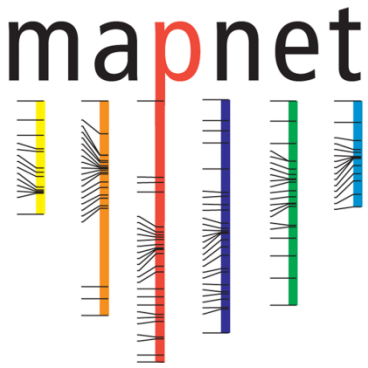

1-2 Sept 2016, Nelson, New Zealand

ABSTRACTS

Oral 1. Predicting routes of horizontal gene transfer using completely sequenced genomes

Heather Hendrickson, Massey University

Horizontal gene transfer (HGT) is a process by which large quantities of genetic information are transferred between bacteria in nature. Despite significant frequencies of DNA admixture there is ecological and phenotypic cohesion amongst higher taxonomic groups in the eubacteria. This suggests that there are mechanisms that impose structure or rules on HGT. Most bacterial genomes have a functional architecture composed of repeat sequences called Architectural IMparting Sequences or AIMS. These are 8 base pair sequences that have distributions in the genome that suggest that they function in orienting DNA associated enzymes to facilitate processes like DNA segregation. AIMS are phylogenetically conserved such that HGT events that occur between closely related bacteria will tend to preserve AIMS structure while more distant transfer events will disrupt AIMS structure therefore be detrimental to the organism.

AIMS are not under equally strong selection across the length of the chromosome. Selection on orientation and abundance is strongest at the terminus and eases towards the origin. In order to understand the degree to which AIMS shape bacterial genomes, we have analyzed naturally occurring disruptions in bacterial chromosomes. We have found that both inversions and insertions appear to be affected by AIMS distributions in bacterial genomes. Last, we used completely sequenced genomes to produce a predictive road map of paths of horizontal gene transfer between species based on AIMS compatibility between donors and recipients. According to these data, some HGT will non-reciprocity of genome compatibility between some donors and recipients. AIMS therefore suggest that some bacterial clades should be expected to be universal donors but not universal acceptors of genetic transfer.

Oral 2. The *Escherichia coli* nucleoid: an organized structure that is shaped by replication and transcription

Ralph S. Grand[1,2,†#], Rachel Bowden[1,†], Tatyana Pichugina[1,†] and Justin M. O'Sullivan[1,*].

Affiliations:
[1]Liggins institute, University of Auckland, Grafton Auckland 1032, NZ, [2]Institute of Natural and Mathematical Sciences, Massey University, Albany, Auckland 0745, NZ.

[#] Present address: Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland.

[†] The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

Contact email:
* To whom correspondence should be addressed. Tel: +64 9 923 9868; Email: justin.osullivan@auckland.ac.nz

The *Escherichia coli* nucleoid is confined within a rod shaped cell many times smaller than the outstretched chromosome. While folding is essential for this process, the chromosome must at the same time remain accessible to essential cellular processes such as replication and transcription. Currently, the individual contributions of cellular confinement, chromosome topology, replication and transcription on nucleoid organization are not well understood. Here we synchronize *E. coli* cells in stationary phase, where replication has ceased, each cell contains only one copy of the chromosome, and transcription is minimal. We then release the cells and capture chromosome contacts and transcript profile immediately following release and through-out one cell cycle. Polymer models of confined and topologically constrained circular polymers revealed that cellular confinement does not contribute extensively to the organization of the *E. coli* nucleoid. Rather local nucleoid structure is established concurrent with replication, the clustering of linearly distant SeqA bound sites, and gene transcription.

Oral 3. PacBio methylome sequencing of *Neisseria meningitidis* – potential associations with disease

Ren, X.[1], Eccles, D.[2], and MacKichan, J.K.[3]

[1]Invasive Pathogens Laboratory, Health Programme, Institute of Environmental Science and Research, Wellington, NZ; [2]Institute of Environmental Science and Research, Wellington, NZ; [3]School of Biological Sciences, *Victoria* University of Wellington, Wellington, NZ.

Neisseria meningitidis, the main causative agent for bacterial meningitis, is a human-specific commensal bacterium which normally resides in the human nasopharynx, but can become invasive causing inflammation of the meninges and septicaemia that can cause death or long-term disabilities. Extensive epidemiology and molecular typing support the presence of bacterial virulence factors, as only a small percentage of meningococcal types cause most cases of invasive diseases. However, attempts to identify virulence factors have not yielded any candidates present exclusively in disease-associated lineages. Virulence in meningococci is likely to be polygenic, and we believe differential regulation of the pathways required for adaptation to different host sites may distinguish carriage and invasive lineages. We used a unique collection of disease and household-contact meningococci isolates to start investigating the bacterial factors which may be involved in distinguishing carriage vs disease.

Between 1996 and 1998 the Auckland Healthcare Public Health Protection Service and Institute of Environmental Science and Research carried out a carriage study among household contacts of meningococcal disease patients. Further analysis of the isolates by Multi Locus Sequence Typing identified instances where healthy household members were colonised by an isolate indistinguishable from that isolated from the patient. Using RNAseq analysis, we identified several transcripts that are differentially regulated in disease-associated (patients) vs carriage-associated (household contacts) isolates within the same household. Though some differences in transcription level can be explained by known regulation mechanisms, many of the observed differences cannot be easily explained.

Epigenetic control of gene expression is better understood in eukaryotes. In bacteria such as *E. coli*, Salmonella and Haemophilus the involvement of DNA methylation in transcription and virulence have also been described. We used PacBio Single-molecule-real-time (SMART) sequencing to identify modified bases and performed initial analyses to integrate the transcriptome and methylome for several disease- and carriage-associated isolates from two different households. Here, we will describe the partial methylome of these isolates, differences between disease and carriage-associated isolates, as well as discussing whether methylation plays a role in the gene expression differences seen.

Oral 4. Causal modeling of relationships between genotype and phenotype using multi-omics data

Heather Cordell, University of Newcastle, UK

Over the past 10 years, genome-wide association studies (GWAS) have been extraordinarily successful at identifying genetic variants associated with common, complex disorders. However, a typical GWAS gives little insight into the underlying biological mechanism through which the associated genetic variants are implicated in disease. Genetic variants identified through such investigations represent the first step along the causal pathway to disease development, and one strategy to help elucidate the underlying causal mechanisms is to make use of data of different types, including genetic data, "omics" data such as measurements of DNA methylation and gene expression, and clinical variables related to disease phenotype. A variety of methods can be used to model the relationships between these different variables. In this talk I will focus on methods that assume at least a proportion of subjects have been measured on all variables of interest. I will outline the methodological approaches that we have been exploring based on causal inference techniques, and present the results of computer simulations and real data analyses illustrating the utility of these approaches.

Oral 5. Dissecting and fine-mapping trans-eQTL hotspots

Karl W. Broman
Department of Biostatistics & Medical Informatics
University of Wisconsin - Madison
http://kbroman.org

Studies of the genetic loci that contribute to variation in gene expression, called expression quantitative trait loci or eQTL,frequently identify loci with broad effect on gene expression: trans-eQTL hotspots, which affect the expression of many genes. In a large mouse intercross with gene expression microarray data on six tissues in nearly 500 mice, we identified one such trans-eQTL hotspot that affects the expression of about 8% of all genes.

We have developed a set of exploratory tools, as well as a formal statistical test, for assessing whether a given hotspot is due to one or multiple polymorphisms, and to define the QTL interval in the case that it appears to be due to a single polymorphism. A useful device is to examine the pattern of gene expression, in the genes that map to a given trans-eQTL hotspot, among individuals that show no recombination event in the region, in comparison to individuals that did have a recombination event. The QTL effects, including the degree of dominance, can also be informative. As part of our formal test for one vs two polymorphisms in a region, we consider a two-QTL model with each expression trait affected by one or the other QTL.

Oral 6. Variance GWAS

Wenhua Wei, University of Otago

Seronegative rheumatoid arthritis is a heterogeneous disorder where patients have no antibodies against citrullinated peptides. GWAS so far have identified only two genome-wide significant loci *HLA-DRB1* and *ANKRD55*. Examining gene-environment interactions (GxE) is desired but traditional approaches are hampered by 1) unknown causal environmental factors and 2) small samples. Genotypic variability based GWAS can prioritize non-additive loci without requiring prior knowledge of causals and thus is a useful alternative to explore.

In this talk, I will first introduce a two-step variance GWAS approach adapted for case-control disease phenotypes. Then I will report how the variance GWAS was applied to study seronegative rheumatoid arthritis to discover a genome-wide significant 'novel' locus *DHCR7*, which is believed important in vitamin D metabolism particularly in northern latitudes with insufficient sunlight but carries only moderate additive effects. I will also discuss the limitation of the study and future perspectives.

Oral 7. Interactive exploration of genome-scale datasets

Nick Burns, Ruth Topless, James Boocock, Michael A Black, Tony Merriman
Department of Biochemistry, University of Otago, Dunedin, NZ

Genome-scale analyses are now well-established in the study of complex human diseases. The variety of statistical tests have extended from the traditional genome-wide association study (GWAS) to include a wide range of functional and spatial traits. There is a need for analytical and exploratory tools to combine and compare results across multiple genome-scale analyses. Unfortunately, the sheer size of these datasets imposes severe restrictions on the types of questions that can be explored. Current genome browsers and exploratory tools are mostly restricted to simple lookups and summaries which miss the potential to discover more compelling relationships.

The Genotype-Tissue Expression (GTEx) project is a ready example of the problems researchers face when dealing with such large, heterogeneous datasets. Combined, the genotype, gene expression and expression quantitative trait loci (eQTL) datasets total more than 500GB of data. The GTEx portal provides a user-friendly interface to explore single-tissue summaries of this data but there are no means by which to compare results across tissues or genes, nor to directly contrast the gene expression profiles with interesting eQTLs. The sheer size of these data necessitates such restrictions, but limits the exploratory potential of such a rich dataset.

We have developed a data warehouse which can facilitate full, unrestricted exploration of genome-scale datasets, which we have applied to the gene expression and eQTL datasets from GTEx together with various GWAS summary datasets. The benefits of deduplication and optimised data access strategies reduce the problems of scale normally associated with these data. This presentation will demonstrate an interactive interface to the data warehouse, using R and Shiny, which provides rich and intuitive visualisations via a web browser, allowing users to quickly explore loci across the GTEx datasets, including investigating relationships to other loci of interest.

Oral 8. Conservation genomics

Shaffer, H.B.[1], McCartney-Melstad, E.[1], Ashander, J.[1], Ralph, P.[2],

[1] Department of Ecology and Evolutionary Biology, and La Kretz Center for California Conservation Science, University of California, Los Angeles, CA, USA, [2] Department of Mathematics, University of Oregon, Eugene, Oregon, USA

The relationship between conservation biology and evolutionary genetics has always been an intimate one, and genomics has both strengthened and clarified that relationship. The primary influence of genomics on applied conservation biology to date has mostly centered on better resolution of traditional problems and questions, ranging from species delimitation to population structure and landscape genomics. These types of analyses remain crucial to decision-makers charged with endangered species management, and they form the bulk of most current research efforts in the field. However, some of the most exciting advances, many centered on using full-genome resequencing to identify genes that are important for climate adaptation, disease resistance and other responses to near-future adaptation, are just emerging on the conservation genomics horizon. In this talk, I will review some of the ways that genomics is contributing to effective conservation management with examples from our current research efforts. Primary empirical examples include range-wide, full genome resequencing work on endangered tortoises to evaluate the effects of mega-scale solar power installations on gene flow and population viability, and genome-enabled analyses of a hybrid invasion of an endangered salamander. Throughout, I emphasize the role that genomics can and should play in helping to focus research efforts on effective, on-the-ground management as well as the underlying evolutionary mechanisms that underlie best practices in conservation biology.

Oral 9. Whole genome sequencing and kākāpō recovery

Bruce C. Robertson
Department of Zoology, University of Otago, Box 56 Dunedin 9054 New Zealand

The critically endangered Kākāpō is genetically managed to avoid inbreeding and to ensure the maintenance of genetic diversity. Coming off a low of 51 individuals in 1996, at the start of the last breeding season, the population numbered 125 Kākāpō. Genetic management was based on parentage analyses and assessment of relatedness coefficients using variation at 25 polymorphic microsatellite loci. Less than ideal matings are redressed with artificial insemination or changing the makeup of the breeding population.

Here I will talk about a new approach to Kākāpō genetic management, the Kākāpō 125 Genomes Project, which aims to sequence the genome of every surviving individual of a species. This ambitious project is a collaboration among researchers from the New Zealand Department of Conservation's Kākāpō Recovery Group, the University of Otago (New Zealand), Duke University (USA), the Genetic Rescue Foundation and NZ Genomics Ltd. The project uses a high-quality reference kakapo genome developed by Duke University and PacBio as a scaffolding to assemble all other 124 kakapo genomes against. This work is being funded by the Genetic Rescue Foundation, a not-for-profit organization dedicated to advancing the scientific techniques required to prevent species extinction through genetic intervention.

A genomics approach to kakapo conservation promises a great boost to the recovery program and allows us to move beyond the limited understanding 25 microsatellite loci can offer. A genome-wide understanding of genetic variation will help to develop breeding strategies to retain variation at genes important for species persistence, such as the immunity genes and their role in Kākāpō diseases. We will also be able to explore the genetic basis of infertility in kakapo; only 60% of eggs hatch (normally this should be about 90% in birds) and sperm abnormalities contribute to infertility. Solving the issue of infertility will greatly aid species recovery by maximising reproductive effort.

Oral 10. A conservation genomic approach for maximising genetic diversity in a critically endangered New Zealand bird.

Galla, S.J. [1], Hale, M.L. [1], Santure A.W. [2], Steeves, T.E. [1]

[1]School of Biological Sciences, University of Canterbury, Christchurch, NZ

[2]School of Biological Sciences, University of Auckland, Auckland, NZ

Endangered species management can utilise captive breeding for translocation as a technique to prevent extinction and enhance species recovery. Captive breeding programmes often make pairing decisions based on available pedigree data to minimise inbreeding and maximise genetic diversity in an effort to maintain the ability to adapt to environmental change (i.e., evolutionary potential). However, captive pedigrees are often shallow (<5 generations deep), incomplete and error-prone. While genetic-based techniques (microsatellites) offer these programmes a way to estimate genetic relatedness among individuals, emerging evidence indicates that genetic-based measures of relatedness based on microsatellites are relatively poor indicators of genome-wide diversity, particularly in genetically impoverished endangered species, and a better indication of genome-wide diversity should be obtained from genomic-based measures of relatedness based on genome-wide single nucleotide polymorphisms (SNPs). Here, I will provide a comparison of pedigree-based, genetic-based and genomic-based estimates of relatedness in kakī (*Himantopus novaezelandiae*) from the 2015/2016 breeding season as part of a larger effort to determine which of these approaches is the most efficient and effective for maintaining evolutionary potential in this critically endangered New Zealand endemic. Using kakī as a proof-of-concept, this research can be applied to other poorly-pedigreed captive breeding translocation programmes in New Zealand, and worldwide.

Oral 11. Assembling the South Island robin genome to understand inbreeding depression: the pitfalls of *de novo* assembly faced by ecological geneticists

Helen R. Taylor[1], James V. Briskie[2], Kim M. Rutherford[1], Neil J. Gemmell[1]

[1] Department of Anatomy, University of Otago, Lindo Ferguson Building, Great King Street, Dunedin, 9016, New Zealand; helen.taylor@otago.ac.nz;
[2] School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

Assembling genome sequences *de novo* has become relatively common practice for ecological geneticists seeking to understand the genomic mechanisms underlying various traits and processes. However, ecological geneticists are not generally bioinformaticians and are often inexperienced in genome sequencing and assembly. Unfortunately, there is a lack of information in the scientific literature regarding guidelines for *de novo* assembly and "best practice" can vary substantially between taxa. We are currently assembling a *de novo* genome for the South Island robin (*Petroica australis australis*). We aim to use this genome to elucidate the links between inbreeding depression and poor male fertility, a subject that has rarely been assessed in wild populations, has almost exclusively been examined in mammals, and has not yet been approached using full genome sequencing. Here, we will outline the steps we have taken to ensure the best quality assembly possible (pre-assembly quality control procedures), as well as comparisons between different assembly software. We will explain some of the pitfalls we have faced as ecological geneticists new to genome assembly, and how we have overcome these issues, emphasising the need for general guidelines for de novo assembly non-model species. Finally, we will explain how we hope to use our robin genome to assist in seeking out the signatures of inbreeding that might explain decreases in male fertility in this, and other, species.

Oral 12. Whole-transcriptome profiling of flexible sexual phenotypes in a model sex-changing fish, the bluehead wrasse.

Erica V. Todd[1], Hui Liu[1], Melissa Slane-Lamm[2], John Godwin[2], Neil J. Gemmell[1]

[1]Department of Anatomy, University of Otago, Dunedin 9016, New Zealand.
[2] Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA.

Sex is increasingly seen as a continuous, rather than a dichotomous trait. Sex is phenotypically plastic in many marine fishes and results from environmentally-sensitive differential gene regulation. Bluehead wrasse (*Thalassoma bifasciatum*) are highly-social reef fish and well-studied models of sexual plasticity. These diandric, protogynous (female-first) hermaphrodites have three sexual morphs as adults whose development is plastic and socially cued. Bluehead wrasse mature as male (primary males) or female, but each have the capacity to become dominant (secondary) males later in life. Large, brightly coloured secondary males actively defend and court a harem of females, whereas primary males are female-mimics that employ a 'sneaker' mating strategy. Using whole-transcriptome RNA-sequencing (RNA-seq) we have explored the molecular basis of plastic sexual phenotypes in the bluehead wrasse brain and gonad. Differential expression analysis identified thousands of genes important in the maintenance of the primary male, secondary male, and female phenotypes. In the brain, secondary males had the most distinct expression patterns, whereas expression profiles of primary males reflect their female-like behaviour, not their male sex. We find that isotocin (homologue of mammalian oxytocin) is overexpressed in secondary males, supporting recent evidence for its regulatory role in teleost social interactions, especially those related to dominance and rank. Gonadal expression profiles were strongly sex-biased, although secondary males upregulated genes involved in androgenesis and in the maintenance of secondary sexual characteristics (i.e., colouration and territoriality). Further investigations into the molecular basis of sexual plasticity are now underway, including transitions between alternative phenotypes and comparisons of gene expression patterns in evolutionarily divergent systems.

Oral 13. Genetic diversity in Tasmanian Atlantic Salmon and prospects for GWAS and genomic prediction

James Kijas [1], Peter Kube [2], Brad Evans [3], Natasha Botwright [1], Harry King [2], Craig Primmer [4], Klara Verbyla [5]

[1] CSIRO Agriculture, Brisbane 4067, Australia, [2] CSIRO Agriculture, Hobart 7004, Australia, [3] SALTAS, Hobart, 7004, Australia, [4] Division of Genetics and Physiology, University of Turku, 20520, Finland, [5] Data61, Canberra, 2601, Australia.

As a precursor to genome wide association and genomic prediction in Tasmanian Atlantic salmon, we collected genotypes at 218,132 SNP in 777 fish from a breeding population. The objective was to assess levels of genetic diversity, the strength of linkage disequilibrium (LD) and imputation accuracy. Genetic diversity was much lower than observed within European populations and the distribution of allele frequencies showed a clear difference, with the Tasmanian animals carrying an excess of low minor allele frequency variants. The strength of observed LD was high a short distances and remained above background for marker pairs separately by large chromosomal distances (hundreds of kb). To estimate imputation accuracy, genotypes from low density marker sets (0.5 K to 5 K) were used to impute genotypes of an increased density SNP set (78 K). This revealed high imputation accuracies (0.89 – 0.97), suggesting use of low density SNP sets will be a successful approach for genomic prediction in this population. The persistence of LD across long physical distances, low levels of genetic diversity and high imputation accuracy in Tasmanian salmon is consistent with known aspects of their population history, which involved a small founding population and no subsequent introgression. The findings of this study represent an important first step towards the design of methods to apply genomics in this economically important population.

Oral 14. Genomes, gene discovery and fish immunity

Steve Bird, Waikato University

The availability of fish genomes and the development of high-throughput sequencing techniques has revolutionized gene discovery. This has allowed researchers to describe the immune system of fish in detail and has shown that they are capable of both innate and adaptive responses, similar to that seen in mammals. However, the actual characterisation of fish immune genes is complicated, due the presence of duplicated and novel genes. Additionally, with over 33,000 species described, the few fish genomes we have available to investigate have also indicated marked differences in immune genes that are present between fish groups. This talk will summarize some recent investigations that have been using genomes to understand the repertoire of immune genes that exist in fish. It will highlight how genomes are proving useful in determining the actual identity of these genes. Lastly, it will touch on current work that is being undertaken to understand fish immune gene function, using zebrafish as a model species.

Oral 15. Making a splash: using genomics to aid the New Zealand aquaculture industry

Rachael Ashby[1,2], Hayley Baird[1], Rudiger Brauning[1], Tracey van Stijn[1], Rayna Anderson[1], Ken Dodds[1], Kim Rutherford[2], Chris Brown[2], John McEwan[1], Neil Gemmell[2] and Shannon Clarke[1]
[1]AgResearch, Invermay Campus
[2]University of Otago

Aquaculture is one of the fastest growing primary industries. The New Zealand Aquaculture industry generates over $400 million per year in revenue, with the aim to increase this to $1 billion by 2025. Exports of the Greenshell™ Mussel ($218m) and the Chinook salmon ($63.4m) are the largest by both value and volume. Utilisation of genomic tools can be used to aid traditional breeding programs by selecting for high value traits. Advances in DNA sequencing technologies now mean we have the ability to cost effectively genotype individuals for thousands of single nucleotide polymorphism (SNP) markers. We are developing Genotyping-by sequencing (GBS) pipelines for both the Chinook salmon and the Greenshell™ Mussel to enable these aquaculture industries to implement inbreeding control, and pedigree construction for genomic selection.

Oral 16. Phenotyping and genomic resources for domestication of the marine finfish snapper (*Chrysophrys auratus*)

David T. Ashton[1], Peter A. Ritchie[2] & Maren Wellenreuther[1,3]

[1]The New Zealand Institute for Plant & Food Research Limited, Nelson, New Zealand
[2]School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand
[3]Molecular Ecology and Evolution Group, Department of Biology, Lund University, Sweden

Mapping the genetic basis of phenotypic variation is a major challenge for biology. The recent development of next generation sequencing technologies has dramatically improved our ability to address this challenge, particularly in non-model species which have limited genomic resources. Australasian snapper (*Chrysophrys auratus*) is one such species that has limited genomic resources and is of interest for fisheries and as a potential aquaculture species. A range of phenotyping and genomic resources are being developed at Plant and Food Research to support domestication and genotype-phenotype mapping in this species. In this talk, I will discuss the development of these resources including 1) genotyping of a three generation pedigree using Genotyping by Sequencing (GBS), 2) development of automated software for rapid phenotyping from images, 3) identification of QTLs associated with growth rate, and 4) development of a recently assembled genome.

Keywords: Snapper, QTL, GBS, genome, New Zealand

Oral 17. Genomics assisted yield improvement in alfalfa - Are we making more hay?

E. Charles Brummer[1], Gitanshu Munjal[1], Scott A. Newell[1], Jingjie Hao[1], Yanling Wei[1], and Xuehui Li[2]. [1]Plant Breeding Center, Department of Plant Sciences, Univerisity of California, Davis, Davis, CA USA and [2]Department of Plant Sciences, North Dakota State University, Fargo, ND, USA. ecbrummer@ucdavis.edu

Alfalfa (aka, lucerne, *Medicago sativa*) is a high value forage legume used extensively in rations for dairy cattle and other livestock in the USA and worldwide. A deep rooted, perennial crop, alfalfa provides positive environmental services when grown in rotation with other crops. The yield potential of alfalfa in the United States has not increased in the past 25+ years, although improvements in disease and insect resistances have enabled stands to survive longer in the field. We have been investigating the potential to improve yield through various strategies including the capture of heterosis for biomass yield, genomic selection to increase selection cycle time, and the manipulation of dormancy. I will discuss these three methods in this paper, focusing most of the time on the last two. Genomic selection, a procedure that enables selection based on whole genome marker analysis, could expedite breeding because field selection typically lasts up to five years per cycle. Our preliminary data suggests that yield gain can be made using genomic selection. Decreasing autumn (fall) dormancy could increase late summer and autumn yield but can lead to winter injury. We have mapping results suggesting that the two traits can be decoupled so that higher yield can be selected without diminishing winter survival. Dormancy also controls or is associated with many other agronomically important traits including regrowth rate following harvest, flowering time, and nutritive value. We are evaluating dormancy from several perspectives to determine the genetic mechanisms behind the trait and to develop rapid methods to modify dormancy. Using biparental genetic mapping, genome-wide association analysis, and selection mapping, we have identified a number of potential candidate genome regions controlling dormancy. Application of genotyping-by-sequencing (GBS), spiked GBS, and target capture GBS have enabled us to more rapidly analyze alfalfa populations than in the past. These technologies offer opportunities for all species to accelerate molecular breeding and trait dissection in the absence of extensive pre-existing genome tools.

Oral 18. Mānuka genome assembly using chromosome conformation capture Hi-C analysis

David Chagné*[1], Amali Thrimawithana[1], Ross Crowhurst[1], Ivan Liachko[2], Shawn Sullivan[2], David Lewis[1], Julie Ryan[1], Munazza Saeed[1], Tracey Van Stijn[3], Jeanne Jacobs[3], Rudiger Brauning[3], Shannon Clarke[3], Cecilia Deng[1], Helge Dzierzon[1], Ella Grierson[1], Elena Hilario[1], Kathy Schwinn[1]

*presenting
1 The New Zealand Institute for Plant & Food Research Ltd
2 PhaseGenomics Ltd
3 AgResearch Ltd

We have evaluated a novel strategy combining capture of chromatin interaction within the nucleus, next-generation sequencing (NGS) and new bioinformatics methods (Hi-C) for developing a near-complete pseudo-chromosome assembly of the mānuka (*Lepstospermum scoparium* 'Crimson Glory') genome. The method relies on the folded confirmation of chromosomes inside the nucleus and the assumption that segments of the same chromosome are in closer 3D proximity with other segments of the same chromosome than to segments of other chromosomes. We fixed and captured chromosome fragments that were in close proximity and then sequenced them using NGS. The Hi-C bioinformatics approach utilizes the probability that chromosome fragments within paired-end NGS reads are close in proximity and this probability can be used to aid scaffolding. Using the Hi-C analysis, we inferred the location of previously unanchored contigs and created chromosome length super-scaffolds, enabling a ~100 times increase in the N50 scaffold length compared with assembly in the absence of this analysis. Furthermore, the Hi-C technique allowed the separation of the mānuka genome from contigs assembled from associated endophytic fungal and bacterial species. The newly assembled mānuka genome was compared to the genome of *Eucalyptus grandis* and high density genetic maps of mānuka constructed using genotyping-by-sequencing. The mānuka genome sequence will help shed new light on the genetic control of unique characters such as nectar and foliage biochemical composition, flowering time and disease resistance.

Oral 19. “Genotyping-by-sequencing” platform to recover genetic relatedness

Klápště, J.[1], Telfer, E.[1], Dungey, H.[1]

[1]Forest Genetics, Scion, 49 Sala Street, Rotorua, New Zealand.

Current developments in next generation sequencing technologies have enabled the implementation of genomics in organisms with no reference genomes, as is the case for most forest trees. The availability of genomic resources and success of genomic selection achieved in animal and crop breeding has now turned the attention of breeders towards the implementation of genomic selection in forest tree breeding programs. Genomic predictions capitalize on capturing genetic relatedness, co-segregation and linkage disequilibrium (LD) between markers and causal variants (quantitative trait loci - QTL). However, common forest tree genome properties such as large size, fast LD decay and large effective population size hinder the ability of genomic prediction models to capture LD between markers and QTLs, which is usually the most stable part of genomic prediction. This means that genomic prediction models in forest trees rather depend on the quality of relatedness recovery, and should be used to predict only related genetic material.

Our study is focused on the investigation of a “Genotyping-by-sequencing” (GBS) platform to recover genetic relatedness. This platform is cost efficient to genotype large training populations but usually suffers a large amount of missing data and low sequence depth. Under such conditions, it can be problematic to perform reliable missing data imputation and relatedness estimations in species with a large effective population size. We deployed tools specifically proposed for GBS data[1] to construct a genomic relationship matrix. Our analysis found there was benefit in avoiding missing data imputation and taking read depth into account for the reliable recovery of genetic relatedness. This strategy can efficiently identify genomic outliers and remove them from the training population, since the informativeness and ability to predict such individuals would be very limited.

1. Dodds, KG., McEwans, JC., Brauning, R., Anderson, RM., van Stijn, TC., Kristjánsson, T, Clarke, SM. (2015). *Construction of relatedness matrices using genotyping-by-sequencing data.* BMC Genomics 16:1047. Doi: 10.1186/s12864-015-2252-3

Oral 20. Predicting the future of C3 plants using a modified small structural protein

N. Roberts, S. Winichayakul, K. Richardson, M. Andrews, S. Rizvi, G. Bryan
AgResearch

Our laboratory is predominately focussed on metabolic engineering of forage plants for the benefit of the NZ pastoral industry. Sometimes seemingly minor modifications to one pathway result in substantial changes to others; the work we present here is an example of this. Through the substitution of 6 residues within an oleosin protein (a seed oil body protein) we have increased CO_2 fixation by 20-25% and increased biomass production of over 50% in Arabidopsis and ryegrass. In both species we measured a number of well documented physiological changes associated with C3 plants grown in elevated CO_2 environments; interestingly, we also observed some relatively controversial changes. This presentation discusses our findings and the ramifications of C3 plant growth with an increase in atmospheric CO_2.

Oral 21. GWAS for weight in Samoans

Steve McGarvey, Brown University, USA

Samoans of Polynesia are a unique founder population with a past history of semi-isolation and a recent history of obesity. They offer opportunities to identify novel genetic contributors to obesity and other cardiometabolic conditions. A genome-wide association study was performed in 3,072 Samoans, identifying a variant on chromosome 5q35.1, rs12513649, strongly associated with body mass index (BMI) ($P=5.3x10^{-14}$). This was replicated in 2,102 additional Samoans from earlier studies ($P=1.2x10^{-9}$). Targeted sequencing revealed a strongly associated missense variant rs373863828 (p.Arg457Gln) in CREBRF (meta $P=1.4x10^{-20}$). This variant is very rare in other populations, but common in Samoans (frequency 0.259). Its effect size on BMI was much larger (1.36–1.45 kg/m^2) than other common risk variants. Surprisingly the variant was inversely associated with fasting glucose (1.65 mg/dl, $P = 9.5 \times 10^{-5}$) and odds of type 2 diabetes (OR = 0.586, $P = 6.68 \times 10^{-9}$). In the 3T3-L1 adipocyte model overexpression of the p.Arg457Gln variant selectively decreased energy utilization and increased fat storage relative to wild-type (WT) CREBRF. There was strong evidence from genomic data for positive selection of the p.Arg457Gln allele. Together these results support a thrifty variant hypothesis as a factor in human obesity and provide evidence for future mechanistic studies, as well as G x E and 'omics investigations in Samoans.

Oral 22. Identification of biologically informative uncommon genetic variants in urate control by re-sequencing in extreme phenotype

Tony Merriman[1], James Boocock[1], Tanya Flynn[1], Callum Tanner[1], Tahzeeb Fatima[1], Hyon Choi[2], David Mount[3], Lisa Stamp[1], Nicola Dalbeth[4], Eli Stahl[5]

1, University of Otago; 2, Harvard University, Boston; 3, Brigham and Women's Hospital, Boston; 4, University of Auckland; 5, Mt Sinai School of Medicine, New York.

Elevated levels of urate (hyperuricemia) are an essential risk factor for gout. Genome-wide association studies (GWAS) have identified >30 loci harbouring common genetic variants with weak effects on urate levels. However, given that the causal genes and genetic variants at the majority of these loci are not obvious, translation of GWAS findings into biological knowledge is challenging. I will describe a re-sequencing experiment of urate GWAS and other candidate loci using 800 Maori and Pacific (Polynesian) and European individuals from opposite ends of the urate spectrum. The primary aim of this experiment was to identify candidate causal genes at GWAS loci – such genes are hypothesized to have a burden of uncommon functional variants. While achieving this aim was hampered by moderate statistical power, we have been able to identify several variants that have been translatable into biological knowledge. In addition a rare genotype combination at the ABGC2 locus provides a clinically translatable possibility.

Oral 23. Insulin resistance and diabetes - insights from monogenic causes

Rinki Murphy
University of Auckland

Identifying those with diabetes or severe insulin resistance disorders that result from a single gene defect has not only enabled improved diagnostic and clinical management of such patients, but has also resulted in key biological insights into the pathophysiology of the increasingly common forms of diabetes and insulin resistance. This presentation will outline the major types of human monogenic disorders that I have found in my clinical practice that result in pancreatic beta cell forms of early-onset diabetes or severe insulin resistance manifesting in absence of obesity at a young adult age. The lessons they provide for current understanding of the molecular pathogenesis of common type 2 diabetes and insulin resistance are highlighted.

Oral 24. The Dynamic and Hypervariable Nature of *LPA*

Taiaroa, G [1,2] Brown, AN [1] McCormick, SP [1]

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

[2]Department of Microbiology and Immunology, University of Otago, Dunedin, NZ

High lipoprotein(a) levels are a major risk factor for cardiovascular disease[1]. Significantly elevated or lowered levels of lipoprotein(a) are caused by variation in the *LPA* gene[2]. Of these, the most significant is copy number variation at a repeat region of LPA (KIV_2), there being 2-40 copies of the repeat at any one allele. The size of this *LPA* repeat region has been shown to be the basis of 30-70% of lipoprotein(a) phenotypic variance[2]. A second major genetic determinant of this phenotypic variance is predicted to be sequence variation located in the *LPA* repeat array[3]. Identifying such variation brings the challenges inherent in sequencing highly repetitive regions.

In this project, next generation sequencing methods were used to reveal the sequence variation across the *LPA* gene in 48 Caucasian individuals. A focus has been placed on the KIV_2 repeat array, including targeted deep sequencing of the two exons encoded in the repeat. Amplicon generation was achieved with PCR-based methods (sequenced at an average coverage of 2,500 reads per exon) and RPA-based methods (sequenced at an average coverage of 12,000 reads per exon). These two methods were coordinated to reduce the significance of any amplification errors or bias towards certain repeat types.

So far, this custom method has identified 10 new sequence variants within the *LPA* repeat array – the total number of known variants in this region now being 28[3]. Two of the novel variants have only been identified in individuals with low or null lipoprotein(a) levels, and are predicted to change the structure of the encoded protein. Other novel variants are illustrative of recent repeat expansion and concerted evolution across the *LPA* repeat region, detailing the dynamic evolution of the *LPA* gene.

1. Kamstrup PR., Nordestgaard BG. *Elevated lipoprotein (a) levels, LPA risk genotypes, and increased risk of heart failure in the general population*." JACC: Heart Failure 4.1 2016

1. Kronenberg F., Utermann G. *Lipoprotein(a): resurrected by genetics.* J Intern Med. 2013

2. Noureen A., Fresser F., Utermann G., Schmidt K. *Sequence variation within the KIV-2 copy number polymorphism of the human LPA gene in African, Asian, and European populations.* PloS one. 10(3), e0121582 2015

Oral 25. Runs of homozygosity in Eastern and Western Polynesians and metabolic disease

Bixley, M.J.[1], Flynn, T.J.[1], Topless, R.[1], Dalbeth, N.[2], Stamp, L.K.[3], Wilson, J.[4], Merriman, T.R.[1]
[1]Department of Biochemistry, University of Otago, Dunedin, NZ
[2]Department of Medicine, University of Auckland, Auckland, NZ
[3]Department of Medicine, University of Otago, Christchurch, NZ
[4]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, UK

Homozygosity caused by consanguineous union has long been associated with an increased prevalence of rare Mendelian disorders. In contrast, the role of homozygosity in relation to quantitative traits and complex disease susceptibility is less well established. Recent work has shown that an increased burden of homozygous DNA segments is associated with reduced height and cognitive ability in diverse human populations. The ROHgen Consortium study investigates the mechanism of runs of homozygosity by exploring different measures of homozygosity. Long homozygous segments arise from recent common ancestry and bring all variants, including rare causal variants, into a homozygous state. In contrast, shorter homozygous segments may arise from more ancient common ancestry and thus capture only the dominance effects of more common variants. Multivariate linear models have been used to show that FROH (the fraction of all autosome in a run of homozygosity >1.5Mb in length) is more predictive of runs of homozygosity than traditional estimates of inbreeding coefficient based on individual SNP homozygosity.

Here, we assess the scope of runs of homozygosity in individuals from Eastern (n=692) and Western Polynesian (n=568) cohorts. Both cohorts show increased frequency in runs of homozygosity when compared to the sample UK population. Trait (urate and renal function as measured by estimated glomurelar filtration rate) residuals from multivariate models were then regressed against the various measures of ROH. In Eastern Polynesians there is an indication that urate may be affected by ROH with several of the models showing significance ($p < 0.05$). Log(egfr) approaches significance ($p=0.07$) in one model. There are no models of significance for urate in the Western Polynesian cohort.

Oral 26. SheepGenomesDB used to identify genomic features impacted by domestication and selection

James Kijas [1], Marina Naval Sanchez [1], Sean McWilliam [1], Ross Tellam [1], Toni Vuocolo [1], Quan Nguyen [1], Francois Pompanon [2], Pierre Taberlet [2], Wahid Zamani [3], Rudiger Brauning [4], Shannon Clarke [4], Alan McCulloch [4], Noelle Cockett [5], Gary Saunders [6], Hans Daetwyler [7]

[1] CSIRO Agriculture, Brisbane 4067, Australia, [2] Laboratoire d'Ecologie Alpine, Universite Grenoble Alpes, Grenoble, France, [3] Department of Environmental Sciences, Tarbiat Modares University, Noor, Iran, [4] AgResearch Ltd, Invermay, Mosgiel, New Zealand, [5] Utah State University, Logan UT, USA, [6] EMBL-EBI, Cambridge, United Kingdom, [7] Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia.

The SheepGenomesDB project has assembled publically available sheep genomes into a single analytical workflow for the purpose of variant detection. The application of a harmonised pipeline for raw read filtering, mapping and variant detection has been used to identify both raw and high quality variant collections from 453 sheep from 63 populations. A total of 41 million high quality variants are annotated against reference assembly OARv3.1 and have been deposited in the European Variant Archive. We view this dataset as an essential resource for gene mutation discovery, imputation and investigations into the consequences of animal breeding and selection. To explore changing patterns of genetic diversity in response to the domestication process, we analysed a subset of domestic sheep genomes along with their wild ancestor (*O. orientialis*). We compared levels of nucleotide diversity and F_{ST} to search for specific genomic regions that have undergone selection, and find clear evidence for a range of genes previously implicated in sheep and other domesticates (eg *KITLG*, *LCORL*, *PLAG1*). To approach a deeper understanding of the impact of domestication, we defined a collection of genomic features including exons, introns, UTRs, transcription start sites and both proximal and distal regulatory elements. For each, we tested for enrichment of selection sweeps and divergent allele frequency between wild and domestic sheep. We found striking evidence for enrichment of sweeps in proximal features such as promotors and enhancer-genic elements (in contrast to distal and repressive marks). This suggests domestication has preferentially acted on proximal elements that positively regulate gene expression. Finally we predict that domestication has impacted a gene regulatory network governed by OVO and NFKB, both transcription factors known to affect aspects of reproduction.

Oral 27. Sequence variation currently not utilized in genomic selection of New Zealand dairy cattle

Couldrey, C.

Livestock Improvement Corporation (LIC), Hamilton, New Zealand

The introduction of genomic selection to dairy cattle breeding is increasing the rate of genetic gain by both increasing the accuracy of selection and by reducing the generation interval. To date, genomic selection has largely focused on the utilisation of single nucleotide polymorphisms (SNPs) located on autosomes, with very little, if any regard for larger structural variations (SVs) and variation on the sex chromosomes. However, SVs accounting for more sequence variation than SNPs, and the large, gene dense nature of the X chromosome, and together these represent a substantial amount of uncaptured genetic variation. The current exclusion of SV and sex chromosome sequence variation is due to the challenges faced in the detection/genotyping of SVs and the non-autosomal nature of inheritance of the X chromosome which is further complicated by inactivation of one copy of this chromosome in females. It is necessary to identify and characterise SVs and understand X chromosome inactivation in the dairy cow before this genetic variation can be utilised in genomic selection. We have combined next generation sequencing (whole genome sequencing and RNAseq), and PacBio long read sequencing technologies together with SNP chip genotypes in order to explore bovine X chromosome inactivation, and SVs. At the population level X chromosome inactivation in the mammary gland appears to be expressed equally from maternal and paternal copies, however, X inactivation differences were observed in individual cows. As observed in other species, different SV detection approaches resulted in identification of different subsets SVs, and the apparent presence of many false positive SVs, however, combinations of sequencing technologies is allowing us to begin building a true SVs present in NZ dairy cattle. These studies represent a first step to inclusion of SVs and X chromosome such genetic variation has the potential to improve current genomic selection strategies.

Oral 28. High-density genotyping of the New Zealand sheep industry

Suzanne Rowe, Kathryn McRae, Patricia Johnson, Rudiger Brauning, Tracey Van Stijn, Hayley Baird, Duc Lu, Ken Dodds, John McEwan and Shannon Clarke

AgResearch, Invermay Agricultural Centre, Private Bag 5034, Mosgiel, 9053, New Zealand

Genomic tools are available to pinpoint the small genetic differences that produce a variety of traits in livestock. These include SNP chips and genome sequencing approaches. We give the current status for the New Zealand sheep industry, discussing the relative merits of the different approaches, providing examples of how genomic technology has been used within different sectors to improve profitability in sheep. One such example, is the use of genomics to increase our understanding of disease, be it individual differences in the ability to withstand infection, or discovery of causative mutations for inherited disease. Large phenotypic datasets have been used in conjunction with high and low density genotyping to interrogate the sheep genome for regions associated with variability in production and disease traits. An alternative approach is the use of whole genome-sequencing to search for causative mutations for potentially monogenic diseases.

Oral 29. Genomic Analysis of a High Vitamin C Trait in *Actinidia* Hybrids

[1]Seal, A., [2] Knaebel, M. , [1] Bulley, S.M., [2]Laing W., [3] Catanach, A. , [4]Datson, P., [3]McCallum, J.A.,

[1]The New Zealand Institute for Plant & Food Research Limited (PFR), Te Puke, NZ; [2]PFR, Christchurch, NZ; [3]PFR, Palmerston North, NZ; [4]PFR, Mt Albert, Auckland, NZ

Kiwifruit of the genus *Actinidia* are still early in domestication and harbour substantial genetic variation that can be accessed by hybridisation. The species *A. eriantha* is notable for containing very high vitamin C concentrations (> 600 mg/100 g fresh weight) but has small fruit. Analysis of vitamin C concentrations in A. *eriantha* hybrid populations revealed evidence for bimodal segregation of vitamin C contents, consistent with a major gene. We are conducting genetic analysis of this trait using whole-genome sequencing methods to develop strategies for tagging and introgressing consumer traits from these polyploid interspecific crosses, to support breeding of novel fruit types.

To identity genome regions linked with VitC content we conducted Pool-GWAS by sequencing four total DNA pools of N>=20 individuals sampled from eleven tetraploid (*A. chinensis* var. *deliciosa x A. eriantha*) x (*A. chinensis* var. *deliciosa x A. chinensis* var. *chinensis*) families. Pool were as follows: high vitamin C/high fruit weight; high vitamin C/low fruit weight; low vitamin C/high fruit weight; low vitamin C/low fruit weight. Global scans for association of allele frequencies with vitamin C content using the Popoolation2 pipeline revealed significant associations across a 7-Mb region on a single chromosome of the *A. chinensis* reference genome. Analysis of gene and repeat density and of recombination in the *A. chinensis* genome support the hypothesis that this is a region with restricted recombination. This suggests that the high vitamin C trait may be suited to marker-aided selection. Inspection of bam alignments revealed that local haplotypes consistent with different donor genome origins could be readily distinguished, confirming that mapping by sequencing will be a highly effective strategy for analysis of *Actinidia* hybrids. Progress in developing validation strategies, and alternative workflows for poolseq analysis will be discussed.

Oral 30. CpG DNA hypomethylation occurs in human Autosomal Dominant Polycystic Kidney Disease

Bowden, S.[1], Bates, M.[1], Chatterjee, A.[1,2], Rodger, E.J.[1,2], Eccles, M.R.[1,2], Stayner, C.[1]

[1]Department of Pathology, Dunedin School of Medicine, University of Otago, 270 Great King Street, Dunedin 9054, New Zealand. [2]Maurice Wilkins Centre for Molecular Biodiscovery, Level 2, 3A Symonds Street, Auckland 1010, New Zealand.

The DNA methylation status of the genome has been shown to contribute to development and progression of many cancers. However, the role of DNA methylation in Mendelian diseases has not been extensively studied. The process of cystogenesis in the Mendelian disease Autosomal Dominant Polycystic Kidney Disease (ADPKD) shares similarities with many neoplasms, due to the characteristic large fluid-filled cysts that develop within the kidney. Most ADPKD patients carry a mutation in one copy of the *PKD1* gene, but loss of function of the second *PKD1* allele contributes to disease progression. It has been hypothesised that an epigenetic mechanism such as DNA methylation could contribute to a change in *PKD1* expression in the second allele, playing a role in cyst progression.

To understand the role of DNA methylation in ADPKD, whole genome-scale methylomes interrogating a 30-fold enrichment of the CpG islands of human ADPKD kidney tissue (n=2) were generated using reduced representation bisulfite sequencing, and compared to those of normal kidney tissue (n=3). Comparative data between the two groups indicated overall hypomethylation in the ADPKD tissue, which is similar to the methylation alteration often seen in cancer. There were 25 significantly differentially methylated fragments (DMFs) identified, and 52% of these were hypomethylated. DMFs were identified with an ANOVA test with FDR <0.25 and a 15% difference in methylation status. There was an overall enrichment of DMFs overlapping gene bodies compared to other genomic elements. A hypermethylated DMF was also found within the *PKD1* gene body, overlapping an intron/exon junction. Methylome analysis of additional ADPKD tissue samples, and cell lines derived from individual cysts, will be used to expand these data sets, to develop a rationale for targeting DNA methylation in ADPKD as a suitable therapeutic strategy.

Oral 31. Old challenges made harder: Optimising linkage map construction in the modern genomics era

P.L. Wilcox[1], J. Boocock[1], E.J. Telfer[2], N. Graham[2], J. Klapste[2], L. Macdonald[2]

[1]Departments of Biochemistry, and Mathematics and Statistics, University of Otago, Dunedin, New Zealand
[2]Forest Genetics, Scion, 49 Sala Street, Rotorua, New Zealand

Genome-wide genotyping is now affordable for most commercial breeding programmes in New Zealand. However, in non-model organisms such as coniferous trees, the lack of comprehensive reference genome sequences means genetic maps need to be developed based on linkage analyses. Linkage maps have many applications, most recently as resource for genomic selection. Using genotypic data generated from a 44K exome capture panel we have conducted linkage analysis in two outcrossed full sib families of *Pinus radiata* D.Don. A total of 31 165 and 21 319 single nucleotide polymorphism (SNP) markers were identified as segregating in each of the families, respectively, based on observed genotypic variation in parents as well as offspring (*n* = 90 - 93). Clustering analyses (LOD 8.0) showed almost all markers mapped to a cluster with >200 markers. Twelve clusters were identified in both families, equal to the haploid number of chromosomes. Marker order in each cluster was determined using a maximum likelihood algorithm implemented in JoinMap 4.1{Van Ooijen, 2006 #3609}. However, the resulting linkage maps were substantively longer than expected based on previously constructed maps, in part due to the accumulation of errors and difficulties placing inaccurately genotyped loci. To obtain biologically realistic estimates of map length, we reduced the number of markers to 100 per cluster, and then reconstructed maps of the three clusters in both families. This resulted in substantively shorter maps than those constructed using all markers. Decreasing the number of markers to 50 per cluster resulted in further reductions in linkage group lengths, indicating further optimisation is needed. Nonetheless our results show that vast increases in the volume of polymorphisms generated by modern sequencing-based genotyping technologies does not necessarily result in higher quality genetic maps, thus substantively fewer – and carefully selected - markers are needed to produce biologically valid genetic maps.

1. Van Ooijen J: *JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations.* Kyazma BV, Wageningen 2006, **33**:10.1371.

Poster G1. Screening loss of function mutations in fgr/badh2 gene for conferring fragrance in rice varieties of Pakistan

Saddia Galani, Shagufta Sahar, Abid Azhar
Karachi Institute of Biotechnology and Genetic Engineering (KIBGE)
University of Karachi, Karachi, Pakistan

Abstract:
Fragrance in rice is conferred by mutations resulting in loss of function of the fgr/badh2 gene product along with accumulation of 2-acetyl-1-pyrroline (2-AP). This fgr/bdh2 gene corresponds 8bp deletions and three single nucleotide polymorphisms (SNPs) in exon 7. Although, many varieties are reported to be fragrant without having these known mutations, suggesting the involvement of other genes/mutations in rice. It is need of the day to identify the multiple mutations sites in fgr/bdh2 gene for 2AP accumulation and characterize the fragrance in a wide range of genetic resources of rice available for this trait. In this study, loss of function mutations of badh2 genes were characterized in aromatic and non-aromatic rice varieties of Pakistan to resolve the ambiguities linked to the genetic basis of fragrance along with quantification of 2 AP through GC-MS. Genetic analysis revealed that many aromatic varieties exhibited 8-bp deletion and 3 SNPs with strong accumulation of 2-AP. However, amount of 2 AP is highly variable among those varieties indicating involvement of additional factors controlling the intensity of aroma. Moreover, badh2 gene was found to be intact in non-aromatic varieties with no AP accumulation. Thus, it is suggested that this mutation is not only responsible for the 2AP synthesis and accumulation but there are other genetic factors for controlling the development of 2 AP metabolism pathways which are needed to explore. Such investigations will be very important in developing molecular-assisted breeding of aromatic rice with multiple genetic sources of 2AP leading to development of more aromatic varieties.

Key Words:

Badh2 gene, fragrance, loss of function mutation, aromatic rice, 2 AP accumulation

Poster G2. Evaluation of Genomic Selection as a Tool for Pea Breeding

Carpenter, M.A.[1], Goulden, D.S.[1], Thomson, S.J.[1], Woods, C.J.[1], Alspach, P.A.[2], Kenel, F.O.[1], Frew, T.J.[1], Cooper, R.D.[1], Timmerman-Vaughan, G.M.[1]
The New Zealand Institute for Plant & Food Research Limited, [1]Lincoln, NZ, [2]Motueka, NZ.

The pea breeding programme at Plant & Food Research (PFR) uses marker assisted selection (MAS) in breeding for resistance to diseases such as powdery mildew, pea seed-borne mosaic virus and pea enation mosaic virus. This is a practical approach where disease resistance is a single gene trait, but not for diseases such as ascochyta blight, where susceptibility is polygenic. A solution may be found in genomic selection (GS), a promising breeding tool that uses genome-wide molecular marker data to predict phenotypic values and has potential to accelerate breeding for polygenic traits.
A training population of 205 individuals from the PFR pea breeding programme and germplasm collection has been genotyped using Genotyping by Sequencing (GBS) and phenotyped for three traits: MYC (ascochyta blight susceptibility), DTF (days to flowering) and TSWT (thousand seed weight). DTF and TSWT were determined from a single field trial whereas MYC was determined from two replicated randomised field trials in 2013 and 2015. Genomic selection analyses have been conducted using the R package BGLR (Bayesian Generalized Linear Regression) to explore several different models (GBLUP, Bayes A, Bayes B, Bayes C, BRR, BL and RKHS) and various parameters. These comparisons were done via cross-validation using replicated randomised testing sets of 50 lines. Overall the mean prediction accuracy did not vary much between the models. However, different testing/training partitions gave very different results such that the distribution of the prediction accuracies for each model is wide. Typical results for mean prediction accuracy were 0.4-0.5 for MYC and 0.6 for DTF and TSWT. The genotype data have also been used in a GWAS analysis of the three traits, which revealed several SNPs putatively associated with the traits. MYC was associated with SNPs in linkage groups 4 and 7, TSWT with linkage groups 1 and 5, DTF with linkage group 5.

Poster G3. QTL Mapping of Durable Stripe Rust Resistance in Wheat

Paul A. Johnston[1], Catherine Munro[1], Matthew Cromey[2], Soonie Chng[1], Vijitha Meiyalaghan[1], Merle Forbes[1], Steve Shorter[3]

[1]The New Zealand Institute for Plant & Food Research Limited, Gerald Street, Lincoln, New Zealand
[2]The Royal Horticultural Society, RHS Garden Wisley, Surrey, United Kingdom
[3]PGG Wrightson Ltd, Kimihia, Tancreds Road, Lincoln, New Zealand

Durable disease resistance is a major target for plant breeders. The hexaploid wheat cultivar 'Monad' was released in New Zealand in 1993 and has remained resistant to stripe rust (*Puccinia striiformis* f. sp. *tritici*) for more than 20 years.

In order to understand the genetic basis for this resistance, a doubled haploid population was created from a cross between 'Monad' and the susceptible cultivar 'Tiritea'. The population of 219 lines was screened for response to natural infections of stripe rust and phenotyped for seed yield traits both with and without fungicide control. Unprotected plots of 'Monad' still achieved more than 94% of the fungicide controlled values for grain number, head weight and seed weight. In contrast, unprotected plots of 'Tiritea' dropped to 72%, 33% and 45% of their respective fungicide controlled values.

Genotyping was performed using the Wheat Illumina Infinium 90k SNP chip[1] and a genetic linkage map was constructed using the R package 'ASMap' from 8,107 polymorphic SNP markers.

Quantitative trait locus (QTL) mapping was performed using the R package 'qtl' and identified a large effect QTL for stripe rust resistance near the central region of chromosome 2B. This region is known to harbour the resistance genes *Yr5*, *Yr44*, and *Yr53*. An additional QTL for stripe rust resistance was also located on chromosome 7AL, but was only significant for two of the four disease severity scores over 2 years.

Identification of these QTL will allow this resistance to be tracked more easily into elite wheat breeding lines and help to maintain high levels of durable stripe rust resistance.

1. Wang, S., Wong, D., Forrest, K. et al. (2014) *Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array*. Plant Biotechnology Jounral 12:787-796.

Poster G4. Genetic correlation between interleukin 23 receptor (*IL23R)* three variants (*rs7517847*, *rs11209026* and *rs11465804*) and gout susceptibility in a New Zealand population

Shaukat, A. [1] , Leaupepe, K. [1] , Phipps-Green, A.[1], Dalbeth, N.[2], Stamp, L.[3], Hindmarsh, J.H.[4] and Merriman, T.[1]

[1]Department of Biochemistry, University of Otago, Box 56, Dunedin, 9054 New Zealand,
[2]Department of Medicine, University of Auckland, Auckland, New Zealand,
[3]Department of Medicine, University of Otago, Christchurch, PO Box 4345, Christchurch, New Zealand,
[4]Ngati Porou Hauora Charitable Trust, Te Puia Springs, New Zealand.

Gout is an auto-inflammatory disorder caused by deposition of monosodium urate (MSU) crystals in and around tissues in the presence of hyperuricemia. However, there are other factors that control progression from hyperuricemia to gout. The interleukin 23 receptor (*IL23R*) plays a role in promoting inflammation. An earlier study associated the *IL23R* variant *rs7517847* G-allele with the risk of gout in a Han Chinese population, [1] while others reported association of variants *rs11209026* and *rs11465804* with crohn's disease (CD) in Dutch cohort and CD and ulcerative colitis (UC) in Jewish and non-Jewish populations, respectively. [(2,3)] The aim of our study was to test for association between three *IL23R* polymorphisms (*rs7517847*, *rs11209026* and *rs11465804*) and gout in New Zealand (NZ) populations. A total of clinically-ascertained 2604 gout cases[4] and 2092 controls of European and New Zealand Māori and Pacific (Polynesian) ancestry were utilised. Taqman® genotyping of *IL23R* polymorphisms was carried out, followed by multivariate-adjusted association analysis in R 3.2.2 with gout as the outcome.

The minor G-allele of *rs7517847* showed a significant protective association with gout in the Western Polynesian group (OR=0.74, P_{OR}=0.029). However, the other *IL23R* variants were not associated with gout in any NZ population. A significant association was also found between the G-allele of *rs7517847* and the risk of gout in Polynesian males (OR=0.83, P_{OR}=0.03).

In conclusion, our study indicates association of the *IL23R* variant *rs7517847* G-allele with gout in the NZ Western Polynesian population, replicating the Han Chinese finding[1]. However, other variants (*rs11209026* and *rs11465804*) did not exhibit association with gout susceptibility in the European or Polynesian populations. This suggests that the *IL23R rs7517847* variant may be involved in regulating the immune-mediated inflammatory response in gout pathogenesis in Western Polynesians.

1. Liu, S., He, H., Yu, R., Han, L., Wang, C. and Cui, Y., et al (2015). *The rs7517847 polymorphism in the IL-23R gene is associated with gout in a Chinese Han male population*. Modern Rheumatology. 25(3):449-52.

2. Weersma, R.K., Zhernakova, A., Nolte, I.M., Lefebvre, C., Rioux, J.D. and Mulder, F., et al (2008). *ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands*. The American journal of gastroenterology. 103(3):621-7.

3. Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S. and Daly, M.J., et al (2006). *A genome-wide association study identifies IL23R as an inflammatory bowel disease gene*. science. 314(5804):1461-3.

4. Wallace, S.L,. Robinson, H., Masi, A.T., Decker, J.L., McCarty, D.J. and Yu, T.F. (1977). *Preliminary criteria for the classification of the acute arthritis of primary gout*. Arthritis Rheum. 20:895–900.

Poster G5. Genetic association of inflammatory gene *PPARGC1B* with gouty arthritis in a New Zealand population

Shaukat, A.[1] , Phipps-Green, A.[1] , Dalbeth, N.[2] , Stamp. L.[3] , Hindmarsh, J.H.[4] and Merriman, T.[1]

1. Department of Biochemistry, University of Otago, Box 56, Dunedin, 9054 New Zealand
2. Department of Medicine, University of Auckland, Auckland, New Zealand
3. Department of Medicine, University of Otago, Christchurch, PO Box 4345, Christchurch, New Zealand
4. Ngati Porou Hauora Charitable Trust, Te Puia Springs, New Zealand

Gout is an auto-inflammatory arthritis caused by deposition of crystallized monosodium urate in and around tissues. Elevated serum urate levels (hyperuricemia) trigger the formation of monosodium urate crystals (MSU). The genetic basis of hyperuricemia is increasingly well-characterised, however, the genetic basis of the innate immune-mediated inflammatory response in gout pathogenesis is still unclear. Previously, Chang et al (2016) reported association of the PPARGC1B missense variant *rs45520937* A-allele with gout susceptibility in a Taiwan Han Chinese population.

Our aim was to replicate this finding in a total of 2680 clinically ascertained gout cases[1] and 2195 controls from the New Zealand Polynesian (Māori and Pacific) population, and Europeans from New Zealand and Europe. Taqman® genotyping of PPARGC1B missense variant *rs45520937* was undertaken, followed by multivariate-adjusted association analysis in R 3.2.2 with gout as outcome.

We found a significant association of the PPARGC1B *rs45520937* A-allele with gout risk in the Polynesian sample set (OR= 1.17, P=0.02*) but not in Europeans (OR= 0.96, P=0.80). A stronger association was observed in males of higher Polynesian ancestry (OR= 1.47, P_{OR} =0.01*).

Our findings support a potential role of the inflammatory mediator PPARGC1B in the pathogenesis of gout.

References

1. Wallace, S.L., Robinson, H., Masi, A.T., Decker, J.L., McCarty D.J., Yu, T.F (1977) *Preliminary criteria for the classification of the acute arthritis of primary gout*. Arthritis Rheum 20:895–900.

Poster G6. Open Source, Open Data and Open Science Policy: Opportunities for New Zealand Genomics Research

Elshire, R.J. [1]

[1]The Elshire Group Limited, Palmerston North, NZ.

Genomics research is underpinned by Free / Libre and Open Source software – from the operating system level to the individual tools used to analyze data. Genomic scientists contribute to large open respositories of sequence data. The global move to open science (which includes transparancy, reproducibility and availability) is gaining traction in New Zealand public policy. The New Zealand Goverment Open Access and Licensing (NZGOAL) framework and soon the Software Extension to that framework (NZGOAL-SE) apply to all Crown Entities.

Getting beyond the buzzwords we will discuss the benefits of working in an open fashion with international examples appropriate to our MapNet community. We will describe the process by which the Software Extension was developed openly and collaboratively this year. We will also present a recap of the efforts of the Biospectra-by-Sequencing project to work openly and collaboratively for Nzinc., giving an update on the status of the project and plans for the future.

Poster G7. New molecular tools for interspecific hybridization of *Gentiana* sp.

Tonya Frew[1], Claudia Wiedow[2], Gail Timmerman-Vaughan[1], Ryohei Kaji [2], David Chagne[2], Ed Morgan[2], Keith Funnell[2]

[1] The New Zealand Institute for Plant & Food Research Limited, Private Bag 4704, Christchurch, NZ
[2] The New Zealand Institute for Plant & Food Research Limited, Private Bag 11-600, Palmerston North, NZ

There are over 400 species of gentians, of which only a few are of economic importance as ornamental crops. Hot pink, red, purple and yellow flowered gentians are among the new genotypes and cultivars bred and developed by The New Zealand Institute for Plant & Food Research Limited (PFR). Developed from wide crosses among various *Gentiana* spp. these hybrids offer the New Zealand and international flower industry new options: different flower colours and forms tap into international demand for novel cut flowers, potted and landscape plants. Within the PFR breeding programme, new molecular tools are being developed and applied; for example, to confirm hybrid status of seedlings generated. Our progress in developing these tools is presented here.

Preliminary studies characterising *Gentiana* spp. using Sequence Characterized Amplified Regions (SCAR) markers, produced some usable assays. One SCAR primer set was identified to be specific for *G. mirandae* and another for *G. scabra*. Subsequent tests across a small number of hybrids verified that such molecular tools can be used to differentiate individuals within these *Gentiana* species. We anticipate further tests using publicly available chloroplastic markers and the High Resolution Melting (HRM) technique as well as microsatellite markers resolved using capillary electrophoresis will help to discriminate among *G. dinarica*, *G. angustifolia 'Alba', G. wilsonii, G. parryi, G. acaulis, G. freyniana, G. przewalskii, G. sino-ornata, G. paradoxa, G. lutea, G. mirandae, G. triflora, G scabra, G. septemfida* and their hybrids.

Poster G8. Genomics in New Zealand Forestry

Dungey, H.[1], Telfer, E.[1], Graham, N.[1], Shearer, A.[1,2], Li, Y[1], Klapste, J[1], Murray, M.[1,3], Macdonald, L.[1]
[1] Forest Genetics, Scion, 49 Sala St, Rotorua, New Zealand
[2]University of Canterbury, 20 Kirkwood Ave, Christchurch 8041, New Zealand
[3] University of Waikato, Gate 1 Knighton Road, Hamilton 3240, New Zealand

Tree breeding in New Zealand has initiated the move to genomics. Tree breeding programmes in NZ are currently all with exotic species, with recorded pedigree depths of between two and four generations from unimproved. The combination of well-structured populations and the ability to clonally test across a number of environments is an advantage in these programmes. However, the limited number of generations means that these populations are still highly heterogeneous. Furthermore, most experimental trials are planted with wind-pollinated trees and have long generation times, both of which present a significant challenge to breeding programmes.

Genomics can help address many of the challenges faced by tree breeders: providing marker-based relationships in wind-pollinated populations where only one parent is known, selecting for wood quality traits that are very expensive to measure, and accelerating the rate of delivery of genetic gain. The Radiata Pine Breeding Company, in partnership with Scion, is leading genomics research for radiata pine, and have developed a 44K exome capture panel to enable the implementation of genomic selection. The development of such extensive genomics resources has historically been unachievable for smaller breeding programmes, such as those for eucalypts and Douglas-fir. However, recent years have seen internationally-developed high density SNP arrays become available for these species {Howe, 2013 #1990;Silva-Junior, 2015 #2383}, which is enabling a paradigm shift in our approach to breeding.

We report on the progress being made towards the introduction of genomics and genomics research in radiata pine, Douglas-fir and in a number of eucalypt breeding programmes. We present early results from the existing multi-species *Eucalyptus* SNP chip EuCHIP60K that has been used in the first application of genomics to a eucalypt breeding programme in NZ. We also report on our plans to effectively incorporate genomics for a number of other forest-based species.

1. Howe, G.T., et al., *A SNP resource for Douglas-fir: De novo transcriptome assembly and SNP detection and validation.* BMC Genomics, 2013: p. 137.
2. Silva-Junior, O.B., D.A. Faria, and D. Grattapaglia, *A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species.* New Phytol, 2015. **206**(4): p. 1527-40.

Poster G9. Giving GWAS the Boot: Teasing out Disease Associations by Bootstrap Sub-sampling

Eccles, D.A.[1]

[1]Gringene Bioinformatics, Wellington, NZ

Genome-wide Association Studies are carried out on a large number of genetic variants in a large number of people, allowing the detection of small genetic effects that are associated with a trait. Natural variation of genotypes within populations means that any particular sample from the population may not represent the true genotype frequencies within that population. This may lead to the observation of marker-disease associations when no such association exists.

A bootstrap population sub-sampling technique can reduce the influence of allele frequency variation in producing false-positive results for particular samplings of the population. In order to utilise bioinformatics in the service of a serious disease, this sub-sampling method was applied to the Type 1 Diabetes dataset from the Wellcome Trust Case Control Consortium in order to evaluate its effectiveness. These results are compared to results from both a low-heritability dataset and a high-heritability dataset, suggesting that the usual noise inherent in GWAS studies might be able to be removed by this method.

The generation of a panel of validated group-specific markers is possible even when using a low-density marker set and small sample group sizes. This method is likely applicable to other genome-wide studies, and provides one way in which false positive associations could be quickly excluded from candidate marker sets.

Genomics Summary of Abstracts for the Poster Session

No.	Title	Presenter	Institutions
G1	Screening loss of function mutations in fgr/badh2 gene for conferring fragrance in rice varieties of Pakistan	Saddia Galani, Shagufta Sahar, Abid Azhar	Karachi Institute of Biotechnology and Genetic Engineering (KIBGE) University of Karachi, Karachi, Pakistan
G2	Evaluation of Genomic Selection as a Tool for Pea Breeding	Carpenter, M.A.[1], Goulden, D.S.[1], Thomson, S.J.[1], Woods, C.J.[1], Alspach, P.A.[2], Kenel, F.O.[1], Frew, T.J.[1], Cooper, R.D.[1], Timmerman-Vaughan, G.M.[1]	The New Zealand Institute for Plant & Food Research Limited, [1]Lincoln, NZ, [2]Motueka, NZ.
G3	QTL Mapping of Durable Stripe Rust Resistance in Wheat	Paul A. Johnston[1], Catherine Munro[1], Matthew Cromey[2], Soonie Chng[1], Vijitha Meiyalaghan[1], Merle Forbes[1], Steve Shorter[3]	[1]The New Zealand Institute for Plant & Food Research Limited, NZ [2]The Royal Horticultural Society, UK [3]PGG Wrightson Ltd, NZ
G4	Genetic correlation between interleukin 23 receptor (*IL23R)* three variants (*rs7517847*, *rs11209026* and *rs11465804*) and gout susceptibility in a New Zealand population	Shaukat, A. [1] , Leaupepe, K. [1] , Phipps-Green, A.[1], Dalbeth, N.[2], Stamp, L.[3], Hindmarsh, J.H.[4] and Merriman, T.[1]	[1]Department of Biochemistry, University of Otago, NZ [2]Department of Medicine, University of Auckland, NZ [3]Department of Medicine, University of Otago, NZ [4]Ngati Porou Hauora Charitable Trust, NZ
G5	Genetic association of inflammatory gene *PPARGC1B* with gouty arthritis in a New Zealand population	Shaukat, A.[1] , Phipps-Green, A.[1] , Dalbeth, N.[2] , Stamp. L.[3] , Hindmarsh, J.H.[4] and Merriman, T.[1]	[1]Department of Biochemistry, University of Otago, NZ [2]Department of Medicine, University of Auckland, NZ [3]Department of Medicine, University of Otago, NZ [2]Ngati Porou Hauora Charitable Trust, NZ
G6	Open Source, Open Data and Open Science Policy: Opportunities for New Zealand Genomics Research	Elshire, R.J. [1]	[1]The Elshire Group Limited, Palmerston North, NZ
G7	New molecular tools for interspecific hybridization of *Gentiana* sp.	Tonya Frew[1], Claudia Wiedow[2], Gail Timmerman-Vaughan[1], Ryohei Kaji [2], David Chagne[2], Ed Morgan[2], Keith Funnell[2]	[1] The New Zealand Institute for **Plant** & **Food Research** Limited, Christchurch, NZ [2] The New Zealand Institute for **Plant** & **Food Research** Limited, Palmerston North, NZ
G8	Genomics in New Zealand Forestry	Dungey, H.[1], Telfer, E.[1], Graham, N.[1], Shearer, A.[1,2], Li, Y[1], Klapste, J[1], Murray, M.[1,3], Macdonald, L.[1]	[1] Forest Genetics, Scion, Rotorua, NZ [2]University of Canterbury, Christchurch, NZ [3] University of Waikato, Hamilton, NZ
G9	Giving GWAS the Boot: Teasing out Disease Associations by Bootstrap Sub-sampling	Eccles, D.A.[1]	[1]Gringene Bioinformatics, Wellington, NZ