outcome on genomic prediction accuracy. Using the simulated and real datasets, we implemented both population-based and pedigree-based

imputation methods. HD and LD marker datasets were created and AlphaSimR tool was used to simulate breeding populations. The University of Nebraska–Lincoln (UNL) soybean breeding program provided datasets consisting of a training population (TP) and validation population (VP) which were both genotyped and phenotyped. There are 190 TP lines and 165 VP lines in the dataset. Both TP and VP lines were F4-derived lines from biparental crosses of between high yielding parental lines. After separate filtering of both TP and VP data sets to identify common SNP markers, 373,348 total SNPs across all 20 chromosomes were used. Similarly, parents of these lines were also genotyped at high density. Markers on the lines comprising the validation population were masked to simulate a situation in which breeding progenies were only genotyped with LD markers. The programs we used for imputation include Beagle 5.4, LD-KNNi (k-nearest neighbor genotype imputation method) imputation and AlphaPlantImpute2 (includes both population-based and pedigree-based imputation) program. Accuracies for both imputation and genomic prediction have been reported for different sizes of LD marker sets when projected to HD marker set.

MULTI-POPULATION GWAS GREATLY ENHANCE POWER OF QTL DETECTION IN A SMALL POPULATION

Kiel Skovbjerg, Cathrine[1]; Sarup, Pernille[2]; Wahlström, Ellen[3]; Due Jensen, Jens[1]; Orabi, Jihad[2]; Olesen, Lotte[3]; Jensen, Just[1]; Jahoor, Ahmed[2]; Ramstein, Guillaume[3]

[1]Nordic Seed A/S, Odder, Denmark,; [2]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark,; [3]Department of Plant Breeding, The Swedish University of Agricultural Sciences, Alnarp, Sweden

Genome-wide association studies (GWAS) serve as important tools for detecting individual genes with effects on traits of interest. The power of GWAS heavily depends on sample size, presenting a challenge for small populations where limited phenotypic and genetic data have been accumulated so far. In this study, we address this challenge by combining data from a small newly established 6-rowed winter (6RW) barley population with data from more advanced barley breeding programs. Focusing on heading date and stem lodging, we compare the findings of single-population GWAS with two different models for multi-population GWAS. The first multi-population model (MP1) employs a univariate approach, assuming genetic identity of the same trait across populations. In contrast, the second model (MP2) accounts for population heterogeneity using a multivariate framework which allows traits to be partly correlated between populations. Our study reveals that while both multi-population models outperform single-population GWAS in terms of power and the number of candidate quantitative trait loci (QTLs) detected, MP2 increases precision by limiting findings to genetic variants with shared effects across population. Specifically, for both traits, single-population GWAS hardly detects any associations within the 6RW population. Instead, MP2 identifies the most promising candidate QTLs, explaining up to 9.2% and 14.3% of the 6RW genetic variance of heading date and lodging, respectively. In conclusion, our study

provides a methodology for identifying genetic variants associated with traits across different populations, thus presenting a potential strategy to launch genomics-based breeding in small populations.

RANDOM REGRESSION MODELLING OF FIBRE DIAMETER MEASURED ALONG THE WOOL STAPLE FOR USE AS A POTENTIAL INDICATOR OF RESILIENCE IN SHEEP.

G Smith, Erin; Waters, Dominic L.; Walkom, Sam F.; Clark, Sam A.

Sheep frequently face challenging environmental conditions, yet there are currently inadequate methods to select for resilience against these disturbances. This study utilised random regression models to analyse the variability in wool fibre diameter measured along the staple, to assess how the genetic and environmental variances of fibre diameter change across different environmental conditions. The study used 4181 Merino sheep measured for fibre diameter at an average of 20 time points along the staple. A model containing a fifth, fourth and second-order Legendre polynomial was used to model the fixed, additive and permanent environmental effects, respectively. Results showed variability in additive genetic and permanent environmental variances along the staple, ranging from 0.55 to 0.65 and 0.30 to 0.35, respectively. The ranking of sire estimated breeding values for fibre diameter was shown to change along the staple and the genetic correlations decreased as the distance between measurements along the staple increased (ranging between 1 to 0.54). This result suggests that some genotypes were potentially more sensitive to changes in the production environment compared to others, and this could be used to inform selection for resilience in sheep.

RANDOM REGRESSION MODELLING OF FIBRE DIAMETER MEASURED ALONG THE WOOL STAPLE FOR USE AS A POTENTIAL INDICATOR OF RESILIENCE IN SHEEP.

G. Smith, Erin[1]; L. Waters, Dominic[2]; F. Walkom, Sam[1]; A. Clark, Sam[3]; Smith, Erin G.[4]; Waters, Dominic L.[5]; Walkom, Sam F.[4]; Clark, Sam A.[5]

[1]School of Environmental and Rural Science, University of New England, Armidale, NSW 35 Australia; [2]Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 35, Australia; [3]Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 351, Australia; [4]School of Environmental and Rural Science, University of New England, Armidale, NSW 235 Australia; [5]Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia

Sheep frequently face challenging environmental conditions, yet there are currently inadequate methods to select for resilience against these disturbances. This study utilised random regression models to analyse the variability in wool fibre diameter measured along the staple, to assess how the genetic and environmental variances of fibre diameter change across different environmental conditions. The study used 4181 Merino sheep measured for fibre diameter at an average of 20 time points along the staple. A model containing a fifth, fourth and second-order Legendre polynomial was used to model the fixed, additive and permanent environmental effects, respectively. Results showed variability

in additive genetic and permanent environmental variances along the staple, ranging from 0.55 to 0.65 and 0.30 to 0.35, respectively. The ranking of sire estimated breeding values for fibre diameter was shown to change along the staple and the genetic correlations decreased as the distance between measurements along the staple increased (ranging between 1 to 0.54). This result suggests that some genotypes were potentially more sensitive to changes in the production environment compared to others, and this could be used to inform selection for resilience in sheep.

HERITABILITIES AND GENETIC CORRELATIONS OF FEED INTAKE PATTERNS AND THEIR DAY-TO-DAY VARIATION IN STATION TESTED GROWING PIGS

Sölkner, Johann; Köberl, Anna-Maria; Draxl, Christian

BOKU University, Institute of Livestock Sciences, Gregor-Mendel-Strasse 33, 1180, Vienna, Austrian Pig Testing Company, Ltd., Streitdorf, Austria

A stable day-to-day pattern of measurable phenotypes, like milk production in dairy cows or feed intake of growing pigs, is currently considered an indicator of resilience in livestock populations. In this study, we used close to 30 million individual feed intake records of Large White (LW), Landrace (LR), Large White x Landrace (LWxLR) and Pietrain (PI) pigs undergoing routine performance testing from 30 to 115 kg of live weight to evaluate daily feeding patterns, including feed intake (FeedIntake), number of visits of the feeding station (NumbVisit) and duration of feeding time (FeedTime), Only animals completing the full test period were considered and polynomial regression with linear and quadratic components was fit for each individual for each of the traits. Standard deviation (STD) and log variance (LogVar) of the deviations of daily records from the regression curve were considered indicators of day-to-day variation for the three traits. Estimation of heritabilities and genetic correlations of means (Mean) and indicators of variation for the tree feed intake traits was performed, also including daily gains and feed efficiency, as routinely recorded. The number of animals included for the different breed types were 2357 (LW), 1078 (LR), 2795 (LWxLR) and 1060 (PI), with deep pedigrees, analysis was performed using VCE6 software. Mean_FeedIntake per breed ranged from 1939g (PI) to 2750g (LWxLR), Mean_NumbVisit from 7.83 (PI) to 9.91 (LW) and Mean_FeedTime from 62.2min (PI) to 66.0min (LWxLR). Heritability estimates for the feed intake traits were Mean_FeedIntake 0.25-0.41, Mean_NumbVisit 0.44-0.66, Mean_FeedTime 0.27-0.63, STD_FeedIntake 0.07-0.23, STD_NumbVisit 0.21-0.49, STD_FeedTime 0.19-0.27, LogVar_FeedIntake 0.07-0.17, LogVar_NumbVisit 0.18-0.52, LogVar_FeedTime 0.17-0.30, for feed efficiency they were 0.37-0.54 and for daily gains 0.29-0.42. Mean number of visits to the feeding station was somewhat more heritable than the other two feed intake variables and the same was true for day-to-day variation of number of visits. Yet, the mean and day-to-day variation of number of visits were highly correlated, indicating that animals visiting the feeding station more frequently on average showed also a higher variation of visits. This study indicates that two novel feed intake traits, number of visits of the feeding station and feeding time per day, are moderately to highly heritable. Use of their day-to-day variation,

either via standard deviation or log variance as novel traits indicating resilience or robustness needs still to be evaluated by correlating them with health and longevity traits of the breeds involved. This seems difficult because the number of health-related losses of animals during the performance testing period is very low, probably due to good management of the testing station.

GENETIC BASIS OF REPRODUCTIVE TRAITS ASSOCIATED TO THERMAL ADAPTATION AT HEAT ENVIRONMENTS USING DROSOPHILA MELANOGASTER AS A STUDY MODEL.

Climatic fluctuation –including temperature- is an environmental stressor that influences most biochemical and physiological functions of the organism. Under the current scenario of global warming, the ability to adapt to high-temperature environments is a relevant issue for understanding potential evolutionary responses to increasing temperature. Several quantitative fitness-related traits, with natural variation attributable to segregating variants at multiple interacting loci, are affected by the ability of organisms to respond to temperature stressors. In this regard, Drosophila melanogaster is a powerful genetic model system for identifying candidate genes associated with trait variations, such as mating success, due to the ability to accurately measure reproductive traits directly associating both genetic background of lines and environmental thermal conditions. In this study, quantitative traits locus (QTL) mapping –as analysis technique- was carried out to estimate the number and locations (usually as marker intervals) of genome regions involved in phenotypic variation of mating traits on heat and benign environments, using two set of recombinant inbred lines selected for low (K-) and high (K+) heat knockdown resistance. Two mating traits were measured: time to courtship initiation and mating efficiency, at both heat (33°C) and benign (25°C) temperatures. Mating traits at heat temperature were highly polygenetic, being influenced by multiple QTLs. This study shows that genetic basis of mating traits at heat temperature is largely different -involving loci and candidate genes related to heat responses- from the genetic basis controlling the variation for mating traits at benign temperature. These thermal-dependent effects may indicating either pleiotropy or genetic linkage between thermotolerance and reproductive traits at high temperature. In addition, many of the identified genome regions involved in reproductive traits under heat temperatures correspond to evolutionarily conserved genes –such as heat shock protein genes- and have orthologous in several organisms. Therefore, considering that the thermotolerance response seems to be partly conferred by evolutionarily conserved genome regions, studying the effects of high temperatures on different fitness-related traits could be broadly relevant, using different model organisms to understand and predict the distribution of populations under the current scenario of globally increasing temperatures.

INTEGRATED GS: PRESCRIPTIVE BREEDING TO GUIDE TRANSGENIC PLATFORM TRANSITION

Sun, Xiaochun; Steckling, Cleiton; Fabiano, Fabiano; Pita, Pita

BASF

A key issue in crop breeding is the genetic gain gap between conventional breeding and transgenic trait introgression. We propose a prescriptive breeding approach that uses genomic selection (GS), genetic simulation, and genetic diversity evaluation to bridge this gap. Our method utilizes a prescriptive optimization algorithm to streamline the integration of transgenic traits into breeding programs. This approach is designed to achieve higher genetic gains in both conventional and transgenic germplasm. This study presents an innovative strategy to address challenges in transgenic crop development. By minimizing the loss of genetic gain during introgression, prescriptive breeding can accelerate the development of elite germplasm enhanced with valuable transgenic traits.

TOWARD A BIOLOGICAL INTERPRETATION OF THE EFFECT OF PROBIOTIC SUPPLEMENTATION EXPRESSED BY THE MICROBIOME COMPOSITION OF THE FISH GUT MICROBIOME

Jakimowicz, Michalina; Mielczarek, Magda; Hajduk, Piotr; Sztuka1, Marek; Jarosz, Lura; Napora, Lukasz; Szyda, Joanna

The microbial communities characteristic for a given environment are more than just a set of particular species, genera, or families. Biologically, they represent unique compositions of metabolites that constitute various metabolic pathways. In our analysis, we focused on bioinformatic modelling of the impact of water and fish feed probiotic supplementation in an aquaculture experiment on Common carp. The experimental setup consisted of 25 ponds that were divided into 5 experimental groups (5 ponds per group) defined by: (1) no supplementation, (2) water and (3) water and food supplementation with supplement A, (4) water and (5) water and food supplementation with supplement B. From each pond microbiome, samples were obtained from the fish gut after six months of supplementation. The implemented bioinformatic workflow comprised: (1) sequencing of two hypervariable regions (v3 and v4) of the gene encoding the S16 subunit of a ribosome, (2) preprocessing of raw sequences, (3) taxonomical annotation of samples to families and the quantification of family abundance in each environment (water, sediment, gut), (4) estimation of within sample family diversity, (5) estimation of pairwise distances between samples. The differential abundance between the experimental units was then assessed, considering two biological scopes: a direct level expressed by differential abundance of microbial families and a functional level defined as differential 'abundance' of metabolic pathways. For the KEGG pathways, the most striking result was that the water supplement A had a much higher impact on gut metabolic composition than other supplements, which resulted in a much larger number of significantly altered KEGG abundance compared to the control. In general, very little overlap was observed in significant pathway sets between experimental setups, indicating that each supplementation exhibited a different effect on metabolism. The same was also observed at the direct level in individual families. The largest number of differentially abundant families was recorded for water supplement A. It is noteworthy that among the significantly differential abundant pathways there is a number of KEGG related to the biosynthesis of antibiotics. This remains well in

line with the expected beneficial effect of probiotic supplementation on the improved immune response of individuals and the corresponding reduction in the application of artificial antibiotic treatment, since the gut microbiome provides extension of the host immune system and therefore prevents infections.

SATURATED GWAS YIELDS NEW FUNCTIONAL ANNOTATION OF THE HUMAN GENOME

Thibaut, Loïc; Yengo, Loïc; Visscher, Peter

Centre for Population and Disease Genomics, Institute for Molecular Bioscience, The University of Queensland

In the most extensive GWAS conducted to date for adult height, 12,111 independent variants associated with height were identified. The genomic regions where these variants cluster, referred to as the height annotation, account for 100% of the common Single Nucleotide Polymorphism (SNP) heritability but only cover 21% of the genome, suggesting that saturation has been reached. This raises the possibility that these loci are not height specific but also capture a substantial portion of the heritability of other traits. We tested this hypothesis using 405 traits measured in UK Biobank (UKBB) participants and 35 common diseases with publicly available GWAS summary statistics. Utilising stratified LD score regression, we evaluated the enrichment of heritability for these traits in the height-associated loci, conditionally to 111 functional annotations covering a wide range of genomic properties. We found that a majority of these traits (52%) exhibit a significant enrichment of their SNP-based heritability in the height-associated loci. As expected, the magnitude of trait-specific enrichment monotonically increases with their genetic correlation with height. However, a similar proportion of traits (51%) not genetically correlated with height also display a significant enrichment for heritability. We further quantified the genetic overlap between these 405 traits and height using a bivariate causal mixture model (MiXeR) and found that traits exhibiting a large enrichment of heritability in the height loci also exhibit a large genetic overlap with height, even if their genetic correlation with height is low. For example, our analyses reveal that 98% of heel bone mineral density causal variants are also (suggestively) causal for height, while the genetic correlation between these 2 traits is low (rg = 0.1, s.e. = 0.02). Similarly, vascular disease (ICD9 code 459) shares 82% of its causal variants with height with a genetic correlation of 0.2 (s.e. 0.04). Overall, our findings suggest that pleiotropy is pervasive in the human genome and that saturated GWAS have the potential to yield new functional annotations, which can further improve discovery of new genetic associations and the accuracy of genomic prediction within and across ancestries.

CONTRASTIVE LEARNING FOR DIMENSIONALITY REDUCTION OF SNP GENOTYPE DATA

Thor, Filip[1]; Nettelblad, Carl[2]; Kovalenko, Max[3]

[1]Division of Scientific Computing; [2]Department of Information Technology; [3]Science for Life Laboratory, Uppsala University

Understanding the population structure of a genetics dataset is an important step in population genetics. One way of doing this is using dimensionality reduction methods. Over the last two decades, 2D visualizations using principal component analysis (PCA) have become commonplace. Recently, criticism has been raised against newer non-linear embedding methods, such as t-SNE and UMAP. The critique has highlighted the inability of these methods to capture admixed samples correctly, and their tendency to artificially inflate distances between populations. In this study, we utilize contrastive learning, where a neural network is trained to embed similar samples close to each other. This is done by comparing each sample with a positive sample (ideally having the same population label), and a negative sample (ideally not with the same label). The model is trained to attract the positive sample while repelling the negative one. Since we assume no prior knowledge of sample labels, we devise our positive and negative examples in a self-supervised way. We generate positive samples by applying data augmentation and choose the negatives as other samples within each training mini-batch. Our model is trained to embed the samples on the unit sphere in 3D, which can be transformed into a 2D visualization using map projections. This means that, regardless of the embedding coordinates, each sample will have neighbors in all directions. As a result, no cluster or population will be considered 'extreme', or 'different', based on their embedding coordinates. Their position can only be viewed in relation to their similarity to other samples, where the similarity is defined or described by which data augmentation is applied. This is opposed to embeddings where there is a clear center of gravity or origin. There exists limited past work on contrastive learning for dimensionality reduction of genetic data. We introduce a novel loss function that outperforms standard approaches in our experiments. Results are presented for a dataset consisting of 1355 dog samples with ~140k SNPs. We attain population classification accuracy almost on the level of t-SNE, while only training on 80% of the samples, which shows good generalization properties. Ongoing work also includes training neural network models for dimensionality reduction on much larger datasets, such as the UK Biobank dataset. Such data volumes come with new computational and methodological challenges, when it comes to stability as well as performance in the training process.

DISENTANGLING NON-CROSSOVER AND CROSSOVER GENOTYPE BY ENVIRONMENT INTERACTION FOR SELECTION

Tolhurst, Daniel; Gorjanc, Gregor

The Roslin Institute, University of Edinburgh

Genotype by environment interaction (GEI) is routinely categorised as either non-crossover or crossover interaction, which reflect changes in the scale of genotype response between environments or changes in rank. Despite the important distinction, however, current methods retrospectively diagnose and test for their presence rather than actively separate them for selection. This talk will explicitly show how non-crossover and crossover GEI can be disentangled

for artificial selection, creating two new independent traits. Here, the first trait exclusively captures the response of genotypes to non-crossover GEI while the second trait captures their response to crossover GEI, so the phenotypes do exhibit re-rankings between environments but these are independent of the changes in scale. The new approach will be demonstrated using two practical examples. The first example utilises genomic selection for breeding datasets with low and high GEI. We show how disentangling GEI can identify high performing stable genotypes across environments, but also how it can identify environments that are well representative of a breeder's target population. The second example then embeds the new approach within a breeding programme simulation for 20 years of multi-trait selection. We show how selection can be achieved using a desired gains index, and how disentangling GEI produces higher performing and more stable genotypes over time than current approaches. The new approach provides a novel way of explicitly disentangling non-crossover and crossover GEI for artificial selection of genotype performance and stability in plant and animal breeding. It is well-aligned to many quantitative genetics concepts, such as (un)-correlated response to selection, and is well-suited to a wide range of genomic selection, genome-wide association and GEI studies.

ADVANCING GENETIC IMPROVEMENT AND STABILITY OF STRIPE RUST RESISTANCE IN AUSTRALIA AND ETHIOPIA WITH THE VAVILOV WHEAT COLLECTION AND HAPLOTYPE-BASED SELECTIONS

Tong, Jingyang[1]; Tarekegn, Zerihun T.[2]; Hickey, Lee T.[1]; Periyannan, Sambasivam K.[2]; Dinglasan, Eric[1]; Hayes, Ben J.[2]; Tong1, Jingyang[1]; Tarekegn1, Zerihun T.[2]; Hickey1, Lee T.[1]; Periyannan1, Sambasivam K.[2]; Dinglasan1, Eric[1]; Hayes1, Ben J.[2]

[1]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Queensland, Australia; [2]School of Agriculture and Environmental Science & Centre for Crop Health, University of Southern Queensland, Toowoomba, Queensland, Australia

Wheat is one of the most cultivated crops worldwide, but wheat production is threatened by rapidly evolving pathogens causing disease. Stripe rust (YR) is an economically important disease that causes significant wheat yield losses globally. Genetic resistance mediated by the host plant is seen an effective avenue to manage the disease, and quantitative disease resistance (QDR) relying on multiple additive major and minor resistance (R) genes and their effective combinations is recognized as a durable host defense mechanism. However, resistance to YR is affected by environment and genotype by environment interaction (GEI) effects. Our aim was to derive selective breeding strategies to achieve rapid improvement for YR resistance that is stable across environments. We investigated disease responses of YR in the Vavilov wheat diversity panel (n = 295) in 11 field experiments, including three in Australia and eight in Ethiopia during multiple years. Significant cross-over GEI of YR resistance was observed in the environments. A Factor analytic (FA) model was used to fit to heterogeneity in both the genetic variances for each environment

and the genetic correlations between environments. Using the FA model, two breeding zones, Queensland and Ethiopia, could be clearly defined based on the genetic correlations. Line-level selection was conducted with the overall performance (OP) and stability of YR resistance in environments, where OP was the genotype score for the first latent factor (F1) expressed in trait units across environments and stability was the root-mean-square deviation (RMSD) from the regression line associated with latent variable F1. To calculate genomic estimated breeding values (GEBV) for these two traits, 34,899 genome wide markers and a GBLUP model using weighted OP and RMSD BLUEs were used. These GEBV were compared with GEBV from a multi-trait GBLUP model, where each environment was fitted as a separate trait, and performance was the average GEBV across traits and stability was the standard deviation of GEBV across environments (traits). Results showed that the GEBV from the FA approach and the multi-trait approach were in good agreement, having high correlations (Spearman r = 0.49 – 0.68). By adopting haplotype-based local GEBV, the haploblocks with the highest variances of haplotype effects were identified to be associated with OP and RMSD of YR resistance, respectively, in combined Queensland and Ethiopia environments, and the haplotypes with favorable effects on both OP and RMSD were identified. Compared with truncation selection, longer-term and higher-ability genetic gain of both OP and RMSD could be achieved by optimal haplotype selection (OHS) to choose parental crosses based on a selection index (SI) of OP and RMSD. Through simulation we also demonstrate that selections in Queensland/ Ethiopia can improve OP and RMSD of YR resistance in the constant breeding zone (i.e. Queensland/ Ethiopia), but not in changing breeding zones (i.e. Ethiopia/ Queensland), although OP to some extend can be improved. Parental selections with OHS and SI in combined environments were predicted to advance long-term genetic improvement for both average performance and stability of YR resistance across global breeding environments, including Queensland and Ethiopia.

ADVANCING GENETIC IMPROVEMENT AND STABILITY OF STRIPE RUST RESISTANCE IN AUSTRALIA AND ETHIOPIA WITH THE VAVILOV WHEAT COLLECTION AND HAPLOTYPE-BASED SELECTIONS

Wheat is one of the most cultivated crops worldwide, but wheat production is threatened by rapidly evolving pathogens causing disease. Stripe rust (YR) is an economically important disease that causes significant wheat yield losses globally. Genetic resistance mediated by the host plant is seen an effective avenue to manage the disease, and quantitative disease resistance (QDR) relying on multiple additive major and minor resistance (R) genes and their effective combinations is recognized as a durable host defense mechanism. However, resistance to YR is affected by environment and genotype by environment interaction (GEI) effects. Our aim was to derive selective breeding strategies to achieve rapid improvement for YR resistance that is stable across environments. We investigated disease responses of YR in the Vavilov wheat diversity panel (n = 295) in 11 field experiments, including three in Australia and eight in Ethiopia during multiple years. Significant cross-over GEI of YR

resistance was observed in the environments. A Factor analytic (FA) model was used to fit to heterogeneity in both the genetic variances for each environment and the genetic correlations between environments. Using the FA model, two breeding zones, Queensland and Ethiopia, could be clearly defined based on the genetic correlations. Line-level selection was conducted with the overall performance (OP) and stability of YR resistance in environments, where OP was the genotype score for the first latent factor (F1) expressed in trait units across environments and stability was the root-mean-square deviation (RMSD) from the regression line associated with latent variable F1. To calculate genomic estimated breeding values (GEBV) for these two traits, 34,899 genome wide markers and a GBLUP model using weighted OP and RMSD BLUEs were used. These GEBV were compared with GEBV from a multi-trait GBLUP model, where each environment was fitted as a separate trait, and performance was the average GEBV across traits and stability was the standard deviation of GEBV across environments (traits). Results showed that the GEBV from the FA approach and the multi-trait approach were in good agreement, having high correlations (Spearman r = 0.49 – 0.68). By adopting haplotype-based local GEBV, the haploblocks with the highest variances of haplotype effects were identified to be associated with OP and RMSD of YR resistance, respectively, in combined Queensland and Ethiopia environments, and the haplotypes with favorable effects on both OP and RMSD were identified. Compared with truncation selection, longer-term and higher-ability genetic gain of both OP and RMSD could be achieved by optimal haplotype selection (OHS) to choose parental crosses based on a selection index (SI) of OP and RMSD. Through simulation we also demonstrate that selections in Queensland/ Ethiopia can improve OP and RMSD of YR resistance in the constant breeding zone (i.e. Queensland/ Ethiopia), but not in changing breeding zones (i.e. Ethiopia/ Queensland), although OP to some extend can be improved. Parental selections with OHS and SI in combined environments were predicted to advance long-term genetic improvement for both average performance and stability of YR resistance across global breeding environments, including Queensland and Ethiopia.

IDENTIFICATION OF POLYGENIC SELECTION FOR DROUGHT STRESS IN EUROPEAN BEECH POPULATIONS

Tost, Mila[1]; Grigoriadou-Zormpa, Ourania[2]; Wilhelmi, Selina[3]; Müller, Markus[4]; Beissinger, Tim[1]; Wildhagen, Henning[2]; Curtu, Alexandru Lucian[3]; Gailing, Oliver[5]

[1]Division of Plant Breeding Methodology, University of Göttingen, Carl-Sprengel-Weg 1, 37075 Göttingen (Germany). Forest Genetics and Forest Tree Breeding, University of Göttingen, Büsgenweg 2, 37077 Göttingen (Germany).; [2]Center for Integrated Breeding Research (CiBreed), University of Göttingen, Von-Siebold-Str. 4, 37075 Göttingen (Germany). 4) Google X, Mountain View, California (USA); [3]Faculty of Resource Management, HAWK, Büsgenweg 1a, 37077 Göttingen (Germany).; [4]Universitatea Transilvania din Brasov, Eroilor 29, Brașov

500036 (Romania); [5]Universitatea Transilvania din Brasov, Eroilor 29, Brașov 500036 (Romania)

Determining the genetic basis of polygenic traits is crucial in forest genetics. Tree breeding is an extremely long and tedious process. Therefore, it is necessary to know how promising the selection will be before breeding programs are implemented. To study the genetic basis of polygenic traits, costly field experiments are implemented. Phenotypic data on traits that are measured at maturity are only available after a long time and juvenile-mature correlations are often unknown. G ̂ is a method that identifies polygenic selection on complex traits by evaluating the relationship between genome-wide changes in allele frequency and their estimates of effect sizes. Genotypic and phenotypic data were previously collected from 100 adult beech trees per stand in five locations in South-Eastern Romanian Carpathians along an altitudinal gradient associated with precipitation and temperature. Different traits related to drought stress or tree physiology were collected. Significant polygenic selection was observed for diameter at breast height (DBH), leaf carbon content and water use efficiency measured as δ13C, while no polygenic selection was observed for stomata density. In a further analysis, we want to determine whether stomata density is really not under selection or whether it is simply not as polygenic as DBH, leaf carbon content and δ13C and therefore cannot be detected as under polygenic selection. Therefore, we are also planning to perform environmental and trait-based association analysis using LFMM (latent factor mixed models) to identify SNP markers associated with precipitation and other environmental variables and the studied traits.

JOINT DISTRIBUTION OF ADAPTIVE AND QUANTITATIVE EFFECTS OF QTLS AFFECTING A COMPLEX TRAIT UNDER SELECTION

Tourrette, Elise; Servin, Bertrand

GenPhySE, Université de Toulouse, INRAE, ENVT

Large-scale genomics data offer the opportunity to better understand the evolution of genetic diversity in response to selection on complex traits (polygenic adaptation). However, the temporal dynamics of an adaptive allele results from a complex interplay between its effect on the trait, its frequency in the population, the effects and frequencies of other alleles causal to the trait and the evolutionary constraints affecting individuals (selection itself but also drift, structure …). Populations under artificial selection (e.g. experimental evolution experiments, plant and animal breeding programs) are powerful experimental designs to study this question: the fitness trait and the pedigree are known and genotyping and phenotyping data are available, usually at multiple time points. In such designs, different, complementary inferences can be done to unravel the genetic determinism of the evolving trait. On the one hand, one can estimate the quantitative effects of QTLs for the selected trait, using GWAS and genomic evaluation methods. On the other hand, one can use population genetics methods to estimate adaptive effects i.e. separate dynamics due to neutral processes from those due to selection on alleles affecting the trait. Such methods can be based for example on allele frequency trajectories or allele frequency

differences between populations selected for different trait values. In principle, the two inferences should be complementary and their combination could improve both quantitative and adaptive effects estimates. The objective of this work is to design simulation experiments of populations under artificial selection to study the best ways to perform this combination. Specifically, we simulate two populations under divergent selection for a quantitative trait, based on the design of a real experiment in livestock. The genetic architecture of the trait is modelled as the sum of the effect of a QTL and of a polygenic value representing the contribution of a very large number of QTLs. The transmission of the QTL alleles follows mendelian rules and the transmission of the polygenic value is Gaussian with parameters accounting for the kinship between individuals (infinitesimal model). Neutral markers are also simulated, to evaluate the effect of selection on allele frequency dynamics solely due to the pedigree structure resulting from selection across generations. The quantitative effects of the QTL and neutral markers are estimated using a GWAS and selective effects with various methods, making use of the particularities of the selection schemes (time-series, divergent lines). Our results reveal the dynamics of a QTL allele evolving in such experimental designs and allow to describe the joint distribution of estimates of quantitative and selective effects. This will be useful in deriving new statistical models to combine quantitative and adaptive effects estimates and help to (1) have a better understanding on how a quantitative trait evolve and (2) possibly improve the estimation of QTL effects. These, in turn, will have practical uses such as for the conservation of genomic diversity or to improve the models used for genomic predictions.

THE CONTRIBUTION OF MOLECULAR VARIATION TO GENETIC VARIANCE COMPONENTS AND PLANT FITNESS IN ARABIDOPSIS LYRATA POPULATIONS

T. Tran, Nhu L.; N. Abraham, Dr. Leen; de Meaux, Prof.Juliette

Cluster for Excellence on Plant Sciences, University of Cologne; Institute for Plant Sciences, University of Cologne

It has long been known that the potential of natural selection lies in the additive variance component, yet it is imperative to recognize the overlooked significance of non-additive variance in the evolutionary landscape. As genetic variation provides the raw material for selection, the interplay between additive and non-additive components shapes a population's adaptive potential. However, our understanding of the genomic and evolutionary factors influencing non-additive variance in natural populations is surprisingly limited hitherto. In the light of elucidating the intricate dynamics of genetic variance and its role in evolution, our study uses Arabidopsis lyrata (A. lyrata) as a study model. A.lyrata is an obligate out-crosser with a close phylogenetic relationship with self-pollinating A. thaliana in which allelic interactions do not contribute to natural variation. From two locally adapted A. lyrata populations, namely, Spiterstulen (SP) line from Norway, and Plech (PL) from Germany, we use full/half-sibling crossing design to generate inter (SP x PL) and intra (PL x PL) population crosses . We hypothesise that crossing two locally adapted populations may show elevated divergence or novel adaptive phenotypes, offering higher statistical power to

quantify genetic variance. Here we generated 400 F1 individuals, resulting from both intra-population and inter-population crosses. This allows us to quantify non-additive variance more heuristically in conjunction with the emergence of adaptive phenotypes in different populations. We apply transcript count of a gene as phenotypes, for example, 17,000 genes would mean 17,000 phenotypes. Follow up on the findings of Takou et al. (2022) showing that genes with high non-additive variance are functionally related to epigenetic programming, our research investigates deeper by integrating plant fitness, methylome data into transcriptome data. The goals are to unravel (1) the amount of additive and non-additive components at the transcriptome, methylome, and phenotype levels; (2) the covariation of genetic variance components at these three levels; and (3) the influence of historical parental populations on these components and their future trajectories. Our most recent result found that 50% variance of fitness traits (rosette size and leaf thickness) of Arabidopsis lyrata inter-population is due to genetics, with one-third attributed to non-additive variance. In intra-population, only 30% of fitness traits variation is due to genetics, more than half of which is attributed to non-additive variance. Our next step is to understand which and how selective force is associated with a population's potential for adaptation at gene expression and epigenetic levels. We will be able to associate the evolutionary factors such as selection, drift, and mutations, with gene length, genomic factors, and molecular interactions. To this end, our study is among the first to reconcile classical quantitative genetics with molecular and population genetics to address a fundamental question in both plant breeding and evolutionary biology: the amount of genetic variation that is truly available for responding to selection and the determinants thereof. The link between observations at molecular and phenotypic levels allows us to identify the molecular and genomic factors that promote the emergence of non-additive variance, potentially limiting the efficiency of selection.

CATEGORIZING GENE-BY-TREATMENT EFFECTS IN MOLECULAR COUNT PHENOTYPES USING BAYESIAN MODEL SELECTION

Harigaya, Yuriko[1]; Matoba, Nana[2]; Le, Brandon[3]; Valone, Jordan[4]; Stein, Jason[5]; Love, Michael[1]; Valdar, William[2]

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [2]UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [3]Carolina Institute for Developmental Disabilities; Carrboro, NC, USA; [4]Department of Biostatistics, University of North Carolina at Chapel Hill; Chapel Hill, NC, USA; [5]Lineberger Comprehensive Cancer Center, Chapel Hill, NC, USA

Genetic effects on a phenotype of interest can differ, sometimes markedly, in response to an applied treatment. Such gene-by-treatment interactions (GxT) can be highly consequential in biomedicine and agriculture. An effective approach to identifying GxT signals and gaining insight into molecular mechanisms is mapping quantitative trait loci (QTL) of molecular count phenotypes, such as gene expression and chromatin accessibility, under multiple

treatment conditions. Current practice, however, exhibits at least two limitations. First, a typical mapping analysis returns a list of feature-SNP pairs with significant GxT interactions but does not provide a principled approach to prioritizing GxT interaction of particular types. For example, treatment may have an impact on individuals of a certain genotype but not on others. In other cases, with similarly significant GxT interactions, treatment may affect all individuals but to different extents depending on their genotype. Formally assigning probabilities to these cases can facilitate the prioritization of response molecular QTLs for further investigation. A second potential limitation is the frequent assumption of linearity between the phenotype and the genotype after a variance-stabilizing transformation, such as the logarithm, which is routinely applied to molecular count phenotypes. This can lead to nontrivial model misspecification and inaccurate inference. Previous studies have shown that, consistent with the biologically reasonable assumption of common allelic additivity on molecular traits, such as expression, the linear relationship holds in the original count scale, but not in the transformed scale (PMID: 29021289, 29073327). To address the first limitation, we have developed a downstream method for categorizing response molecular QTLs. Our method uses Bayesian model selection and assigns ior probabilities to different types of GxT interactions for a given feature-SNP pair. To address the second limitation, we have adapted nonlinear regression to account for the inherent relationship between the genotype and phenotype. After simulation analysis, we apply our method to response molecular QTLs previously identified in human primary neural progenitor cells from genetically diverse fetal donors with and without growth stimulation (PMID: 36798360). Our method provides an intuitive way to report the evidence for different types of GxT interaction across a set of feature-SNP pairs.

MOC-BAYESW: MULTI-OMICS PENALIZED BAYESIAN REGRESSION FOR TIME-TO-EVENT PHENOTYPES

Villanueva, Ariadna; Bajzik, Jakub; Robinson, Matthew

Institute of Science and Technology Austria

Recent advancements in molecular diagnostic techniques have enabled the retrieval of massive amounts of multiple types of omics data from patients, including genomics, epigenomics, proteomics, and transcriptomics. However, we lack methods for integrating all these different data types and combining them with clinical outcomes to study the molecular mechanisms that govern pathological phenotypes. We present multi-omics BayesW, a penalized Bayesian regression method that can handle general omics data for survival analysis of time-to-event phenotypes. Our method can: (1) accommodate missing data by allowing censored individuals, (2) use continuous time-to-event data to test associations of markers with a phenotype and (3) estimate effects jointly while allowing for independent groups of biological markers. Extensive simulations using planted signals on real data demonstrate that our model performs well, accurately retrieving the true parameters of the model while controlling for false discoveries and maintaining the expected prediction accuracy. We address data

correlations by estimating the effects jointly, even between omic groups, while also estimating the individual variance explained by each group. We apply our model to two datasets. Using 18,000 individuals from the Generation Scotland study we model the association of time to onset of Type 2 Diabetes, Stroke, Ischemic Disease, and Osteoarthritis from baseline study entry, with both SNP and CpG methylation array data. We find that large proportions of variation in disease onset times can be attributed to methylation as measured in whole blood at baseline in individuals without disease symptoms. We then apply our model to The Cancer Genome Atlas (TCGA) pan-cancer dataset, in which we use 5 types of omics: copy number variation, epigenetics, tumor mutations, miRNA, and gene expression. When studying cancer survival time we find that, when fitting the 5 groups together, almost all variation attributable to "omics" data is explained by DNA methylation. We find only 3 genes that are significantly associated (95\% ior inclusion probability) with cancer survival time, conditional on all other genome-wide omics data variation. Owing to the vast variability of mechanisms characterizing different cancers, there are likely few specific genes with a strong signal in a pan-cancer setting. Nonetheless, we find previously identified protein drug targets in the methylation group and known cancer markers such as HER2. Moreover, we identify pseudogenes associated with the immunoglobulin heavy chain's variable region, potentially indicating shared mechanisms within the general immune response across different cancers. Taken together, we show the applicability of our multi-omics BayesW model to a wide-range of biological questions in multi-omics data.

TESTING THEORIES OF GENETIC VARIATION IN COMPLEX TRAITS USING MACHINE LEARNING TECHNIQUES

Villiers, Kira[1]; Huang, Helen[2]; Chen, Zhi[1]; Hayes, Ben J.[2]

[1]Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, St Lucia, Queensland, Australia;; [2]School of Electrical Engineering and Computer Science, The University of Queensland, St Lucia, Queensland, Australia

The theory of how genetic variation is established, maintained, and lost is not settled in quantitative genetics. A large barrier to comparing different proposed theories and their predictions is the theories' dependence on hard-to-measure parameters, like strength of stabilising selection and magnitude of mutational effects. We investigated the potential of simulating population genetic histories according to theories of genetic variation, and using these to train machine learning models for predicting the values of parameters of genetic variation for real datasets. Simulated histories corresponding to the breeding program design of the University of Illinois' Long Term maize Selection Experiment were reproduced under different values of 10 parameters of genetic variation. Of the search methods and machine learning methods tested, the long short-term memory (LSTM) recurrent neural network model produced the highest accuracies of prediction of variation parameters on simulation datasets. However, the parameters the LSTM predicts for a real dataset are parameters that, in simulation, produce trends that are divergent from the behaviour

observed in the real dataset. The low number of dimensional features, and hence low power, of the agricultural-style simulated datasets seem a likely cause of this gap. The use of several varied datasets or higher-detail datasets could be measures that would extend this method into a useful approach for estimating the parameters underpinning genetic variation. This would ultimately allow breeders to make informed decisions about management of genetic diversity.

PREDICTING HIDDEN INDIVIDUAL INBREEDING DEPRESSION LOAD FROM MENDELIAN DECOMPOSITION OF INBREEDING IN DAIRY SHEEP

Vitezica, Z.G.[1]; Antonios, S.[2]; Rodri´guez-Ramilo, S.T.[3]; Legarra, A.[4]; Astruc, J.M.[1]; Varona, L.[2]

[1]INRAE GenPhySE, 31326 Castanet Tolosan, France; [2]CDCB, 20716 Bowie MD, USA; [3]Institut de l'Elevage, 31321 Castanet Tolosan, France; [4]Universidad de Zaragoza, Instituto Agroalimentario de Aragón, 50013 Zaragoza, Spain

As well as lethal mutations, other (usually unknown) recessive conditions that are not lethal are carried by individuals, in what is known as load. If this "hidden" genetic load varies among founders (unequally distributed among the founder genomes), or if the founder families were exposed to different selection pressures, the offspring of different founders may be differentially affected by inbreeding. Individual variability exists in this load of unknown deleterious alleles. In other words, some individuals carry less recessive deleterious mutations than others. Two individuals could have a different genetic load (inbreeding has been accumulated in different regions of the genome) with the same inbreeding coefficient. Clearly, inbreeding is an imperfect measure of the hidden load of an individual because it cannot distinguish the accumulation of favorable homozygosity from deleterious alleles. There is therefore a polygenic individual inbreeding depression load (IDL) for which individuals differ. In complex pedigrees, each individual possesses parts of inbreeding coming potentially from different ancestors. These fractions can however be computed using pedigree by the Mendelian decomposition of inbreeding. Using these fractions, a linear model then predicts the IDL of the individuals, which is an additive trait that is expressed only in inbred individuals. Thus, even if the contributing loci are unknown, IDL can be predicted in the same manner that we do for additive genetic values. We predicted the IDL for milk yield of each individual in dairy sheep breeds. The full decomposition of the co-ancestries shows that the founders and non-founders animals contribute to the actual inbreeding. Results confirm the presence of heterogeneity in the IDL among individuals. Animals with less load could be selected to avoid inbreeding depression in future generations and undesirable matings can be discarded.

INBREEDING DEPRESSION AND PURGING INFERENCE CONSIDERING THE EFFECTS OF AUTOSOMAL AND SEX-SPECIFIC INHERITANCE ON MILK PRODUCTION IN DAIRY COWS

In highly selected dairy cattle populations, inbreeding is unavoidable, leading to an accumulation of harmful mutations and inbreeding depression. Inbreeding

depression in milk production is a well-documented phenomenon in many dairy cattle populations. Nevertheless, the assessment of inbreeding depression due to a higher sex-specific inbreeding coefficient is very rare, while in dairy cattle breeding the intensive use of elite bulls can lead to extremely high sex-specific inbreeding. Furthermore, we are not aware of a single study in which the elimination of high sex-specific inbreeding has been documented. Our main objective in this study was to analyze the inbreeding depression for milk yield in Czech Holstein cattle (17129 cows with recorded performance and good pedigree quality – equivalent complete generation equal to 12.3), broken down by classical and sex-specific inbreeding coefficient. In addition, we also analyzed whether purging is present in both classical and sex-specific inbreeding depression. Before performing purging analyses, we improved our gene-dropping based software GRain (to GRainSX) to allow calculations of classical and ancestral sex-specific inbreeding coefficients (classical sex-specific inbreeding coefficient, new Kalinowski and ancestral sex-specific inbreeding coefficient). We ran several different models – combinations of inbreeding coefficients - all based on a Bayesian inference approach. We used the software gibbs2f90 with two parallel Gibbs sampling chains with 500,000 iterations, where the first 50,000 iterations (burn-in period) were discarded, while the dependence between successive iterations was adjusted by tinning at every twentieth iteration. Negative linear regressions (median) of milk yield were observed in the model with classical (bF CD99% from -36.6 to -22.9 kg milk) and sex-specific (bFsex CD99% from -3.9 to 2.3 kg) inbreeding coefficients (per 1% increase). Similar results were obtained in the model with Kalinowski new (bFnew CD99% from -73.5 to -47.0 kg milk) and Kalinowski new sex-specific (bFnew-sex CD99% from -6.0 to 2.7 kg milk) inbreeding coefficients (per 1% increase). However, the linear regression estimates of sex-specific inbreeding coefficients were not significant, probably due to the lack of statistical power. Similarly, purging in other complex models was not significant. Although our work has methodologically enriched the general knowledge on estimating inbreeding depression and performing purging analyses in the context of sex-specific inbreeding, much work is still needed to draw reliable conclusions.

EXPLOITING THE MULTI-OMICS DATA TO DECIPHER FUNCTION GENES OF THE EGG QUALITY TRAITS IN CHICKENS

Wang, Xiqiong; Zhong, Conghao; Consortium, TheChickenGTEx; Sun, Congjiao; Yang, Ning

State Key Laboratory of Animal Biotech Breeding and Frontier Science Center for Molecular Design Breeding, China Agricultural University, Beijing, 100193, China

Chicken eggs provide abundant protein and nutrient for human at a relative low price. A large number of eggs are produced annually for human consumption worldwide. The egg quality traits include egg weight, yolk color, albumen height, Haugh unit, egg shape index, eggshell weight, eggshell thickness, and eggshell strength. Egg quality is subject to degrade with the aging process of laying hens, which hinders the developmental trend to prolong the laying cycles of egg-type chicken in the future. Therefore, understanding the genetic control for dynamic

egg quality with aging process is of great economic and biological importance. The combination of GWAS and multidimensional epigenetic annotation information, as well as molecular QTL regulation information, can help identify causal variants and decipher the genetic mechanisms of complex traits. In this study, an F2 resource population was derived from reciprocal crosses between White Leghorn (WL) and Dongxiang chickens (DX), a Chinese indigenous strain. In total, 1512 hens were chosen for SNP genotyping. In this study, we conducted GWAS analysis in the chicken population with longitudinal egg quality traits at 11 age points from the age of the first egg (AFE) to 72 weeks of old. The final ChickenGTEx data will release included 28 chicken tissues and 5 types of molecular quantitative loci (PCG expression, lncRNA expression, exon expression, splicing variation and 3'UTR alternative polyadenylation). We systematically integrated molQTLs and regulatory elements annotation with GWAS result of egg quality traits by applying four complementary methods, including fastENLOC, summary-data-based MR (SMR), single-tissue transcriptome-wide association study (sTWAS), and multi-tissue TWAS (mTWAS) to decipher function genes. The results showed that GWAS analysis for dynamic egg quality traits identified thousands of significant associations, which were much more than those screened by previous GWAS studies focused on certain age point in chicken. It was notable that more associations were discovered for old hens (40-72 weeks) than young hens (AFE to 36 weeks), suggesting certain genetic variants were age-dependent. In addition, we found that each type of molecular phenotypes can make a specific contribution to complex traits at distinct levels of gene regulation. For example, the eQTL of AREL1 in duodenum exhibited a strong association with yolk weight at AFE, and a colocalization between GWAS loci of egg weight at 40-week-old and molQTL of IRF5 was observed for a spleen sQTL. Most egg production traits were significantly enriched in intestine-specific EnhAs, except for AFE (enriched in brain-shared EnhAs), which was consistent with a previous report that the hypothalamus-pituitary-gonad axis plays important roles in AFE. The resources generated by the ChickenGTEx consortium and FAANG offer unprecedented opportunities to advance our understanding of the biology of chicken traits. These data thus represent a valuable resource, enabling novel biological insights and facilitating follow-up studies of causal mechanisms. Moreover, the new candidate genes or causal DNA variants we found in this study will help improve genomic prediction accuracy and could be helpful to future marker-assisted selection and genomic selection in chickens.

T-COJO: A TRANS-ANCESTRY CONDITIONAL AND JOINT ANALYSIS APPROACH

Wang, Xiaotong[1]; M Visscher, Peter[2]; R Wray, Naomi[3]; Yengo, Loic[3]

[1]Department of Psychiatry, Univeristy of Oxford, Oxford UK; [2]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU; [3]Nuffield Department of Population Health, Univeristy of Oxford, Oxford, UK

Extensive linkage-disequilibrium (LD) between genetic loci poses challenges for estimating the number of independent causal variants represented at a trait-associated locus. Methods such as conditional and joint multiple-SNP of GWAS

summary statistics (GCTA-COJO) have been developed to identify additional association signals within a locus. However, most of these methods can only be applied to GWAS summary statistics derived from a single inferred ancestry. GWAS meta-analyses across multiple inferred ancestry groups are becoming increasingly available, but the resulting GWAS summary statistics are not readily analysable using standard methods. Here, we present a trans-ancestry conditional and joint analysis method (T-COJO), designed to perform step-wise conditional and joint association analysis in the trans-ancestry setting. By leveraging diverse LD structures from different ancestries, we expect to gain better resolution in tagging true causal variants. In simulations, we demonstrate that by modelling diverse LD structures in different ancestries, we are, on average, able to tag significant more true causal variants when the sample consists of 50% Europeans and 50% Africans, compared with a sample of 100% Europeans of the same sample size. Our research also informed that under certain conditions, such as when the density of causal loci is very high in a specific region (e.g., the Lipoprotein A locus of chromosome 6q25.3-26), current single ancestry methods might overestimate the number of independent signals. Our approach allows the input of the LD reference to be either individual genotypic level data (default in COJO) or an LD correlation matrix. Allowing an LD correlation matrix as input should promote sharing information globally, especially when sharing individual genotypic level data is prohibited due to policy/ethical restrictions.

NEW MATHEMATICAL TOOLS FOR ANALYZING GENETIC CONSTRAINTS WITH G MATRIX

Watanabe, Junya

Autonomous University of Barcelona, Spain; University of Cambridge, UK

Theory on multivariate character evolution predicts that a population's response to directional selection is biased toward the major axis of the additive genetic covariance matrix (G matrix). The genetic constraints in this sense are often analyzed by using certain statistics, collectively known as evolvability measures, derived from G matrices and response vectors. However, it has not been fully understood how these statistics can be meaningfully interpreted or even accurately calculated in empirical analyses. Here, I present results from my recent work on analytic properties of evolvability measures, aiming to enhance our toolkit for analyzing genetic constraints. A key in deciphering analytic properties of evolvability measures is their structure as ratios of quadratic forms in selection gradient vectors. We can borrow useful theoretical results on this form of statistics from the statistical and econometric literature to derive explicit expressions for their moments and probability distributions. These enable us to calculate descriptive statistics such as the mean conditional evolvability and mean autonomy/integration, for which only approximate evaluation methods were known in the biological literature. The traditional "random skewers" method for pairwise comparison between G matrices can also be replaced by a more accurate analytic version. Theory of antieigenvalues provides the theoretical minimum of autonomy and flexibility, which can be used to quantify

how freely phenotypes can respond to selection. Retrospective inferences on genetic constraints from phenotypic divergence vectors can be enriched by utilizing the probability distributions of evolvability measures with respect to varying directions of divergence vectors. Most of these theoretical results have been implemented in my R package qfratio, which hopefully serves as a useful infrastructure for empirical analyses.

REVIVING THE DESIRED GAINS INDEX: AN OPTIMAL SOLUTION FOR PARENT SELECTION IN PUBLIC PLANT BREEDING PROGRAMS

Werner, Christian R[1]; Gardner, Keith A[2]; Gemenet, Dorcus C[3]; Tolhurst, Daniel J[3]

[1]Accelerated Breeding Initiative (ABI), Consultative Group of International Agricultural Research (CGIAR), Texcoco, Mexico.; [2]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico.; [3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, United Kingdom.

Plant breeding programs aim to simultaneously improve multiple traits of commercial importance to release varieties that meet the complex and dynamic requirements of growers, processors, and end-users. However, simultaneous improvement of multiple quantitative traits is challenging. Traits are often unfavorably correlated, and selection on one trait affects the response to selection that can be realized in other traits. Index selection has been shown to be more efficient for selecting parents to improve multiple quantitative traits in plant breeding populations compared to other parent selection strategies, such as independent culling and tandem selection. In index selection, the traits of interest are first weighted by their importance in terms of a breeding objective, and then combined into a single value which represents the merit of an individual. The breeding objective can be defined, for example, in terms of total economic profit or gain per trait. The individuals with the highest selection index are the best parents to achieve the breeding objective in the next generation. Despite the widespread adoption of the Smith-Hazel index to maximize economic profit in animal breeding programs, the application of index selection in public plant breeding programs remains underutilized. Assigning precise economic weights to all quantitative traits of interest is a complex, laborious, and costly process. Therefore, many public plant breeding programs face difficulties in quantifying an individual's value in terms of profitability. Moreover, economic gain does not necessarily translate into genetic gain, and poorly chosen economic weights can lead to adverse genetic progress in traits intended to exhibit positive gains. We advocate for the desired gains index as an optimal solution for simultaneous selection on multiple quantitative traits to identify crossing parents in public plant breeding programs, where economic weights are usually unknown and difficult to obtain. Breeders find the concept of desired gains highly intuitive, and the integration of the index into the selection process is straightforward once the desired gains are defined. Additionally, the desired gains index can be directly integrated with other strategies for optimizing parent selection, such as optimal cross to selection, to harmonize long-term genetic

gain and genetic variance. While this is known among quantitative geneticists, our intention is to bring the desired gains index back into the spotlight of the public plant breeding community and use stochastic simulation to demonstrate the efficiency and user-friendliness of this strategy to parent selection without the utilization of mathematical equations and selection theory.

IDENTIFYING RECESSIVE RESISTANCES AGAINST POWDERY AND DOWNY MILDEW IN THE DOMESTICATED GRAPEVINE

Wettstein, Gregor[1]; Wullschleger, Gianna[1]; Melgar, Aurélie[1]; Schnée, Sylvain[2]; Gindro, Katia[2]; Patocchi, Andrea[2]; Wuest, Samuel[2]

[1]Breeding Research Group, Research Division Plant Breeding, Agroscope, Switzerland; [2]Mycology Group, Research Division Plant Protection, Agroscope, Switzerland

Cultivation of grapevines originating from the common grape (Vitis vinifera) requires frequent phytosanitary treatments to protect them from fungal diseases like powdery and downy mildew. The breeding of more resistant cultivars through the introgression of dominant resistant (R) genes from related resistant species is challenged by the pathogens' ability to rapidly evolve and evade resistance mechanisms conferred by R-genes. A more durable breeding approach could be the employment of recessive resistances. These rely on the loss of susceptibility (S) factors, which, when hijacked by biotrophic pathogens, facilitate host compatibility and thus susceptibility. The well-known Mildew Locus O (MLO) gene family contains several S-genes whose recessive alleles confer durable and broad-spectrum resistance against powdery mildews in a wide variety of plants. Besides few known S-genes, plant genomes are presumed to carry many more S-genes constituting their susceptibilities to various biotrophic agents. Due to clonal propagation and a prior history of obligate outcrossing, self-fertile domesticated grapevines have accumulated many recessive mutations – typically occurring in a heterozygous state –, some of which potentially reside in the presumed S-loci. By selfing grapevines, we aim to discover such S-genes by genetic mapping of recessive resistances, followed by cloning of the respective mutant alleles. We have selfed a selection of over hundred grapevine cultivars and have assessed the susceptibility of the selfing progeny to powdery and downy mildew infection. From this, we have identified several candidate cultivars where the susceptibility phenotype segregated consistent with recessive resistances of large effects. We aim to map the responsible gene(s) by QTL mapping using SNP markers yielded from shallow whole-genome sequencing data. Obtained QTL will be fine mapped and candidate genes validated by targeted mutagenesis. The identified recessive resistances could be used to breed new resistant grapevines, improve existing cultivars through genome editing, and might even be utilized in other crops if these carry homologous genes.

QUANTITATIVE GENETIC APPROACHES FOR THE MANAGEMENT OF HIGHLY INBRED DOG POPULATIONS WITH MULTIPLE GENETIC DISEASES

Windig, Jack J.; Doekes, Harmen P.

Centre Genetic Resources, Animal breeding and Genomics, Wageningen UR

Pedigreed dog breeds often suffer from genetic disorders and hereditary diseases. Improving their genetic health is nowadays one of the most important aspects of dog breeding. There are two causes for impaired genetic health: 1) excessive inbreeding rates and 2) strict adherence to breeding standards that ignore health. An example of the first is the Saarloos Wolfdog which has had inbreeding rates above 12% per generation and as a consequence its fertility level dropped to dangerously low levels. An example of the second is the large body size of the Irish Wolfdog causing a short lifespan. To combat these effects the veterinary approach has been dominant. It focuses on individuals, their diagnosis and developing DNA tests to detect carriers of genetic defects. There is, however, a risk of excluding too many dogs and thereby further reducing the effective population size. For example, 95% of the dogs of the Bouvier des Flandres are diagnosed suffering from eye disorders. Furthermore, detrimental alleles with a small effect are generally not detected and their frequency may continue to increase. Especially for polygenic disorders this can pose a problem. A quantitative genetic approach at the population level, however can provide a long term perspective. Computer simulations show that exclusion of dogs with a high Mean Kinship (MK) to all other dogs in the population is generally the most effective way to increase the effective population size. Another example is the estimation of breeding values for polygenic disorders such as hip dysplasia in breeds from the UK, or longevity in the Irish Wolfhound. In case of multiple disorders developing a health index in which the different diseases and defects are weighted by severity and publishing its breeding values is the best option. A major challenge is, however, to explain the often difficult concepts of population and quantitative genetics to owners and breeders, in order to get them accepted and implemented.

A RESOURCE OF BOVINE GENETIC MAPS AND A USE CASE OF SELECTION SIGNATURES

Wittenburg, D.[1]; Ding, X.[2]; Melzer, N.[3]; Abdollahi Sisi, N.[1]; Schwarzenbacher, H.[2]; R. Seefried, F.[3]

[1]Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany,; [2]ZuchtData GmbH, Dresdner Straße 89/B1/18, 1200 Vienna, Austria,; [3]Qualitas AG, Chamerstrasse 56, 6300 Zug, Switzerland

Genetic diversity among cattle breeds exits due to their demographic history and different breeding objectives for meat, dairy or dual purpose. Taking this into account, targeted breeding strategies can be further enhanced by employing breed-specific genetic maps. The aim of this study was to derive genetic maps for a selection of commercial breeds. We analysed genotype data from eight cattle breeds with sample size ranging from 4,181 to 367,056. Because various assays were used for genotyping the animals, we streamlined the data preparation and analysed the data with a standardised workflow. We investigated the frequency of paternal recombination events and derived genetic-map coordinates of about 50K SNP markers. Additionally, estimates of

recombination rate between intra-chromosomal pairs of markers enabled the localisation of further putatively misplaced markers or regions in the bovine genome assembly ARS-UCD1.2 which have been excluded from genetic-map estimation. Estimates of map length obtained from a deterministic approach varied from 24.0 M to 27.4 M between breeds. To explore recombination activity interactively and to evaluate differences between breeds, we implemented all results in an R Shiny app "CLARITY", constituting the resource of genetic maps. As a particular use case for genetic maps, selection signatures based on iHS and XP-EHH statistics were investigated and revealed 73 unique regions of discrepant haplotype homozygosity within breed or between breeds. In future research, it may be investigated whether the breed purpose is also visible from these footprints in the genomes by evaluating information about genes located in candidate regions.

OPTIMAL IMPLEMENTATION OF GENOMIC SELECTION IN POTATO BREEDING PROGRAMS

Po-Ya, Wu[1]; Benjamin, Stich[2]; Juliane, Renner[3]; Katja, Muders[4]; Vanessa, Prigge[5]; Delphine, Van Inghelandt[6]

[1]Institute of Quantitative Genetics and Genomics of Plants, Heinrich Heine University, Düsseldorf, Germany;; [2]Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich Heine University, Düsseldorf, Germany;; [3]Max Planck Institute for Plant Breeding Research, Köln, Germany;; [4]Böhm-Nordkartoffel Agrarproduktion GmbH & Co. OHG, Hohenmocker, Germany;; [5]NORIKA GmbH, Sanitz, Germany;; [6]SaKa Pflanzenzucht GmbH & Co. KG, Windeby, Germany; (7) Present address: Institute for Breeding Research on Agricultural Crops, Federal Research Centre for Cultivated Plants, Sanitz, Germany

Genomic selection (GS) has emerged as a powerful tool to increase the genetic gain of complex traits in breeding programs of various animal and plant species. However, its optimal integration especially in clone breeding programs (for example potato), and its combination with the cross-selection (CS) method in heterozygous and tetraploid crops to balance genetic gain and diversity in long-term breeding programs are still unclear. In this study, we performed computer simulations based on an empirical genomic dataset of tetraploid potato to (i) investigate how the weight of GS relative to phenotypic selection, the stage of the GS implementation, the correlation between an auxiliary trait and a target trait, and the prediction accuracy affect the genetic gain of the target trait, (ii) determine the optimal allocation of resources maximizing the genetic gain of the target trait, and (iii) assess how different CS methods incorporating GS with or without consideration of genetic variability affect both short- and long-term genetic gains and diversity. Our results suggest that implementing GS with optimal selection intensities had a higher short- and long-term genetic gain compared to the phenotypic selection solely. In addition, implementing GS in consecutive selection stages largely increased genetic gain compared to using GS in only one stage. Furthermore, our results suggest that the optimal selection intensity per stage requires to be adjusted under different scenarios in

dependence of the selection strategies, prediction accuracy of the GS model, correlation between the traits, etc. When studying the long-term selection response, the CS method considering additive and dominance effects to predict progeny mean based on simulated progenies (MEGV-O) reached the highest accuracy in predicting progeny mean and the highest long-term selection gain among the assessed mean-based CS methods. However, MEGV-O also resulted in a quick loss of genetic variability. The linear combination of usefulness criteria (UC) and genome-wide diversity (called EUCD) kept the same level of genetic gain but a higher genetic diversity, compared to UC and MEGV-O. Therefore, EUCD showed a high efficiency in converting diversity into genetic gain. However, choosing the most appropriate weight to account for diversity in EUCD depends on the genetic architecture of the target trait and the breeder's breeding objectives. Our results provide breeders with concrete methods to improve their potato breeding programs and are presumably also helpful for other clone breeding programs.

THE GENOMIC SELECTION SIGNATURES OF INTERACTING ENVIRONMENTAL STRESSORS

Xiao, Changyi[1]; Duarri-Redondo, Sara[2]; Nolte, Viola[1]; Schlötterer, Christian[2]

[1]Institut für Populationsgenetik, Vetmeduni Vienna, Austria;; [2]Vienna Graduate School of Population Genetics, Vienna, Austria

Adaptation of natural populations involves the concurrent response to multiple stressors, yet statistical tests aimed to identify selection from sequence polymorphism data do not account for the potential interactions among these stressors. To assess the impact of such interactions we evolved two groups of polymorphic Drosophila simulans populations in a new complex environment that differs only in the temperature regime. One group of populations evolves under a constant 23°C while the other experiences fluctuations around the same mean temperature. Despite both groups exhibiting a similar increase in fitness, indicating comparable adaptation to the shared novel environment, their genomic responses were notably divergent: only 6.4% of the SNPs showing a significant response are common between the two temperature regimes. We propose that genetic redundancy coupled with environment-specific pleiotropy of selected SNPs underlies the utilization of alternative paths of adaptation. Overlooking these interactions could have resulted in the interpretation of temperature-specific adaptation, rather than revealing distinct adaptive trajectories to the same stressors. Our findings highlight the necessity for selection scans to incorporate stressor interactions to avoid erroneous conclusions.

ASSEMBLY OF A CHROMOSOME-SCALE AND HAPLOTYPE-RESOLVED SUGARCANE REFERENCE GENOME

Xu, Song[1]; Wang, Jing[2]; Lu, Fei[3]

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing,

China.; [2]University of Chinese Academy of Sciences, Beijing, China.; [3]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. #These authors contributed equally to this work. .

Rapid changes in the global climate and high demand for sugar have prompted sugarcane breeding to shift from traditional methods to precision design. The lack of a complete sugarcane cultivar reference genome is a substantial challenge. Given the highly allopolyploid and autopolyploid characteristics of the sugarcane genome, the current assembly either focuses on diploid species within the Saccharum complex or founding Saccharum species. Here, we report on the assembly process of GT42, the leading sugarcane cultivar planted in China. By generating deep coverage and high-quality PacBio HiFi reads, utilizing Oxford Nanopore 100-kb ultra-long sequencing and high-throughput chromatin conformation capture technologies, we expect to phase allelic contigs and assemble a haplotype-resolved and chromosome-scale genome in a few months. We believe this work will represent a major step forward in understanding the complexity of the allopolyploid and autopolyploid sugarcane genomes and improving the efficiency of modern sugarcane breeding.

GENOMIC MATE-ALLOCATION STRATEGIES EXPLOITING ADDITIVE AND NON-ADDITIVE GENETIC EFFECTS TO MAXIMISE TOTAL CLONAL PERFORMANCE IN SUGARCANE

Yadav, Seema[1]; M. Ross, Elizabeth[2]; Wei, Xianming[3]; Powell, Owen[4]; Hivert, Valentin[5]; T. Hickey, Lee[6]; Atkin, Felicity[7]; Deomano, Emily[5]; S. Aitken, Karen[6]; P. Voss-Fels, Kai[7]; J. Hayes, Ben[7]

[1]Queensland Alliance for Agriculture and Food Innovation, Queensland Bioscience Precinct, 306 Carmody Rd., St. Lucia, Brisbane, Queensland 4067 Australia;; [2]Sugar Research Australia, Mackay, QLD 4741, Australia;; [3]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4067, Australia;; [4]Sugar Research Australia, Meringa Gordonvale, QLD 4865, Australia;; [5]Sugar Research Australia, 50 Meiers Road, Indooroopilly, QLD 4068, Australia;; [6]Agriculture and Food, CSIRO, Queensland Bioscience Precinct, St. Lucia, Brisbane, QLD 4067, Australia;; [7]Department of Grapevine Breeding, Hochschule Geisenheim University, Germany

Sugarcane (Saccharum spp.) is the world's primary source of sugar, accounting for over 70% of global sugar consumption and contributing approximately AUD 4 billion annually to the Australian economy. Despite its economic significance, the rate of yield improvement in sugarcane lags other major crops. Genomic selection (GS), a revolutionary breeding tool that has significantly advanced animal and plant breeding, offers a promising solution to accelerate genetic gains in sugarcane. This research explores the potential of GS to improve genetic gain in sugarcane by focusing on three key industrial traits: tonnes of cane per hectare (TCH), commercial cane sugar (CCS), and fibre content. This study involved genotyping almost 3,000 clones for 26,000 genome-wide SNPs. We found large non-additive effects, which accounted for almost two-thirds of the

total genetic variance for TCH. Additionally, genome-wide heterozygosity significantly influenced TCH. By incorporating heterozygosity and non-additive genetic effects into our models, we achieved a 31% improvement in the prediction accuracy for TCH clonal performance. Furthermore, we evaluated the impact of using continuous genotypes, serving as a proxy for allele counts, in genomic predictions for this highly polyploid crop. Our research also investigated mate-allocation strategies that consider both additive and non-additive genetic effects to optimize progeny performance in GS-assisted breeding programs. Using these strategies, progeny performance for TCH, CCS, and fibre improved by 57%, 12%, and 16%, respectively. However, we observed a decline in additive gains, particularly for TCH, likely due to large epistatic effects, highlighting a trade-off in breeding strategies. In conclusion, our findings demonstrate that significant improvements in clonal performance can be achieved in sugarcane by exploiting non-additive effects and employing continuous genotypes. Implementing these strategies in breeding programs could substantially benefit sugarcane growers worldwide by accelerating the development of new, high-performing varieties. This research underscores the transformative potential of genomic technologies in enhancing the efficiency and outcomes of sugarcane breeding programs.

ESTIMATING THE GENOMIC LANDSCAPE OF HUMAN COMPLEX TRAITS USING SUMMARY ASSOCIATION STATISTICS

Yao, Yue; Lin, Wenzhuo; Shen, Xia

Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China; State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, School of Life Sciences, Fudan Univer

Genetic variants across the genome play a crucial role in determining complex traits, which have a polygenic architecture. To better understand this complexity and improve genomic prediction, it's essential to infer the genetic effects distribution throughout the genome. The heteroscedastic effects model (HEM) is a powerful tool that estimates genome-wide SNP effects using a generalized ridge regression framework. However, HEM is limited by its need for individual-level genotype and phenotype data, which makes it challenging to apply in large-scale human genetics studies. To address this issue, we developed a new model called SumHEM, which enables HEM to be fitted using summary statistics from genome wide association studies (GWAS). We tested SumHEM using simulations and real data analysis and found that it outperforms state-of-the-art methods such as LDpred2 in heritability parameter estimation, genomic effects distribution estimation, and genomic prediction. SumHEM is particularly effective for highly polygenic traits, as demonstrated in our analysis of 300 phenotypes from the UK Biobank. In fact, SumHEM's out-of-sample prediction for more polygenic traits with higher heritability was significantly better than that of LDpred2. Additionally, SumHEM provided comparable heritability estimates to those of LD score regression (LDSC) while revealing a more accurate genome-

wide genetic effects profile of each complex trait underlying the heritability model.

EVALUATION OF HERITABILITY PARTITIONING APPROACHES IN LIVESTOCK POPULATIONS

Yuan, Can[1]; Gualdrón Duarte, José-Luis[1]; Charlier, Carole[1]; Takeda, Haruko[1]; Georges, Michel[1]; Druet, Tom[1]; Yuan, Can[2]; Gualdrón Duarte, José Luis[2]; Takeda, Haruko[2]; Georges, Michel[2]; Druet, Tom[2]

[1]Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège, Liège, Belgium; Walloon Breeders Association, Rue des Champs Elysées, 4, 5590 Ciney, Belgium; [2]l'Hôpital, 1, 4000 Liège, Belgium, 2Walloon Breeders Association, Rue des Champs Elysées, 4, 5590-Ciney, Belgium

Background. Heritability partitioning approaches estimate the contribution of different functional classes, such as coding or regulatory variants, to the genetic variance. This information allows a better understanding of the genetic architecture of complex traits, including complex diseases, but can also help improve the accuracy of genomic selection in livestock species. However, methods have mainly been tested on human genomic data, whereas livestock populations have specific characteristics, such as high levels of relatedness, small effective population size or long-range levels of linkage disequilibrium. Results. Here, we used data from 14,762 cows, imputed at the whole-genome sequence level for 11,537,240 variants, to simulate traits in a typical livestock population and evaluate the accuracy of two state-of-the-art heritability partitioning methods, GREML and a Bayesian mixture model. In simulations where a single functional class had increased contribution to heritability, we observed that the estimators were unbiased but had low precision. When causal variants were enriched in variants with low (<0.05) or high (> 0.20) minor allele frequency or low (below 1st quartile) or high (above 3rd quartile) linkage disequilibrium scores, it was necessary to partition the genetic variance into multiple classes defined on the basis of allele frequencies or LD scores to obtain unbiased results. When multiple functional classes had variable contributions to heritability, estimators showed higher levels of variation and confounding between certain categories was observed. In addition, estimators from small categories were particularly imprecise. However, the estimates and their ranking were still informative about the contribution of the classes. We also demonstrated that using methods that estimate the contribution of a single category at a time, a commonly used approach, results in an overestimation. Finally, we applied the methods to phenotypes for muscular development and height and estimated that, on average, variants in open chromatin regions had a higher contribution to the genetic variance (> 45%), while variants in coding regions had the strongest individual effects (> 25-fold enrichment on average). Conversely, variants in intergenic or intronic regions showed lower levels of enrichment (0.2 and 0.6-fold on average, respectively). Conclusions. Heritability partitioning approaches should be used cautiously in livestock populations, in particular for small categories. Two-component approaches that fit only one functional category at a time lead to biased estimators and should not be used.

UPDATED CATTLEGTEX RESOURCE FOR DISSECTING COMPLEX TRAITS IN CATTLE

Zhang, Huicong; Li, Houcheng; Cai, Zexi; Sahana, Goutam; Sørensen, Peter; Fang, Lingzhao

Center for Quantitative Genetics and Genomics, Aarhus University

Genome-Wide Association Studies (GWAS) have identified lots of loci that contribute to phenotypic variation in both farm animals and humans. However, dissecting the molecular mechanisms underlying such variants can be extremely difficult. In human genetics, projects such as the Genotype-Tissue Expression (GTEx) project have discovered lots of functional variants regulating the transcriptome, which contributes to understanding of regulatory mechanisms behind complex traits and diseases. While in livestock, the characterization of these regulatory variants remains limited. One of the most effective strategies to bridge this gap is molecular quantitative trait locus (molQTL) mapping in natural populations. Inspired by the human GTEx project, the Farm Animal Genotype-Tissue Expression (FarmGTEx) project was established to build a comprehensive public resource for regulatory variants discovery and molecular phenotype prediction in farmed species. Previously, the pilot phase CattleGTEx project has described the genetic regulatory mechanisms across 23 tissues. However, the resource also exhibits several limitations in number of molecular phenotypes, sample size and tissue types, thereby hindering the detection of tissue-specific regulatory patterns of small effect size. Moreover, the influences of contextual factors such as sex, age, and cell type are not well elucidated. The overall aim of our study is to fully develop the CattleGTEx phase 1 project. The workflow of CattleGTEx data analysis was improved, taking additional molecular phenotypes together with the contextual factor effects into consideration, and organized into a Nextflow-based pipeline, which can be readily shared in the community. Furthermore, the large-scale public cattle RNA-seq was analyzed using this pipeline to build a new atlas of regulatory variants in cattle, bringing the total sample size to 19,000, representing over 40 tissue types. To impute genotypes from RNA-Seq data, we newly built the multi-breed genotype imputation panel using over 3.000 publicly available whole genome sequencing data in cattle. These updated resources will contribute to illustrating genetic and molecular mechanisms underlying a wide range of complex phenotypes in cattle.

XWAS USING MULTI-GENERATIONAL DATA FROM CROSSBRED BOS INDICUS-BOS TAURUS CATTLE

Zhang, Nan; Gill, Clare A.; Riggs, Penny K.; Herring, Andy D.; Sanders, Jim O.; Riley, David G.

Texas A&M University Department of Animal Science

In modern beef cattle production, birth weight and weaning weight are two heavily studied beef cattle production traits because they are both good economic indicators. Birth weight (or prenatal growth) is inherited in an unusual way in Bos indicus-Bos taurus crosses. Bos indicus sired calves out of Bos taurus

dams have heavier average birth weight and larger sex differences when compared to those sired by Bos taurus and out of Bos indicus dams. Heavy birth weights often result in dystocia and cause economic loss due to death or injury of cows and/or their progeny. Previous genome-wide association studies have identified quantitative trait loci on the autosomes that may influence birth weight and weaning weight. In spite of its large size (6% in females), the associations of markers on bovine chromosome X (BTA X) with calf birth and weaning weight have not been extensively investigated. We conducted a X chromosome-wide association study (XWAS) in a multi-generation, closed herd of Bos taurus indicus x Bos taurus taurus crossbred cattle. For birth weight analysis, we identified 28 significant SNP in the pseudoautosomal region (PAR) and 9 significant SNP in the X-specific region in females ($P<0.05$ after Bonferroni correction for multiple testing). For weaning weight analysis, neither PAR nor female X-specific region returned significant marker-trait associations. Strong genomic inflation was observed for the X-specific region in males from this divergent cross, and we will report outcomes of our attempts to control for population stratification. Potential future adaptations include alternative parameterizations of marker effects and consideration of BTA X marker interactions with autosomal markers. These results highlight the importance of BTA X for birth weight as an indicator of prenatal growth and further refinement of valid analysis methodology for markers on this chromosome.

LEVERAGING A TRANSFORMER-BASED LARGE LANGUAGE MODEL ON QSAR MOLECULAR DATASETS

Zhang, Joia[1]; Martinez, Paola[2]; Gilman, Teddy[3]; Ancelin, Kason[4]; Zhu, Yanqiao[5]; Ponzoni, Luca[5]

[1]Cornell University,; [2]Universidad Nacional Autonoma de Mexico,; [3]Rice University,; [4]UCLA; [5]Relay Therapeutics, NSF and Relay Therapeutics funded project at the Institute of Pure and Applied Mathematics (IPAM) Research in Industrial Projects for Students (RIPS) at UCLA

Predicting biological functions solely from DNA sequences or molecular structures present a formidable challenge. However, computational methods utilizing textual representations of molecules in drug discovery may offer valuable insights into these complex pathways. Understanding the effects of chemical compounds on biological metrics is essential for drug discovery, yet the traditional approach of testing the biological effects of molecules is impractical, given its laborious, expensive, and time-consuming nature. In silico simulations present a promising avenue to narrow down potential targets for practical biological testing, offering a more efficient and cost-effective approach. IBM's transformer-based large language model, MoLFormer, has been trained on a vast repository of over one billion molecules represented in the textual format of simplified molecular input line entry systems (SMILES). Our study aimed to evaluate MoLFormer's performance on quantitative structure-activity relationship (QSAR) datasets, which provide structured information about chemical compounds and their biological activities. These datasets are instrumental in computational drug discovery and chemical biology, facilitating

the development of predictive models for compound optimization, virtual screening, and lead identification in pharmaceutical research. Our investigation delved into the performance of MoLFormer on phenomena such as activity cliff datasets, which represent subsets of QSAR datasets characterized by compounds with significant differences in biological activities or properties despite structural similarity. This discrepancy underscores crucial structure-activity relationships and unveils critical molecular features responsible for observed differences in biological activity, a property important for any computational method to capture. Additionally, we explored the utilization of symmetries in molecular representations to enhance MoLFormer's performance and learning, aiming to better interpret results and propose modifications to address common mistakes in the model's learning process. The implications of our findings for the drug discovery pipeline include potentially enhancing its efficiency and effectiveness. By leveraging advanced language models like MoLFormer and gaining a clearer understanding of its performance, we aim to facilitate more informed decision-making in drug discovery. Ultimately, our study offers insights into the capabilities and limitations of chemical large language models, providing guidance for their optimal use in pharmaceutical applications. Implementing our results could reduce the labor intensity and costs associated with the drug discovery process, while simultaneously expanding the computational toolkit available for evaluating the effects of molecules on quantitative biological metrics. Keywords: Large Language Model, MoLFormer, QSAR datasets, drug discovery

IDENTIFYING THE FULL DISEASE-RESISTANCE GENE REPERTOIRE USING POOLED HIFI SEQUENCING

Zhang, Zhiliang[1]; Xu, Song[2]; Xu, Jun[3]; Jiang, Liping[1]; Zhang, Jijin[2]; Kang, Lipeng[3]; Guo, Yafei[1]; Niu, Zelin[2]; Dong, Jiayu[3]; Song, Xinyue[1]; Qiu, Xuebing[2]; Wang, Jing[3]; Yin, Changbin[2]; Lu, Fei[3]

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.; [2]University of Chinese Academy of Sciences, Beijing, China.; [3]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.

The genomes of flowering plants have an extremely diverse and substantial gene family of NLR (nucleotide-binding leucine-rich repeat), playing a pivotal role in plant immunity. However, species-wide pan-NLRome in major crops was inadequately identified due to the intricate nature of NLR genes and the limitations of sequencing technologies. To better understand the architectural diversity of NLRome, we developed the pooled NLR population-scale HiFi sequencing (NLR-poolseq) to identify the full disease-resistance gene repertoire in plants. Also, we introduced the NLR Analysis Toolkit (NATK) for NLR discovery and genotyping, applying which in test data achieved a high performance with a precision of ~98.2% and a recall of ~91.9%, respectively. We futher developed the NLRome of rice from the USDA minicore population, including 108,372 NLR

genes, with 99.3% of singletons and 95.5% of paired genes, respectively. Furthermore, we identified known disease-resistance gene Pi5 through genome-wide association studies (GWAS), unveiling the functional haplotypes within the NLRome. These results demonstrate an ultralow-cost and effective method to profile the species-wide pan-NLRome to study plant environmental adaptation and enhance disease-resistant breeding.

POLYGENIC ADAPTATION TO GRADUALLY CHANGING ENVIRONMENTS

An important goal of evolutionary biology is to understand how populations adapt to new environments, especially in the context of global climate change. The adaptation of quantitative traits, controlled by a large number of loci with varying effects, is crucial for populations facing selective pressures. However, the precise role of the rate of environmental change in shaping evolutionary dynamics of quantitative traits remains unclear. To explore the impacts of trait architecture and environmental change rate, we employ forward-in-time simulations to study adaptive changes in quantitative traits under stabilizing selection subjected to continuous environmental shifts. Our findings indicate that despite varying evolutionary responses to selection, most trait architectures eventually reach a stationary state, while few lead to extinction. By examining this stationary state, we find that many summary statistics, including genetic variance and mean effect size of fixations, depend on the combination of the strength of stabilizing selection and the mutational effect size when environmental shifts are slow. Notably, as shift speeds approach an extinction threshold, the strength of stabilizing selection has only little effect. This leads to a similar adaptive pattern across multiple trait architectures. The effect size of new mutations and the rate of change appear to be the most important factors for successful adaptation to continuously changing environments. Taken together, our study highlights the diverse evolutionary trajectories of quantitative traits during environmental change and underscores the critical influence of environmental change rates on polygenic adaptation.

EXPLOITING MULTITISSUE MOLQTLS TO DECIPHER THE MOLECULAR MECHANISMS OF AGE-DEPENDENT GENETIC ARCHITECTURES FOR BODY WEIGHT CHANGES IN CHICKENS

Zhong, Conghao[1]; Li, Xiaochang[2]; Guan, Dailu[3]; Zhang, Boxuan[4]; Wang, Xiqiong[1]; Qu, Liang[2]; Zhou, Huaijun[3]; Fang, Lingzhao[4]; Sun, Congjiao[3]; Yang, Ning[4]

[1]State Key Laboratory of Animal Biotech Breeding and Frontier Science Center for Molecular Design Breeding, China Agricultural University, Beijing, 100193, China; [2]Department of Animal Science, University of California, Davis, CA, 95616, USA; [3]Jiangsu Institute of Poultry Science, Yangzhou, Jiangsu, 225125, China; [4]Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, 8000, Denmark

Introduction Body weight as a major growth trait has economic significance for farm-animals and is always the focus in breeding programs. For egg-type

chicken, body weight at sexual maturity is an important indicator in breeding programs for its influences on the onset of egg production. In addition, heavy body weight at late laying period, to a large extent, would decrease egg production performance. Hence, it is fascinating to accurately control or manipulate the body weight at each growth stage, and it will be a key event to understand the age-dependent genetic architecture of body weights across the whole life circle. Here, we provide guidelines for employing the ChickenGTEx resources to dissect regulatory mechanisms underlying genetic associations with growth traits. Materials and method An F2 resource population was derived from reciprocal crosses between White Leghorn (WL) and Dongxiang chickens (DX), a Chinese indigenous strain. In total, 1512 hens were chosen for SNP genotyping. In this study, we conducted GWAS analysis in the chicken population with longitudinal body weights at 30 age points from hatch to 72 weeks of old. The final ChickenGTEx data will release included 28 chicken tissues and 5 types of molecular quantitative loci (PCG expression, lncRNA expression, exon expression, splicing variation and 3'UTR alternative polyadenylation). We systematically integrated molQTLs with GWAS result of growth traits by applying four complementary methods, including fastENLOC, summary-data-based MR (SMR), single-tissue transcriptome-wide association study (sTWAS), and multi-tissue TWAS (mTWAS). Results and discussion By performing an integrative analysis of body weight at 30 age points with 5 types of molecular QTLs across 27 chicken tissues, we constructed a comprehensive gene atlas and enhanced the functional interpretation of body weight changes in chickens. We demonstrated that genetic determinants change with the growth phases of animal life cycles. The growth cycle of chickens can divide into three stages based on body weight. Genes expressed in various tissues play a role in weight changes during the three stages, with key influences observed on chromosomes 1, 4, and 27. Specifically, we found that the SLC25A30 gene in the hypothalamus and NEK3 in the retina play significant roles across all growth stages. Moreover, the ST3GAL4 gene in the intestine and CAB39L in adipose tissue contribute to weight regulation during the second stage (8 to 22 weeks), while body weight gain in the third stage (23 to 72 weeks) is mainly influenced by the CKAP2 and KAT7 genes in the ovary. Overall, our use of multidimensional molecular phenotypic data allows for a more comprehensive analysis of complex traits. These findings provide novel insights into the genetic and biological mechanisms underlying longitudinal changes in body weight, advancing the application of marker-assisted selection and genomic selection in future chicken breeding programs.

RNASEQ PROFILING AND MOLECULAR QTLS MAPPING FOR REPRODUCTIVE-RELATED TISSUES ACROSS VARIOUS EGG-LAYING STAGES IN CHICKENS.

Zhu, Di[1]; Shi, Kai[2]; Zhang, Ran[3]; Jiang, Ziqin[4]; Li, Houcheng[1]; Feng, Chungang[2]; Wang, Yuzhe[3]; Fang, Lingzhao[4]; Hu, Xiaoxiang[4]

[1]State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China; [2]College of Animal Science and Technology, Nanjing Agricultural University, Nanjing, China; [3]Center for

Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark; [4]National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangdong, China

Chickens play a pivotal role in global agriculture due to their contributions to meat and egg production. Understanding the cyclical nature of egg-laying is essential for optimizing production and breeding programs. In this context, our comprehensive study employs multi-omics analysis to dissect the regulatory genetic variations in five molecular phenotypes across different egg-laying stages in chickens. This research provides critical insights into the genetic controls of egg production, a key aspect of poultry science with significant economic implications. We collected 1,272 tissue samples from 359 hens across three critical egg-laying stages: the pre, peak, and late stage of egg production. Each phase included RNA-seq data of four key tissues: the hypothalamus, pituitary, ovaries, and liver. All 359 hens were subjected to whole-genome sequencing at a depth of over 20×. After quantified the gene expression, we identified 6,421 tissue-specific and 4,559 laying stage-specific expressed genes. Gene Ontology (GO) enrichment analysis showed that tissue-specific genes were significantly enriched in biological processes relevant to their tissue's functions, whereas laying stage-specific genes were significantly enriched in development and cell differentiation pathways. Notably, we found that genes involved in defending against microbial invasion were significantly high expressed in the ovary during the peak laying stage. Using co-expression network and time-course analysis, we found remarkable similarities in gene expression patterns among the reproductive relate tissues—hypothalamus, pituitary, and ovaries—across various egg-laying stages. We quantified the expression of protein-coding genes, long non-coding RNAs, exons, enhancers, and alternative splicing events and then performed molecular QTL mapping analysis. In total, 13,866 (85.5%) of 16,227 tested protein-coding genes (eQTLs), 8,393 (75.2%) of 11,156 lncRNAs (lncQTLs), 40,722 (38.8%) of 104,931 exons (exQTLs), 10,831 (52.1%) of 20,798 genes with alternative splicing events (sQTLs), and 8,214 (41.9%) of 19,570 expressed enhancers (enhQTLs) were significantly (FDR < 0.05) regulated by at least one genetic variant in at least one tissue/stage. Colocalization analysis confirmed the limited sharing of regulatory control among difference molecular phenotypes. Furthermore, we found that the majority of molQTLs were shared across different egg-laying stages, for instance, 69.29% of eQTLs were ubiquitous throughout the three egg-laying stages, in contrast, there is considerable variation in molQTLs among tissues; for example, 32.29% of eQTLs are tissue-specific. Our study revealed the expression patterns of genes in different tissues during various egg-laying stages in chickens, identified many tissue- and period-specific expressed genes, and associated molecular QTLs with five types of molecular phenotypes, providing strong data support for further analysis of the regulatory mechanisms behind complex traits in chickens.

WHAT DETERMINES LEVELS OF ADAPTIVE GENETIC DIVERSITY?

Zijmers, Lillith[1]; Abson, Katie[1]; Mittell, Lizy[2]; Hadfield, Jarrod[2]; Eyre-Walker, Adam[2]

[1]School of Life Sciences, University of Sussex, Brighton, BN 9QG; [2]Institute of Ecology and Evolution, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL.

A population's ability to adapt is determined by its levels of quantitative genetic variation (QGV), and while it is agreed that most organisms have QGV for most traits, the extent to which it varies between species is unknown. Previous studies have found low interspecific variation in both heritability and evolvability, and no factor has been shown to correlate to evolvability. In order, to investigate the matter further we compiled estimates of heritability and evolvability from more than 190 species. We aim to address several questions with our larger dataset using phylogenetically controlled general linear mixed models. We will test whether there is systematic variation in the level of evolvability between species and investigate whether evolvability is correlated to the mutation rate, the effective population size and a variety of life-history traits (e.g. longevity, body size). Preliminary results will be presented.

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26 Location

University of Vienna
Universitätsring 1, 1010 Vienna
AUSTRIA

Editor

Institute of Science and Technology Austria (ISTA)
Am Campus 1, 3400 Klosterneuburg
AUSTRIA

Responsible for content
Matthew Robinson, Ph.D.
Photos Gerhard Sengmüller /
https://layout.univie.ac.at/fotopool/fotos-von-gebaeuden/

Contact

Theresia Hammerl
Event Project Manager

Institute of Science and Technology Austria (ISTA)
Am Campus 1, 3400 Klosterneuburg
AUSTRIA
Telefon +43 664 883 264 84
icqg7@ista.ac.at