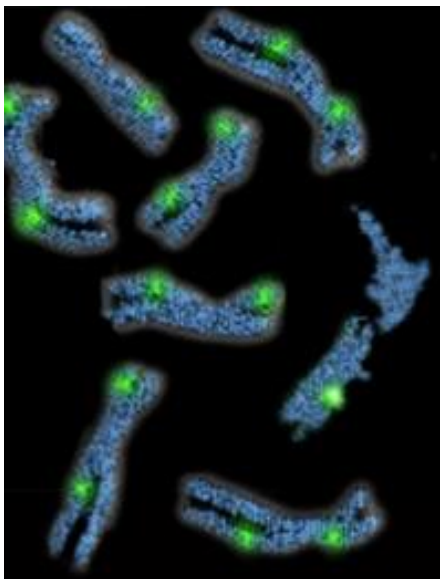
MapNet 2023 Program



Mihimihi

E nga mana, e nga reo, e nga karangataha maha, tēna koutou! Nau mai, haere mai ki te hui nei o MapNet mō te tau 2023! Ko te hui nei i Te konika o te Matamata. Ko Agresearch te whakahaere manaaki. Nōreira, e nga kairangahau kua tae mai, nau mai, mihi mai, karanga mai ki a koutou katoa.

Translation: All peoples, all languages, all authorities, greetings! Welcome to MapNet 2023 workshop, held in Te Konika o te Matamata (Mosgiel), hosted by AgResearch Ltd. So, to all researchers gathered here: greetings and welcome to you all!

About MapNet:

Established in 2005, MapNet is a semi-formal Aotearoa/NZ-wide collective of researchers in various areas of gene mapping and genetical genomics. The key purpose of MapNet is '**encourage pre-commercial, cross-sector, gene mapping related R&D in a range of organisms valuable to NZ's economy, ecology and culture**'. Participation in MapNet is voluntary, at an individual researcher level. All major sectors are typically represented in activities: health, environment, and primary sector (agriculture, horticulture, aquaculture, etc). These activities have included annual workshops that have typically been attended by 50-100 participants, as well as collaborative research activities such as the Virtual Institute of Statistical Genetics, which was funded 2008-2013 and is currently hosted by Genetics Otago. MapNet also holds one-off workshops and has undertaken various advocacy-type activities on behalf of the community of researchers involved in MapNet.

Organising Committee members

Rachael Ashby	Mik Black	Rebecca Clarke	Shannon Clarke
Ken Dodds	Jeanne Jacobs	Jie Kang	Maddi Post
Phillip Wilcox			

Locations

Wednesday's venue is the Auahi Ora room in the University Union, University of Otago.

Thursday/Friday venue is the Cullen Conference Room at AgResearch Invermay, Mosgiel.

Social Functions

Nibbles function on Wednesday evening from 4:30-6 is occurring at the same venue as the talks, in the Auahi Ora room in the University Union, Ground Floor, University of Otago.

Conference dinner on Thursday evening from 6:30-9 at Joe's Garage, 19 Frederick Street, North Dunedin.

Social Media

If you are using X (a.k.a. the social media platform formerly known as Twitter), tag your conference posts #mapn23

Please respect the requests of speakers and conference attendees that ask or suggest not to be included in social media posts.

MapNet Website

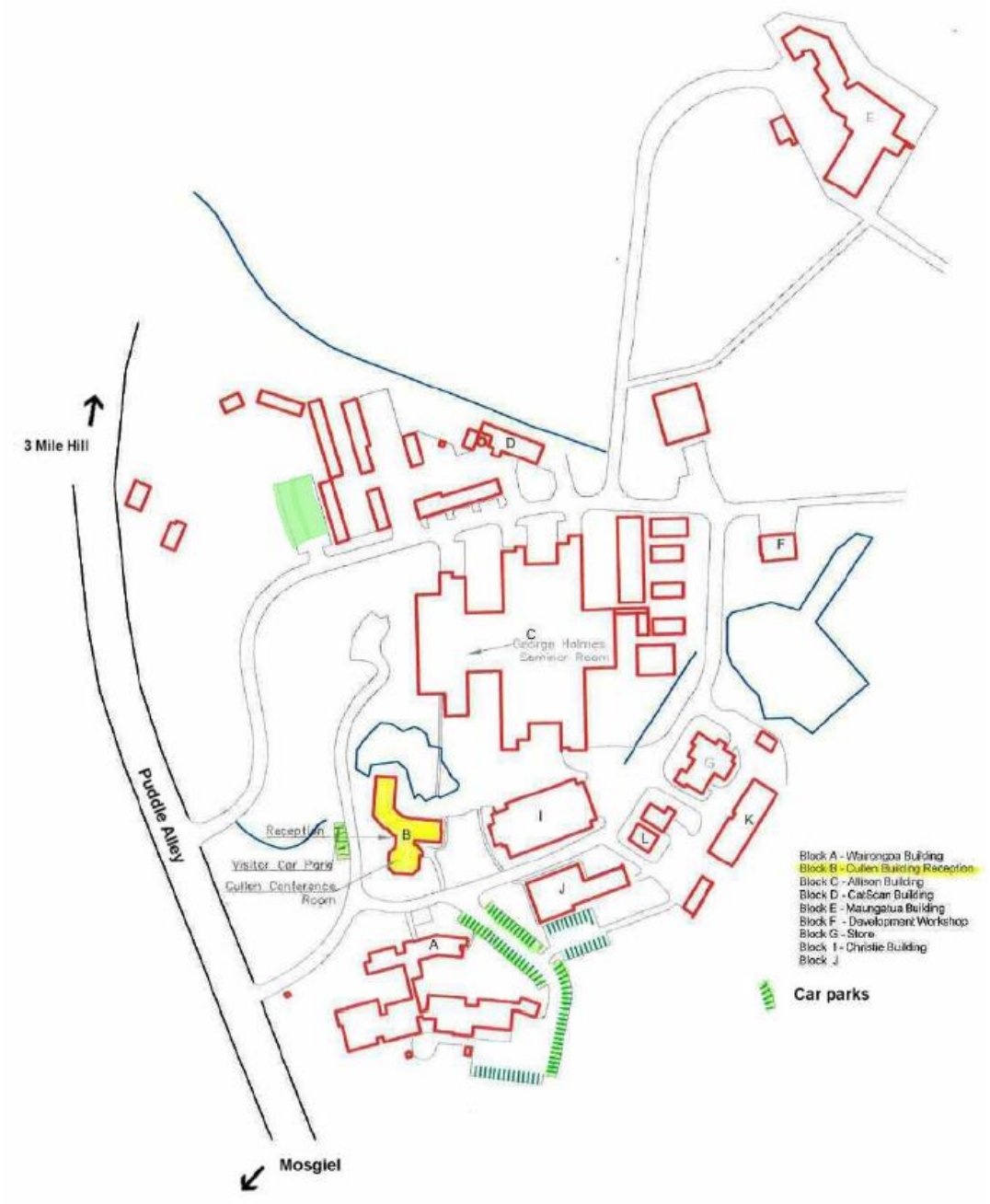
MapNet website is located at https://mapnet2023.github.io/

Code of conduct

We are dedicated to providing a harassment-free meeting for everyone, all attendees, speakers, and sponsors are required to abide by the code of conduct located on the MapNet website. Organisers will enforce the code of conduct and expect cooperation to ensure a safe environment for all.

Otago University Venue Map

AgResearch Venue and Car Parking Map



Wednesday 22nd November	
GA & MapNet shared session	
Session Chair: Mik Black	
12:30pm	**Registration Desk Opens,** please be seated 1:25pm at the latest
1:30pm	**Keynote:** Elinor Karlsson (Broad Institute of MIT and Harvard) T*he Future of Comparative Genomics: Finding Meaning in DNA Sequence in a Million Genome Age*
2:10pm	**Keynote:** Alison Van Eenennaam (University of California) *Global Status of Gene Edited Food Animals and their Products*
2:50pm	Joseph Guhlin (University of Otago) *Exploring the NZ Bee PanGenome*
3:10pm	**Afternoon tea**
3:30pm	**Keynote:** Suzanne Rowe (AgResearch) *Using Molecular Phenotypes to Lower Global Methane Emissions*
4:00pm	Kimberley Dainty (University of Otago) *Predicting Successful Biocontrol; What We Can Learn from Microctonus hyperodae Genomics*
4:20pm	Karakia Whakamutunga
4:30pm	Social Function

Thursday 23rd November	
8:30am	Karakia and Welcome
Illumina Session	
Session Chair: Rebecca Clarke	
8:45am	John McEwan (AgResearch) *Genomic Improvement in Dairy Goats: How Can We Start Immediately?*
9:05am	Chris Couldrey (LIC) *Got Milk? Got Microbes!*
9:25am	Timothy Bilton (AgResearch) *Profiling the Rumen Microbiome in New Zealand for Potential Application in Ruminants*
9:45am	Paul Johnston (Plant and Food) *Reducing Gluten Allergenicity through Wheat Breeding*
10:05am	**Morning Tea – Decode Science Sponsored**
Primary Production	
Session Chair: Monica Vallender	
10:30am	Megan Scholtens (Cawthron Institute) *Current Status of Genomics in New Zealand Aquaculture*
10:50am	Jane Symonds (Cawthron Institute) *Climate Adapted King Salmon*
11:10am	Josie Athens (AgResearch) *Multi-omics Analysis of Lactobacillus rhamnosus Growth*
11:30am	Pavithra Ariyarathne (AgResearch) *Heritabilities for Tail and Bareness Traits Estimated Based on Genomic Relationships in New Zealand Sheep*
11:50am	Jordan Clarke (AgResearch) *Investigating Rumen Microbial Communities in New Zealand Farmed Deer*
12:10pm	Satish Kumar (Plant and Food) *Genetics and Genomics of Apple Phytochemicals*
12:30pm	**Lunch – GenomNZ Sponsored**

Statistical Methods and Bioinformatics	
Session Chair: Ben Perry	
1:30pm	**Keynote:** Vinzent Boerner (GHPC Consulting and Services Pty Ltd) *Past, Present and Future of High Performance Computing in Animal Breeding*
2:10pm	Jie Kang (Beef and Lamb New Zealand) *Enhancing Genomic Selection through the Integration of SNPs and Haplotypes in Single-Step Hybrid Genomic Best Linear Unbiased Prediction (ssHGBLUP)*
2:30pm	Michael Lee (University of Otago) *Genomic Predictions for a Large Complex Population: New Zealand Sheep*
2:50pm	Roy Costilla (AgResearch) *Estimating Genotype-by-Environment Interactions using Genomic Data: An Application to Smallholder Dairy Farms in India*
3:10pm	**Afternoon tea – Decode Science Sponsored**
Session Chair: Rachael Ashby	
3:30pm	Marcus Davy (Plant and Food) *Barcode Metadata Considerations*
3:50pm	Charles Hefer (AgResearch) *Analytical and Sequencing Approaches to Improve the Quality of a Heterozygous Ryegrass Genome*
4:10pm	Setegn Alemu (AgResearch) *Streamlining Methane Emission Data Management: A Comprehensive Shiny Application for PAC Data in Sheep Breeding*
4:30pm	Annabel Whibley (Bragato Research Institute) *Characterising Clonal Genomic and Epigenomic Variation at Scale in Grapevine*
4:50pm	Ken Dodds (AgResearch) *Correcting Duplication Effects in Sequencing-based Genotypes*
5:10pm	Finish
6:30pm	Dinner – Joe's Garage

Friday 24th November	
8:55am	Welcome
Indigenous Science	
Session Chair: Madeleine Post	
9:00am	**Keynote:** Phil Wilcox (University of Otago) *Development of Genomic Resources for Māori Health*
9:40am	Alastair Lamont (University of Otago) *Development of a Population Simulator to Optimise Study Designs and Estimate Polygenic Disease Risk/Resilience in Aotearoa New Zealand Māori Populations*
10:00am	Anna Edwards (AgResearch) *Indigenous Ethics in Research: The Protection and Correct Treatment of Taonga Species*
10:20am	**Morning Tea – Decode Science Sponsored**
New Technologies	
Session Chair: Jie Kang	
10:50am	Astra Heywood (Plant and Food) *Asymmetrical Inheritance in Honeybees: A Polyhybrid Species*

11:10am	Guillermo Rodriguez Piccoli (Victoria University of Wellington) *Novel Tools for Omics Data Analysis in Pseudo-haploid Species*
11:30am	Ben Perry (AgResearch) *TaFFE: A Snakemake Workflow for Taxonomic and Functional Analysis of Meta-Omics Data Derived from Sheep Rumen Microbiomes Divergently Selected for Methane Emissions*
11:50am	**MapNet Discussion: Phil Wilcox**
12:50pm	**Closing remarks and Karakia**

1. The Future of Comparative Genomics: Finding Meaning in DNA Sequence in a Million Genome Age

Karlsson, E[1].

[1]Broad Institute of MIT and Harvard

The rapid advancement of genomic technology has resulted in an explosion of genomic data. Zoonomia is one of the largest comparative genomics resources produced to date. By comparing the genomes of 240 different mammal species, we harness the power of 100 million years of evolution to explore genome function. We can predict which bases, when changed, are most likely to impact health. We can find changes in genes and regulatory elements that are associated with exceptional traits such as hibernation, and changes that are specific to humans. Earth's vast and imperilled biodiversity offers unique power for advancing biological discovery, and, through collaborations with ecologists and conservation scientists, can also help protect that biodiversity.

2. Global Status of Gene Edited Food Animals and their Products

Ledesma, A.V.[1], Van Eenennaam, A.L.[1]

[1]Department of Animal Science, University of California, Davis

A search using Gene Editing in the Medical Subject Heading (Mesh), or gene edit*, or genome edit* or base edit* in the title or abstract and targeted to agricultural animals was performed in PubMed on July 21, 2023; and resulted in over 1,200 publications. After reviewing each publication, those that were exclusively for biomedical purposes, or where the edits were performed only in cells, or embryos that did not result in a live animals, or where edited animals did not survive beyond birth were excluded. The remaining publications were categorized by editing system, species, purpose, type of edit (SDN-1,2,3), & country of first author. A total of 195 applications were identified in the peer-reviewed literature. Fifty-nine percent were targeting food mammals, followed by 29% aquatic species, 8% avian species, and 4% insects (honeybee, and silkworm). The most common trait category targeted was meat and fiber yield, followed by reproduction, biotic stress, production of hypoallergenic products, multiple traits, color, product quality, abiotic stress, and other. The large number of applications focused on reproduction is due in part to the interest in both single-sex offspring in numerous industries, and also research into surrogate broodstock and germline complementation in mammalian, avian and aquatic species. The majority (147) of the applications were SDN-1, with 18 SDN-2, and 30 SDN-3 applications involving a repair template. Likewise, the majority were CRISPR/Cas9, followed by TALENs, ZFNs, DNA base editor, CRISPR/Cas12a, and other systems. There were 59 applications (30%) where the editing was done in cell lines followed by cloning to produce an animal, all in mammals; 118 publications (61%) that edited developing embryos, and 18 “other” approaches (9%) to editing, the majority of which were publications with avian species where editing was done in primordial germ cells. The majority of the 195 papers had first authors based in China, followed by the United States, Japan, South Korea, Norway, and the United Kingdom. To date, several developed or ongoing research applications have been authorized for commerce, or judged to be “non-GMO” hence conventional in at least one country, including knockout tiger pufferfish and red sea bream in Japan; tilapia, cattle, pigs and horses in Argentina; cattle and tilapia in Brazil, and two knockout PRLR-SLICK cattle were given enforcement discretion in the United States. One application, the CD163 deletion PRRSV resistant pig lines is formally in the precommercial stage, and there are proof-of-concept applications in multiple food species testing gene targets for traits of commercial interest.

3. Exploring the NZ Bee PanGenome

Guhlin, J.[1], Dearden, P. K.[1]

[1]Genomics Aotearoa, Department of Biochemistry, University of Otago, Dunedin, NZ,

The honeybee (*Apis mellifera*) produces honey and beeswax, as well as playing a critical role in crop pollination, generating $NZ 449M in hive produces and underpinning pollination-requiring industries worth $NZ 6.7B towards New Zealand's economy annually. Bees are also eusocial haplodiploid insects, providing unique opportunities to explore these understudied life history traits.. Increasing productivity and disease resistance through selective breeding requires understanding the vital genetic fundamentals at play and a comprehensive overview of the variation underlying the variation of these phenotypes found within the population. An *Apis mellifera* pangenome has been created using nanopore sequencing of 16 haploid males, combined with the chromosome-length high-quality reference genome. This pangenome provides a better view of population-wide variation, including larger indels and structural variants, gene presence and absence, copy number variation, and haplotypes. This pangenome has 23 million nodes, 33 million edges, and a total length of 250 Mbp, 25 Mbp larger than the published reference genome, with 3.2 million indels, 581 thousand MNPs, and 3.73 million SNPs. The pangenome also identifies an additional 42,454 indels fixed in our population. This new resource will allow us to identify novel sequence variation, identify constrained and fast evolving genes, explore how NZ-specific genomes are changing, and support future work in genotyping larger populations using short-read sequencing. This work serves as an example of creating and working with pangenomes in haplodiploid species which can be applied to species for agricultural outcomes.

4. Using Molecular Phenotypes to Lower Global Methane Emissions

Rowe, S.J.[1], Bilton, T.P.[1], McRae, K.M.[1], Perry, B.[1], Jonker, A.J. [2], Janssen, P.H.[2], Henry H.M.[1], Hickey, S.M.[3], Johnson, P.L.[1], McEwan, J.C.[1]

[1]Invermay Agricultural Centre, AgResearch, Mosgiel, NZ, [2]Grasslands Research Centre, AgResearch, Palmerston North, NZ, [3]Ruakura Research Centre, AgResearch, Hamilton, NZ.

Microbial fermentation of feed in ruminant livestock accounts for more than 11% of all global anthropogenic greenhouse gas emissions. We have shown that individuals vary in the composition of the microbial communities that they harbour, and so the dominant fermentation pathways and the resultant amount of methane also vary. These pathways in turn, control energy sources for the animal and have a downstream effect on metabolic processes such as tissue deposition and milk composition. Genetic selection based on these variations has been shown to be an effective mitigation strategy for reducing enteric emissions but is limited by availability of high throughput phenotypes. All of these processes offer candidates for molecular phenotypes that have potential as predictors of methane emissions and other feed related traits. We describe a program of work to determine the impact of genetic selection for methane emissions, correlated changes in downstream metabolites, and discuss the utility of a range of biological and spectral predictors of methane emissions in livestock breeding schemes.

5. Predicting Successful Biocontrol; What We Can Learn from *Microctonus hyperodae* Genomics

Dainty, K.R.[1,2], Inwood, S.[1,2,3], Goldson, S.L.[4], Dearden, P.K[1,2,3]

[1]Biochemistry Department, University of Otago, Dunedin, Aotearoa, New Zealand, [2]Genomics Aotearoa, University of Otago, Dunedin, Aotearoa, New Zealand, [3]Bioprotection Aotearoa, Lincoln University, New Zealand, [4]AgResearch Ltd, Weeds, Pests and Biosecurity Group, Lincoln, New Zealand

The insect pasture pest, the Argentine stem weevil (*Listronotus bonariensis*, ASW), causes an estimated NZD$200M of damage to New Zealand pasture annually. In the 1990's the endoparasitoid wasp, *Microctonus hyperodae,* was collected from eight South American locations and released throughout NZ as a biocontrol agent against ASW. Initially, parasitism rates of ASW by *M. hyperodae* were as high as 80%, effectively reducing ASW population sizes and resultant pasture damage. Despite this initial success, parasitism rates have now declined by over 50% in multiple locations, resulting in increased ASW populations and return of pasture damage. This biocontrol decline may be a result of ASW evolving parasitism resistance, against which the asexual *M. hyperodae* was unable to coevolve. Alternatively, *M. hyperodae* may have become less parasitic of ASW by way of adaptation to novel New Zealand environments.

Here, for the first time, we were able to understand the population relatedness of the eight ecotypes of *M. hyperodae* at the time of deployment throughout New Zealand. Our data suggests that, despite their differing geographical origins and previous identified phenotypic variation, there is very little genetic difference amongst and between the eight ecotypes. This lack of genetic diversity may be due to the asexual reproduction of *M. hyperodae,* or the need for tight regulation of genomes in a species which occupies niche environments.

We next compared these eight historical populations to recent collections of *M. hyperodae* in New Zealand, in order to understand how *M. hyperodae* have been selected for and evolved under the pressures of novel New Zealand environments. Again, preliminary results indicate little genetic differences between the historical and recently collected *M. hyperodae* genomes. This work critically informs upon the population structure of a declining New Zealand biocontrol agent and provides insights for considerations of future biocontrol efforts in New Zealand.

6. Genomic Improvement in Dairy Goats: How Can We Start Immediately?

McEwan, J.C.[1], Wheeler, M.[2], Brauning, R.[1], Van Stijn, T.C.[1], Baird, H.J.[1], Anderson, R.M.[1], Foote, B.J.[3], Foote, J.F.[3], Cameron, A.W.N.[4], Blichfeldt, T.[5], Jakobsen, J.[5], Dodds, K.G.[1], Caulton, A.[1], Clarke, S.M.[1]

[1] AgResearch, Invermay Agricultural Centre, Mosgiel, Otago, New Zealand, [2] AgResearch, Ruakura Research Centre, Hamilton, New Zealand, [3] 859 Russell Road, RD4, Hikurangi 0184, New Zealand, [4] Cameron Road, Meredith, Victoria 3333, Australia, [5] NSG, Box 104, 1431 Ås, Norway

In dairy goats, industry uptake of genomic technologies has been slow due to the small size of the industries coupled with the limited market for SNP array-based technologies. However, the potential benefits of genomic selection in dairy goats are large, because key traits are sex limited, recorded post selection and pedigree recording in large dairy goat herds is problematic. This has led our laboratory to utilise two low-cost genotyping strategies based on sequencing: RE-RRS and GT-seq. Historically, adoption of these approaches has been slow, due to missing and probabilistic genotyping calls. This made them difficult to integrate with existing genetic evaluation software. The sequencing-based technologies described above are currently used for separate genetic evaluations in Australia, New Zealand and Norway, with more than 87,000 samples genotyped to date using RE-RRS with more than 56K SNPs reported. Evaluation methodology and results for an example herd genetic trend will be presented, which for a 290 lactation length lactation show improvements of 26 litres/doe/year and 4.51kg cheese/doe/year over the last 5 year period.

7. Got Milk? Got Microbes!

Couldrey, C[1].

[1]Research and Development, Livestock Improvement Corporation, Hamilton, New Zealand

With over 10 million herd test samples coming through LICs doors every year, we are working towards understanding the milk microbiome and using such information to monitor individual cow and herd health. Research towards the ultimate goal of proactive animal health monitoring with minimal effort on farm has begun with shotgun sequencing of over 9000 herd test samples and 1300 vat milk samples from over 300 farms across New Zealand. We have compared shotgun and targeted sequence data with traditional bacteriology and PCR-based testing. DNA sequencing results correlate well with these traditional testing methods while also allowing increased sensitivity and specificity.

With over 4000 species identified to date, this presentation will address how we are undertaking analysis, addressing the common challenges in understanding microbiomes and the unique challenges of working with milk samples in order to turn DNA sequence data into a meaningful understanding of how the milk microbiome varies by region, season, and farm management, and what this could mean in terms of animal health for the farmer. While this research is still in infancy, high throughput DNA sequencing to understand the milk microbiome appears to be a promising approach for revolutionizing the approach to animal health in the dairy industry.

8. Profiling the Rumen Microbiome in New Zealand for Potential Application in Ruminants

Bilton, T.P.[1], Perry, B.J.[1], Alemu, S.[1], McRae, K.[1], Henry, H.M.[1], Janssen, P.H.[2], Hess, M.K.[4], Hickey, S.M.[3], Bain, W. [1], Pile, G.[1], Baird, H.[1], French, M.[1], Ferguson, S.[1], Vallender, M.[1], van Stijn, T.C.[1], Sandoval, E.[2], Searle, A.[1], Bryson, B.[5], Veenvliet, B. [1], Watson, T.[3], Peers-Adams, J.[1], Booker, F.[1], Dodds, K.G.[1], McEwan, J.C.[1], Johnson, P.L.[1], Rowe, S.J.[1]

[1]Invermay Agricultural Centre, AgResearch, Mosgiel, NZ, [2]Grasslands Research Centre, AgResearch, Palmerston North, NZ, [3]Ruakura Research Centre, AgResearch, Hamilton, NZ, [4]University of Nevada, Reno, Nevada Bioinformatics Center, Reno, USA, [5]Woodlands Research Farm, AgResearch, Invercargill, NZ

The conversion of feed in ruminant animals is driven by microbial fermentation in the rumen. This produces a mix of volatile fatty acids (VFA) that are a major energy source for the animal. Changes in rumen microbial composition can affect the composition of the VFA mix, and therefore affect the overall performance and health of the animal. The relationship between rumen microbiome community (RMC) profiles and livestock traits have been previously investigated and associations with methane, performance and feed intake traits identified. However, the rumen microbial composition is highly variable across environments which may change the relationship of the RMC profiles with livestock traits. Here, we investigate the variation of RMC profiles of ruminants located across a diverse range of farms and flocks in New Zealand. Over 10,000 rumen samples from cattle, deer, goats, and sheep were collected from across New Zealand and sequenced using a restriction enzyme reduced representation sequencing approach to generate RMC profiles. Sequences were classified with the Genome Taxonomy Database (GTDB) using the GBS-TaFFE (https://github.com/BenjaminJPerry/GBS-TaFFE) pipeline to determine microbial taxonomy. Results suggest that feed followed by species are the main sources of variation in the RMC, although the top 10 most abundant genera was consistent across these variables. Nevertheless, RMC profiles collected on the same type of ruminants grazing similar feeds were found to be similar, even though samples were collected at geographically distant farms and different seasons and years. Associations between direct measurements and a proxy trait predicted from RMC profiles for methane emissions and feed efficiency was also examined, where moderate-to-high genetic correlations and moderate phenotypic correlations were observed on a subset of the samples collected on sheep. More importantly, we found that RMC profiles could be used as a proxy trait for methane and feed efficiency in selective breeding programs.

9. Reducing Gluten Allergenicity through Wheat Breeding

Macalister, J.[1] and Johnston, P.A.[1]

[1]The New Zealand Institute for Plant & Food Research

Several immune–related disorders prevent people enjoying or even consuming wheat–based food products. Gluten epitopes are short peptide sequences (13–33 amino acids) that can trigger an inflammatory response in susceptible consumers and in some people, exposure can result in celiac disease and a life–long avoidance of gluten. Gluten itself is a complex matrix made up of proteins from several large gene families with many allelic variants that differ in their abundance of these immunogenic epitopes. Gluten epitopes occur primarily in the alpha gliadin fraction of gluten and as such are an important part of the functional properties of gluten in baking. In this study, a genomic selection approach was used to evaluate whether wheat varieties with reduced gluten allergenicity could be developed whilst still retaining all the functional requirements for producing high quality food products.

A wheat training population was assembled from breeding and germplasm resources at Plant & Food Research and genotyped using the Illumina 90K wheat single nucleotide polymorphism chip. Gluten epitope concentrations were measured by liquid chromatography mass spectroscopy for six specific epitopes from seed harvested over three field seasons and baking quality metrics were measured from one field season. The data from this training population was then used to develop statistical models for the prediction of gluten epitope concentration in a related validation population.

All the epitopes measured had at least a three–fold variation in concentration across the training population and heritability estimates ranged between 0.44 and 0.75. Most baking quality metrics had only low–level correlations with epitope concentrations. However, a moderate to strong correlation was identified between one epitope (P4) and flour protein content (a key baking quality metric). Final realised genomic prediction accuracies ranged from 0.22 to 0.82 depending on the specific epitope and the type of analysis performed.

10. Current Status of Genomics in New Zealand Aquaculture

Scholtens, M.R.[1], King, N. [1], Symonds, J.E.[1]

[1]Aquaculture Group, Cawthron Institute, Private Bag 2, Nelson 7042, NZ

Genomics has the potential to revolutionize genetic improvement programs and production systems within New Zealand's aquaculture industry. Leveraging realized genomic relationships and genome-wide markers linked to important traits, genomic selection enables accurate prediction of complex polygenic traits, while simultaneously maximizing genetic gain and minimizing inbreeding. Genomics can help the industry to gain greater understanding of challenging traits such as feed conversion efficiency, disease resistance and environmental tolerance, which are often hard to measure.

Aquatic species remain in the early stages of domestication which is advantageous as it is often linked to higher within-species genetic diversity. In addition, the fecund nature of aquatic species, often undergoing external fertilization, provides flexibility in breeding program design and widespread dissemination of selectively bred strains without the need for extensive multiplication tiers.

Despite considerable investment in aquaculture genetics and technologies, there are limited off-the-shelf commercially available arrays for aquatic species farmed in New Zealand. A genomic approach that is more widely used is genotyping-by-sequencing (GBS). This is a low-cost, high-through-put genotyping method for non-model species which is proving invaluable for species lacking reference genomes or access to commercial genotyping arrays, and as such, is rapidly becoming one of the genomic tools of choice for aquaculture species in New Zealand. We will discuss the technical advances and commercial applications that have been made in the genomic space across various aquaculture industries in New Zealand with particular focus on the Pacific oyster (*Crassostrea gigas*), Greenshell mussel (*Perna canaliculus*) and Chinook salmon (*Oncorhynchus tshawytscha*).

11. Climate Adapted King Salmon

Symonds, J.E.[1], Walker, S.P.[1], Waddington, Z.[2], Mota-Velasco J.C.[3], Diaz Gil, C.[3], Looseley, M.[3], Dodds, K.G.[4], Clarke. S.M.[4]

[1]Aquaculture Group, Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand, [2]The New Zealand King Salmon Co. Ltd, Nelson 7010, New Zealand, [3]Xelect, Horizon House, Abbey Walk, St Andrews, Fife, KY16 9LB, Scotland, UK, [4]AgResearch, Invermay Agricultural Centre, Puddle Alley, Mosgiel 9053, New Zealand

The New Zealand aquaculture industry is vulnerable to climate change and is already impacted by marine heatwaves, causing increased summer mortalities, including farmed king (Chinook) salmon in the Marlborough Sounds in 2022. More extreme storm events also have impact, including damage to infrastructure with increased flooding, freshwater runoff and sediment accumulation. To overcome these challenges, the industry recognises they need to develop and implement climate change adaptation plans, including new husbandry practices, new farm locations (e.g., open ocean) and future adaptive breeding strategies that look beyond business as usual.

As part of the pathway towards producing climate adapted king salmon, a tank based temperature challenge model has been established. To date this model has successfully tested over 230 pedigree king salmon families. Heritabilities for time to death at elevated temperature were high (0.34 to 0.48) suggesting that selection for improved thermotolerance is possible.

The next stage is to develop an improved phenomics to genomics pipeline for king salmon to enable genomic selection and gene discovery for thermotolerance and other resilience traits in king salmon. Building on the genotyping-by-sequencing pipeline already established by GenomNZ, a new MBIE Endeavour research programme “Fast-tacking climate adaption in finfish” will generate a functionally annotated reference genome for king salmon, alongside transcriptomics and methylomics resources. These multi-omics resources will be applied to analyse multiple resilience phenotypes generated through commercial on-farm field studies and tank-based challenges for multiple environmental stressors. Ultimately salmon farmers will apply the genomic breeding values generated to produce climate adapted king salmon.

12. Multi-omics Analysis of *Lactobacillus rhamnosus* Growth

Athens, J.[1], Hefer, C.A.[2], Vignes, M.[3], Maes, E.[2], Villamizar, L.[2], Ross, A.[2], Fraser, K.[4]

[1] Systems Biology Enabling Platform, AgResearch, Invermay Agricultural Centre, Mosgiel, NZ, [2]AgResearch, Lincoln Research Centre, Mosgiel, NZ, [3]School of Mathematical and Computational Sciences, Massey University, Palmerston North, NZ, [4]Systems Biology Enabling Platform, AgResearch, Grasslands Research Centre, Palmerston North, NZ

The genus *Lactobacillus* is a heterogeneous group of bacteria consisting of many species associated with the fermentation of plants, milk, and meat. They are some of the most economically important food and feed biotechnology genera. We aimed to generate a high-quality, multi-omics dataset for an integrative analysis approach to characterise and understand the molecular changes during a growth curve of *Lactobacillus rhamnosus* GG (ATCC 53103).

Lactobacillus rhamnosus GG (ATCC 53103) was propagated at 37°C in MRS agar (Sigma-Aldrich) and incubated anaerobically at 37°C for two days. Single colonies were removed from the plates and transferred to a Hungate tube containing 10 ml MRS broth (SigmaAldrich), which was incubated at 37°C for 24 hours. Four samples were taken at times 1 hour, 6 hours, 12 hours and 24 hours; however, only the last three time points were used in the analysis. The following omics-related technologies were measured for each sample: transcriptomics (RNA-seq), proteomics and metabolomics. We estimated the median and relative dispersion (coefficient of variation) for each sample per sampling time. As a quality control, we removed samples whose relative dispersion was greater than a defined arbitrary threshold. Next, we transposed the matrices and interpolated values for 9 and 18 hours, i.e., five samples per transcript, protein or metabolite. We used two different approaches to identify the top molecules from each platform (transcriptomics, proteomics and metabolomics) relevant to the growth of *Lactobacillus rhamnosus*: integration of multi-omics data via sparse Generalised Canonical Correlation Analysis (sGCCA) and inference of phenomenological networks.

The top transcripts and proteins found in the analysis were chaperonin GroEL, phospo-carrier protein HPr, phosphor-glycerate kinase and pyruvate kinase. In the case of metabolites, we found a negative trend in the dynamics of carbohydrates and a positive trend in the dynamics of amino acids. We inferred phenomenological networks for each omics technology and their integration; for the latter case, glucose and galactose were the nodes with the higher degree and higher betweenness centrality.

The inference of phenomenological networks gave more consistent and biologically relevant results from all the integration technologies we used.

13. Heritabilities for Tail and Bareness Traits Estimated Based on Genomic Relationships in New Zealand Sheep

Ariyarathne, H.B.P.C.[1], Dodds, K.G.[2], Costilla, R.[3], Clarke, S.M.[2], McRae, K.M.[2], Johnson, P.L.[2].

[1]AgResearch, Grasslands, Palmerston North, New Zealand, [2]AgResearch, Invermay, Mosgiel, New Zealand, [3]AgResearch, Ruakura, Hamilton, New Zealand.

Improving welfare of animals is an important issue in the New Zealand sheep industry. Selection on traits such as tail length (TLEN), the amount of skin under the tail (TSKIN), bare breech (BBREECH), and bare belly (BBELLY) could improve the welfare of sheep including their resistance to disease. Recently, published literature based on pedigree relationships indicated that these traits are highly heritable. However, the use of genomic information will help to better understand the genetic architecture of these traits. In general, genomic relationships estimate heritabilities (h^2) more accurately in comparison to h^2 estimated using the pedigree relationships. Therefore, the objective of the current study was to estimate the h^2 of TLEN, TSKIN, BBREECH, and BBELLY traits with genomic relationships estimated using single nucleotide polymorphisms (SNP) chips.

We used data from sheep genotyped using five different low-density chips (four 18K and one 15K SNP chips) and imputed it to mid-density (60K SNPs) using Beagle (V 5.2) software. The average allelic R^2 of imputation over all five chips was 0.86 and SNPs that had allelic R^2<0.85 were discarded to improve allelic R^2 up to 0.96 with 55,004 SNPs remaining. These quality controlled, imputed genotypes were then used to estimate the genomic relationship matrix and genetic parameters using GCTA (V 1.94.1) software. The estimated h^2 of the traits ranged from high to moderate; TLEN (0.41±0.01, n=22,646), TSKIN (0.39±0.01, n=22,646), BBREECH (0.32±0.01, n=21,331), and BBELLY (0.26±0.01, n=13,531). These findings confirm that these traits are moderate to highly heritable using genomic relationships, and therefore, can be implemented in future genomic selection.

14. Investigating Rumen Microbial Communities in New Zealand Farmed Deer

Clarke, J.L.[1], [2], Bilton, T., Black, M. Henry, H., McRae, K., Booker, F., Veenvliet, B., Bain, W., Ward, J A, Thompson, B., McEwan, J.C., Rowe, S.J.

[1]AgResearch, Invermay, Mosgiel, NZ, [2]Department of Biochemistry, University of Otago, Dunedin, NZ.

Grazing ruminants are able to digest complex starches such as plant cell walls through a symbiotic relationship with microbes that inhabit the fore-stomach or 'rumen'. Fermentation pathways and the resulting energy sources and waste products differ depending on the microbial community in the gut. Although diet is the main driver, Rumen microbial communities (RMC) have been shown to be affected by the host genetics and to be heritable in sheep and cattle and are predictive of feed related traits such as methane emissions, a by-product of the fermentation process. The aim of this project was to determine the variability in RMC amongst individual deer and to estimate heritability of the deer RMC. The rumen content of 853 red and wapiti cross deer were sampled, and microbial DNA was sequenced using restriction enzymes reduced representation sequencing, which is a reduced sampling method that captures ~1% of the microbial genomes. A genome taxonomy database was used to determine taxonomy assignment. Sequences were used to determine similarity of RMC amongst individuals and presence of known microbes. Contemporary group, herd and farm were shown to significantly affect RMC composition. Similar to other ruminants, *Prevotella* overall was the most abundant genera present in the rumen. Similarly, *Sodaliphilus*, a genus linked to increased methane production is found in the top 10 most abundant microbial genera across the samples examined. The heritability of the RMC was estimated by using principal component analysis for dimension reduction. Heritability estimates were ~3% (+/-(6.150%). These were lower than reported in other species. Next steps are to determine whether there are components of the RMC that are unique to deer and associations with key production traits such as growth and methane emissions.

15. Genetics and Genomics of Apple Phytochemicals

Kumar, S.[1], Molloy, C.[1], Hunt, M.[3], Deng, C.H.[2], Wiedow, C.[3], Andre, C.[2], Dare, A.[2], McGhie, T.[3].

[1]The New Zealand Institute for Plant and Food Research Limited, Hawke's Bay Research Centre, Havelock North, NZ, [2]The New Zealand Institute for Plant and Food Research Limited, Mount Albert Research Centre, Auckland, NZ, [3]The New Zealand Institute for Plant and Food Research Limited, Palmerston North Research Centre, Palmerston North, New Zealand.

Breeding for phytochemicals in fruit crops is becoming an important goal. Understanding the genetic architecture of apple phytochemicals, and their genetic correlations with the fruit eating quality traits, is critical for the development of new apple cultivars with enhanced health benefits. A sample of 344 accessions, representing the domesticated and wild apple genepools, were genotyped using the Apple 20K SNP Array for a genome-wide marker-trait association study. Fruit samples were phenotyped for several targeted metabolites, including a stable vitamin C glycoside 'ascorbic acid 2-β-glucoside' (AA-2βG). Several fruit quality traits, including firmness, crispness, juiciness, and red skin over-colour, were also assessed. The average content of some metabolites (e.g., ascorbic acid, chlorogenic acid, phloridzin, and trilobatin) was at least 2-fold higher in wild accessions than domesticated genepool. Several new genomic regions of interest were identified, and the percentage of phenotypic variance explained by the best SNP ranged between 3% and 21% for the different metabolites – suggesting a complex polygenic control of most metabolites. Genomic heritability estimates suggested that most of the phytochemicals in this study have high heritability. Genetic correlations between phytochemicals and sensory traits were moderate. This study will assist in identifying candidate genes for functional genomics studies and help with the selection of accessions to establish genomics-based breeding strategies for the development of bio-fortified apple cultivars.

16. The Past, Present and Future of High Performance Computing in Animal Breeding

Boerner,V[1].

[1]GHPC CONSULTING & SERVICES PTY. LTD., Armidale, Australia

Since the introduction of quantitative genetics principles, animal breeding has always been at the forefront of high performance computing and it had to develop it's very own knowledge base for analysing big data. This resulted in specialists at the intersection of mathematics, statistics, animal breeding and computer science. The 70's and 80's were characterized by small computers with simple CPU designs, few specialised computing libraries, and scripting languages had yet to be developed. Huge volumes of in-house code were written in FORTRAN, covering even the most basic algorithms. Generally, people with a genetics background became self-trained computer scientists. In the 90's powerful scripting languages, third-party high-performance libraries for basic algorithms, and third-party high-performance linear model analysis software became available, which when combined were able to replace an entire in-house code ecosystem. This allowed geneticists to shift the focus of programming away from "algorithmic" towards "data manipulation", which has led to a general deterioration of knowledge about core programming, software-hardware interaction and efficient algorithm implementation, all critical ingredients to high-performance computer programs. Contrarily to that deterioration, programming linear model software has increased in complexity requiring the detailed understanding of core programming and algorithms from mathematics, statistics and quantitative genetics, which has resulted in a remarkable concentration of specialised knowledge in very few people worldwide, primarily located at universities. The exponential growth in data, combined with hardware developments (cache hierarchy, GPU, FGPA), and the replacement of the once almighty FORTRAN by C++ with it's mind-boggling programming techniques now requires a level of specialisation which is maybe be even out of reach for quantitative geneticists with an aptitude for computer programming. This may mean a move away from publicly funded software development at universities to businesses with a specialised mission to provide sophisticated and leading-edge computing solutions, and a further concentration of providers.

17. Enhancing Genomic Selection through the Integration of SNPs and Haplotypes in Single-Step Hybrid Genomic Best Linear Unbiased Prediction (ssHGBLUP)

Kang, J.[1], Zhang, X.[2]

[1]Beef and Lamb New Zealand, Dunedin, NZ, [2]AbasucsBio, Dunedin, NZ

Single Step Genomic Best Linear Unbiased Prediction (ssGBLUP) is a pivotal method in the realm of genomic selection, harmonising pedigree and genomic data (e.g., SNPs). Its applications extend to animal and plant breeding programs, facilitating accurate selection decisions.

Recent advancements in genomics led us to revisit the methodology of ssGBLUP. Meuwissen (2014) highlighted that haplotypes, unlike SNPs, capture younger genetic relationships, making it valuable resources for improving predictions. In situations where recording pedigree information is challenging, the substitution of pedigree data with haplotypes becomes a promising solution. Hence, we propose to integrate the historical relationships represented by SNPs with more recent relationships akin to pedigree, creating a unified framework called single step hybrid GBLUP (ssHGBLUP).

This presentation delves into the method and the performance evaluation of ssHGBLUP, comparing it to traditional SNP- and haplotype-based GBLUP methods. The results of this assessment demonstrate the enhanced accuracy of ssHGBLUP, highlighting its potential as a game-changer in the realm of genomic selection. By leveraging both SNP and haplotype information within a single-step framework, ssHGBLUP can pave the way for more precise and efficient breeding decisions, particularly in situations where pedigree data may be lacking or incomplete.

18. Genomic Predictions for a Large Complex Population: New Zealand Sheep

Lee, M.A.[1&2]

[1]Department of Mathematics and Statistics, University of Otago, Dunedin, NZ, [2]Beef+Lamb New Zealand Genetics, Queens Gardens Court, 3 Crawford Street, Dunedin 9016.

The aim of genetic evaluation is to use data collected from a population of individuals to predict genetic merit for a specific trait. The prediction is more-often a random effect from a best linear unbiased prediction (BLUP) model and called a breeding value. The breeding value is used to make selection decisions, where this forms the basis for genetic improvement via selective breeding in many commercial species.

In New Zealand, the sheep industry, via Beef + Lamb New Zealand Genetics, uses single step genomic BLUP (ssGBLUP) to predict breeding values. This allows both genotype and pedigree to be included and is the basis for genomic selection. Breeding values are reported on 49 different traits weekly. The models for each trait have been developed by many researchers and data recorded by numerous seed stock breeders mainly over the last two decades. Currently, there are about 12 million pedigreed animals in the evaluation of which about 400,000 have genotypes.

The New Zealand sheep population has a complex and heterogenous population structure. Data recording is sparse for many of the traits evaluated. This has created a number of challenges for its' evaluation via ssGBLUP. This paper will describe this population, some of the challenges encountered and the solutions to allow routine evaluations of this population.

19. Estimating Genotype-by-Environment Interactions using Genomic Data: An Application to Smallholder Dairy Farms in India

Costilla, R.[1], Al Kalaldeh, M. [2], Gaundare, Y. [3], Potdar, V. [3], Joshi, A. [3], Bhave, K. [3], Joshi, S. [3], Swaminathan, M. [3], Gibson, J. [2], Ducrocq, V. [4], Warburton, C[5]., Hayes, B.[5].

[1]AgResearch Limited, Ruakura, Hamilton, New Zealand, [2] Centre for Genetic Analysis and Applications, School of Environmental and Rural Science University of New England, Armidale, Australia, [3]BAIF Research Foundation, Pune, India, [4] Universite Paris-Saclay, INRAE, AgroParisTech, UMR GABI, 78350 Jouy-en-Josas, France,[5] Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia.

Genotype-by-environment interactions (GxE) are prevalent in most livestock industries. In smallholder dairy farms in India, this potential GxE is particularly important because of the huge variety of environments animals are raised in for these production systems. Here we aim to estimate GxE by state in India using simulations based on real genotyped data from crossbred dairy cows from the BAIF research foundation. In particular, we simulate phenotypes under the null hypothesis of perfect genomic linkage between environments (genetic correlation=1) and we use multi-trait models to estimate the resulting genetic correlations. In addition, we use the fixation index (Fst), to measure the genomic similarity among animals in different states. Using data from over 5,000 animals across five Indian states, we show that the estimated genetic correlations among states in the simulations are high (>=0.89) for all possible combinations of states, and thus provide evidence of enough genomic linkage/connectedness across the analysed states. Moreover, Fst values are also very low providing additional evidence of similar genetic backgrounds among animals in different states. These findings support the feasibility of estimating GxE by state in India using real milk yield phenotypic records.

20. Barcode Metadata Considerations

Davy, M.W.[1], Deng, C.H.[2], Hilario, E.[2], Kumar, S.[3], Gapper, N.E.[2]

[1]The New Zealand Institute for Plant and Food Research Limited, Te Puke, New Zealand, [2]The New Zealand Institute for Plant and Food Research Limited, Auckland, New Zealand, [3]The New Zealand Institute for Plant and Food Research Limited, Hawke's Bay, New Zealand

As sequencing technologies evolve, the quantity of data is increasing along with the number of samples researchers want to study in NGS experiments such as RNA-Seq studies.

This talk discusses considerations in DNA barcode design, and potential bias implications due to the misclassification of sequencing reads if compromised Barcodes are used with low edit distance in NGS experiments where demultiplexing of barcoded samples is required.

21. Analytical and Sequencing Approaches to Improve the Quality of a Heterozygous Ryegrass Genome

Hefer, C.A.[1], Weston, M.K[2] and Jacobs, J.M.E.[2]

[1]Digital Agriculture, AgResearch, Lincoln, NZ, [2] Resilient Agriculture, AgResearch, Lincoln, NZ

Lollium perenne L. (perennial ryegrass) is widely planted in the pastoral sector and serves as the main source of ruminant nutrition. Traditional breeding for desirable traits has led to distinct cultivars being developed in Aotearoa/New Zealand, compared to their European ancestors. We have previously sequenced and assembled the draft genome of one such cultivar, One50. The advent of third generation sequencing technologies has provided the opportunity to generate high-quality, long-read sequence data which can be utilized to improve historical genome assemblies predominately made from short-read data. We have generated high quality DNA and direct RNA sequencing libraries using Oxford Nanopore Technologies, and will discuss the analytical strategies employed to improve the assembly and annotation of the draft genome.

22. Streamlining Methane Emission Data Management: A Comprehensive Shiny Application for PAC Data in Sheep Breeding

Alemu, S.[1], Bilton, T, Bain, W. [1], Pile, G.[1], Neville, C. A[2], McEwan, J.C.[1], Johnson, P.L.[1], Roy, C [2], Rowe, S.J.[1]

[1]Invermay Agricultural Centre, AgResearch, Mosgiel, NZ, [2]Ruakura Research Centre, AgResearch, Hamilton, NZ

Portable Accumulation Chambers (PAC) offer a more affordable method compared to respiratory chambers for capturing methane emissions in livestock. However, the challenge lies in processing, cleaning, and organizing PAC data for integration into breeding programs. Addressing this, we present a Shiny application designed to efficiently manage methane data from PAC. Our Shiny App offers automatic data processing capabilities, allowing users to easily download cleaned data ready for further analysis. Additionally, it provides quality control tools, such as outlier detection, and interactive visualization capabilities for a comprehensive examination of the data across various variables. As the field of animal breeding evolves, with a special emphasis on genomic selection, there is a growing need for tools that can handle large datasets while also offering advanced analytical features. Our application, built on Shiny and enhanced with JavaScript, meets these needs. In its current iteration, the app excels at data input management and quality control. Our vision for its future includes the seamless integration of genetic statistical software, like ASReml, BLUPF90, and DMU, ensuring accurate breeding value estimation. We aim to deliver a platform that streamlines everything from data capture to quality checks, eventually preparing the data for genomic selection analysis. Once the data has been thoroughly analysed, we aim for our application to present results in an interactive format, ensuring clarity and accessibility for a diverse audience.

23. Characterising Clonal Genomic and Epigenomic Variation at Scale in Grapevine

Whibley, A.C.[1], Vanga, B.R.[1], Hill, A.[1], Wante S.[1], Liau, Y.[1], Carvajal, J. I. [2], Hilario, E. [2], Thompson, M. [3], Wang, L. [3], Fulton, B. [3], Barrell, P. [3], Bicknell, R. [3], Winefield, C. [4], Lizamore, D. [1].

[1]Grapevine Improvement, Bragato Research Institute, Lincoln, NZ, [2]Plant and Food Research Auckland, NZ, [3]Plant and Food Research, Lincoln, NZ, [4]Lincoln University, Lincoln, NZ.

Sauvignon Blanc is the dominant wine grape variety grown in Aotearoa New Zealand and accounts for around 85% of the export wine market. New Zealand Sauvignon Blanc vines are derived from a limited number of plant importation events and consequently exhibit very little clonal diversity. This genetic uniformity is a point of vulnerability as any new pest, disease or environmental challenge that negatively affects one Sauvignon Blanc vine could potentially affect every vine. Furthermore, conventional plant breeding approaches cannot be used to introduce genomic diversity since the products of any inter-crossing would no longer be classed as Sauvignon Blanc.

We have embarked on a Sauvignon Blanc improvement programme which harnesses endogenous transposable elements and their propensity to mobilise under conditions of temperature or chemical stress. With carefully calibrated treatments, we are generating a large clonal diversity population. This project will yield new insights into grapevine genome biology but also promises to generate tangible benefits to the wine industry by creating and characterising plants with traits that fall outside the spectrum currently available in Aotearoa New Zealand and building genomic resilience into our Sauvignon Blanc germplasm.

Alongside phenotypic evaluations, we will characterise the genomic variation in these clones and this work is underpinned by our Oxford Nanopore Technologies PromethION sequencing platform. The earliest stages of this project have involved building foundational genomic resources, including a high-quality phased diploid genome assembly. As we await the arrival of the first tranche of new clones, we've explored the genomic and epigenomic variation in a set of commercially available clones. I will present our learnings from this work and how they inform our approaches and the challenges we face as we move from genotyping a handful of grapevine plants to several thousand.

24. Correcting Duplication Effects in Sequencing-based Genotypes

Dodds, K.G., McCulloch, A.F., Brauning, R., van Stijn, T.C., Clarke, S.M.

AgResearch, Invermay Agricultural Centre, Puddle Alley, Mosgiel 9053, New Zealand

Most pipelines for calling genotypes or providing allelic counts from sequencing data assume that each of the possible alleles have been sampled at random. For a biallelic SNP, this corresponds to binomial sampling with the observed read depth as the sample size. However, the processes involved in generating the sequence reads can sometimes lead to duplication events whereby a particular sequence is copied and read multiple times. This leads to greater variation in allele counts than would be predicted by the binomial model, resulting in false inference of homozygous genotypes. A common practice with randomly sheared DNA fragments is to remove exact duplicate reads, but for restriction enzyme-based reduced representational sequencing (RE-RRS) some exact duplicate reads are expected. We investigate duplicate effects on an Illumina NovaSeq 6000 for RE-RRS. Duplication events specific to patterned flowcells result in duplicates tending to be spatially close to the original read, allowing for bioinformatic deduplication involving unsupervised spatial clustering of candidate duplicates. We have found that this may need to be complemented by using statistical models that allow for extra-binomial variation. Model parameters can be estimated using results from parents and their offspring or from multiple results on the same individual. Combining these two approaches supports suitable downstream analyses despite the presence of duplications in the sequencing results.

25. Development of Genomic Resources for Māori Health

Wilcox, P.L.[1], Robertson, S.[2], Watson, H.[3], Lamont, A.L.[1], Black, M.[4], Aotearoa Variome Project team[5], Rakeiora Project Team[5].

[1]Department of Mathematics and Statistics, University of Otago, Dunedin, NZ, [2]Department of Woman and Children's Health, University of Otago, Dunedin, NZ, [3]Ngāti Porou Oranga, Gisborne, NZ, [4]Department of Biochemistry, University of Otago, [5]Various.

Genomic resource development for Māori health applications has been underway for a number of years via projects such as Aotearoa Variome and Rakeiora. In this talk I will provide an update on those projects, as well as other significant developments, including steps toward development of a Māori pangenome in collaboration with UC Santa Cruz, and simulations of genetic structures of extant Māori populations including the evaluation of mātauranga whakapapa (genealogies) for estimating disease risk. I will also provide an overview of genetics education efforts in both university and Māori learning environments that collectively seek to increase Māori participation as researchers, and increased control over DNA information that is considered a taonga, and therefore subject to Article 2 of Te Tiriti o Waitangi.

26. Development of a Population Simulator to Optimise Study Designs and Estimate Polygenic Disease Risk/Resilience in Aotearoa New Zealand Māori Populations

Lamont, A.I.[1], Wilcox, P.L.[1], Black, M.A.[2].

[1]Department of Mathematics and Statistics, University of Otago, Dunedin, NZ, [2]Department of Biochemistry, University of Otago, Dunedin, NZ

Disease risk/resilience (DR) prediction requires statistical models that are typically generated from empirical studies. For commonly occurring polygenically inherited conditions such as gout, type 2 diabetes, and cardiovascular conditions, risk/resilience estimates have most often been derived from GWAS (genome-wide association studies). Such studies require large sample sizes ($n > 10^4$ participants) genotyped with 10^4-10^7 DNA markers.

However, such datasets often do not include indigenous peoples, who can have important genetic differences from more commonly represented populations of predominantly European descent. Moreover, existing datasets from Māori (and Pasifika) domiciled in New Zealand are few, and those that could be utilised, consist of fewer than two thousand individuals – typically from case-control studies – thus are underpowered for clinically accurate DR prediction. In addition, establishing sufficiently large GWAS is highly unlikely in Aotearoa/NZ because of substantive costs associated with generating genotypic data and reluctance of many Māori to participate in such studies.

In order to offset further health inequities arising from lack of Māori-specific DR prediction models, new studies are required. Such studies require both (a) optimal designs that incorporate known genetic relationships on non-genotyped as well as genotyped individuals, and (b) analytical methods that more accurately predict phenotype than GWAS-based methods such as polygenic risk score (PRS).

We have used a population simulator (SLiM) to model genetic structures of Māori communities (i.e., whānau/hapū/iwi), incorporating estimates of effective population sizes prior to European admixture, as well as post-colonisation admixture with Europeans. We are using these simulations to explore what features of study design and analytical methods lead to optimal DR prediction. I will illustrate and present current results on this.

27. Indigenous Ethics in Research: The Protection and Correct Treatment of Taonga Species

Edwards, A.[1], Holloway, B.[2], Clarke, S[2].

[1]AgResearch Māori Partnerships Team, [2]AgResearch Animal Genomics.

Over the last 10 years, many advances have been made in our understanding of the Māori view of Aotearoa's native flora and fauna and the significant mātauranga Māori that exists within our iwi, hapū and whānau throughout the country. In this time, many claims, frameworks and guidelines have been developed to aid in the transition for carrying out genomic research with taonga species in an appropriate and respectful way. This includes the preservation and protection of the mātauranga Māori associated with a taonga species when it is used in research e.g. collection, storage and disposal of native flora and fauna. This also encompasses the mātauranga Māori (traditional Māori knowledge) surrounding these species, and the protection of any extension of that mātauranga. This presentation will speak about our recognised taonga species, how they are categorized, why they are significant and finally, what you can do in your research to ensure this significance is acknowledged. These points will be supported by Pūrakau (Māori traditional narratives) that contain mātauranga related to taonga species.

28. Asymmetrical Inheritance in Honeybees: A Polyhybrid Species

Heywood, A.[1], Guhlin, J.[2], Petersen, G., Gilligan, J.[2], Fennessy, P.[3], Dearden, P.[2]

[1]Plant and Food Research, Lincoln, NZ, [2]Department of Biochemistry, University of Otago, Dunedin, NZ, [3]AbacusBio, Dunedin, NZ.

Honeybees (*Apis mellifera*) are a poly-hybrid species where queens and female workers are diploid and drones are haploid. Queen bees practice polyandry by mating with multiple drones in flight. The mix of genetics from different drone fathers maintains the genetic diversity of the colony. Genetic diversity is vital to colony health as it improves disease resistance, productivity, adaptation to changing conditions, and overall resilience. Due to their complex mating systems, haplodiploidy and high genetic diversity, identifying genetic relationships between honeybees is a complex task. As honeybees are agriculturally significant, a method to estimate breeding values and select traits that enhance their pollination efficiency and disease resistance is of great importance. Here we present an experiment based on the managed mating of honeybee queens over 5 generations in 5 different lineages and investigate the parental meiotic contribution to the genotype of these honeybees. Drones and queens that contributed to each pedigree as well as the fifth-generation offspring were whole-genome-sequenced at a read depth of approximately 30 x. As honeybees require polyandry to maintain colony health, queen progeny from each progenitor queen were inseminated from several drone sources. Determination of drone fathers was done by cross-examination of relationship matrices that were based on genetic variants identified by the paternal candidates. Genetic variants provided the basis by which genomic contribution was calculated along the pedigree lines. The maternal contribution was estimated to be ~56% vs ~24 % from the paternal line. However, increased heterozygosity in queen offspring was also detected. The majority of queens had a higher proportion of heterozygous alleles than homozygous alleles, suggesting that despite reduced genomic contribution from the paternal line, genetic diversity is maintained within the population. These findings suggest interesting implications for our understanding of meiosis in honeybees.

29. Novel Tools for Omics Data Analysis in Pseudo-haploid Species

Rodriguez Piccoli, G. [a, b], Hilton, Z. [b], King, N. [b], Gardner, J.P.A. [a], Ritchie P. A. [a, §]

[a] School of Biological Sciences, Victoria University of Wellington, New Zealand, [b] Cawthron Institute, Nelson, New Zealand, [§] Corresponding author peter.ritchie@vuw.ac.nz; guillermo.rodriguezpiccoli@vuw.ac.nz; zoe.hilton@cawthron.org.nz

Statement of the Problem

Genome assemblies of species which contain high levels of duplication, high content of repetitive elements, those with genome-wide hemizygosity, and/or pseudo-ploidy such as molluscs present unique challenges that hinder accurate analysis of genetic variation. There is an urgent need for comprehensive bioinformatic tools to address these complexities.

Brief Methods

We have developed: A Pipeline for Haplotig-resolved Assembly Sequence Elucidation (PHASE); a robust bioinformatic pipeline. An Iso-Seq annotation module called IsoScope that employs EnTAP and InterProScan to annotate PacBio CCS Iso-Seq data. A conceptual pipeline called GenPop for population genetics analysis using low-coverage whole-genome sequencing with short reads. The solution is distributed through a Docker virtualised server based on the Butterfly development cycle principle for enhanced data presentation.

Results

PHASE successfully resolved an assembly from DNB Hi-C / HiFi reads sequenced by BGI, in a highly duplicated and hemizygous mollusc species. IsoScope annotated Iso-Seq data from BGI obtaining a saturated gene count curve, while GenPop identified critical genetic variation metrics in three geographical populations of interest, using Illumina reads.

Conclusions and Significance

PHASE provides an all-encompassing solution for mollusc genomic research, from sequence assembly to population genetics. Its modular design and Docker-based presentation make it a scalable and user-friendly resource, thereby advancing the field of genetic variation analysis in complex genomes.

30. TaFFE: A Snakemake Workflow for Taxonomic and Functional Analysis of Meta-Omics Data Derived from Sheep Rumen Microbiomes Divergently Selected for Methane Emissions

Perry, B.J. [1], Kim, A. [1], Henry, H. [1], Booker, F. [1], Bilton, T.P. [1], McRae, K.M. [1], McCulloch, A. [1], Alemu, S. [1], Clarke, S. [1], Janssen, P.H. [2], McEwan, J.C. [1], Rowe, S.J. [1]

[1]Invermay Agricultural Centre, AgResearch, Mosgiel, NZ, [2]Grasslands Research Centre, AgResearch, Palmerston North, NZ

We present a bioinformatic pipeline for Taxonomic and Functional Feature Extraction (TaFFE) from shotgun metagenomic and metatranscriptomic sequencing data, built using the Snakemake workflow control language and open-source software. Here we used the TaFFE pipeline to analyse 250 bp paired-end shotgun metagenomic data generated from the rumen contents of selection lines of high and low methane emitting sheep. Briefly, the TaFFE pipeline trims, entropy filters, quality filters, host-depletes, and finally profiles sequencing reads both taxonomically and functionally using kraken2 and humann3. Taxonomic profiling utilises the Genome Taxonomy Database (GTDB) whereas functional profiling utilises the UniProt Reference Clusters (UniRef) database. Comparisons of the features extracted with TaFFE from the metagenomes from rumen contents samples from high and low methane-emitting sheep identified significant taxonomic and functional differences. Biodiversity of the rumen microbiome was significantly higher in the high-methane line, possibly due to a higher prevalence of archaeal genera. Methanogens and methanogenic functions were associated with the high-methane line while the low-methane line was associated with producers of propionate and butyrate, products of fermentation that lead to less methane. Future work will establish the utility of these features for the prediction of sheep methane emission phenotypes directly from taxonomic and functional features extracted from rumen microbiome data such as these.