Abstracts for MapNet 2017 (2nd and 3rd November)

1. Session: Bioinformatics

Major threads of the cotton genome

Joshua A. Udall
Brigham Young University, Provo, UT, USA
jaudall@byu.edu

Every genome has its own story. In cotton, the four major threads (grand challenges) are polyploidy, plant genome evolution, domestication, and plant productivity. Of these, we have recently been focused on domestication and plant genome evolution. In the first thread, genome resequencing of cotton germplasm can uncover the natural history of cotton from its origination to its domestication. Numerous accessions were selected from the USDA collection for resequencing with high coverage, including wild, federal and cultivar tetraploid and diploid cotton. A comprehensive variation map for more than 1,000 cotton accessions has been constructed, including SNPs and InDels. These resources will assist cotton breeders efforts to improve the fiber quality, disease resistance, and yield of modern cotton varieties. In the second thread, we have combined three different genome technologies to create robust whole genome sequence assemblies for representatives of all of the genome groups in *Gossypium* (A-, B-, C-, D-, F-, G-, and K-genomes). These assemblies are being used for comparative, structural genomics. First, PacBio sequence of the diploid genomes is assembled. Second, the sequence contigs are scaffolded using Hi-C. Third, nick-based physical map data is used to validate the sequence assembly, then to correct the order and orientation of scaffolded contigs. Generating a chromosome-level sequence from this final product leverages the best features of each technology.

Getting ahead in the cloud: predicting compute requirements for the 25 Gb *Pinus radiata* megagenome assembly

Shane Sturrock

Forest Industry Informatics, Scion, Rotorua, NZ
shane.sturrock@scionresearch.com
Shane Sturrock[1] , Emily Telfer[2] , Richard Winkworth[3] , Natalie Graham[2] , Lucy Macdonald[1] , Tancred Frickey[1] , Phillip Wilcox[4] and Heidi Dungey[2]

[1]Forest Industry Informatics, Scion, Rotorua, New Zealand,
[2] Forest Genetics , Scion, Rotorua, New Zealand,
[3]Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand,
[4]Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand.

A robustly assembled and well annotated reference genome can facilitate a range of research activities, including marker development for breeding and selection, functional genomics for understanding complex gene interactions, and comparative evolutionary studies. Conifer genomes dwarf those of other tree species. For example, the *Pinus taeda* genome is 21 Gb in size compared to 420 and 640 Mb for *Populus trichocarpa*and *Eucalyptus grandis*, respectively. Much of this size difference reflects paleo-polyploidy in conifers, a phenomenon that has resulted in diverse gene families and massive amounts of repetitive DNA. Development of an annotated resource for the 25 Gb *Pinus radiata*, therefore, represents a significant commitment. Indeed, even using the closely related *P. taeda* as a reference, the computational resources needed to assemble these data have outstripped those available in New Zealand. Using a fraction of our available data, we performed scaling tests on cloud resources to define with high accuracy the server specifications needed for completing the assembly and annotation. These scaling tests have also allowed us to compare costs between continuing with cloud resources and purchasing a standalone resource. As a result, we have now purchased a high end server and have initiated our second iteration assembly to include PacBio long read data. Here we present an update on our progress towards annotated reference genomes for the nuclear and organellar compartments of *P. radiata*.

Bioinformatics analysis methods developed for the sugarcane mycome

Ruy Jauregui
AgResearch
Ruy.Jauregui@agresearch.co.nz
Ruy Jauregui[1], Lilia Carvalhais[2], Priya Joyce[2], Shamsul Buhiyan[2], Siva Ganesh[1] and Stuart Card[1]

[1]Plan-Fungal Interactions, AgResearch, Palmerston North, New Zealand
[2]Molecular Genetics, Sugar Research Australia, Brisbane, Australia

The bioinformatics analysis of amplicon-based high throughput sequencing data from plant-associated fungal communities presents specific challenges. The variability in the communities' complexity in different plant organs/tissues, and the similarity of plant and fungi-derived marker sequences commonly used to classify fungi (ITS1, ITS2) pose fundamental issues to the development of a semi-automated analysis pipeline. Here we describe the steps of a novel discovery pipeline that includes data quality filtering and trimming, clustering, background and contamination evaluation and cleaning. Taxonomic annotation and operational taxonomic unit abundance evaluation, as well as principles of statistical analysis specifically developed to analyse amplicons of the sugarcane mycome are also included. These methods can easily be adapted to other studies of plant-associated fungal communities. The pipeline presented here has been applied in the analysis of the mycome in samples recently obtained from sugarcane and its wild relatives collected in Papua New Guinea and Australia.

Variations in mutation accumulation in grape chimeras revealed through whole genome analysis

Dr Susan Thomson
Plant & Food Research
susan.thomson@plantandfood.co.nz
Susan Thomson[1], Tim Millar[1], Darrell Lizamore[2], Philippa Barrell[1] and Chris Winefield[2]

[1]The New Zealand Institute for Plant & Food Research,Canterbury Agricultural & Science Centre, Lincoln, 7608, New Zealand
[2]Lincoln University, Lincoln 7647, Canterbury, New Zealand

Grape (*vitis vinifera*) is a domesticated crop that has been clonally propagated for centuries. The varietal Pinot noir is one of the oldest, grown for wine production for at least 700 years. Chimerism and the accumulation of genome mutations is known to exist within clonally propagated grape vine. To date this has been shown through the application of low resolution SSR markers to selected loci. Through the use of callus cell culture derived from diploid anther material, we have regenerated several plants with reduced chimerism. We present whole genome variant analysis (sequence polymorphisms as well as transposon fingerprinting) of parent meristem and 6 regenerants highlighting distinct genome regions that may be under selective pressure through vegetative propagation.

Report from the Bioinformatics Computational Science Workshop

Robyn Johnston

The Elshire Group Limited

Robyn@ElshireGroup.co.nz

David Winter[1], Robyn Johnston[2], Rob Elshire[2], Helge Dzierzon[3] and Jenny Draper[4]

[1] Institute of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand

[2] The Elshire Group, Palmerston North, New Zealand

[3] Plant and Food Research - Rangahau Ahumara Kai, Palmerston North 4474, New Zealand

[4] Institute of Environmental Science and Research Limited [ESR], Kenepuru, Porirua 5022, New Zealand

Bioinformatics and Computational Science underpin much of modern biological research. They are scientific disciplines unto themselves with their own challenges and open research questions. The Palmerston North Bioinformatics Meeting group has been exploring the question 'Does our software do what we expect it to?' Unfortunately, what we find is often not what we would hope. We have put together a workshop to explore the current state of practice in scientific computing and the issues researchers encounter. The workshop explores how data generation in the lab affects downstream analyses and highlights potential pitfalls in bioinformatic analyses. It then presents demonstrated methods to improve the reliability of the code we use and write, and resources available within New Zealand to help improve computational science practices. A focal point of the workshop is the panel discussion, the purpose of which is to spark a conversation about the common challenges in computational science and identify ways to improve our research. This talk will summarise the key points from the workshop with an emphasis on how they might be applied. The Palmerston North Bioinformatics Meeting is a monthly gathering to discuss the practice and profession of bioinformatic and computational science. Meetings have been held every month since September 2015. It is an informal forum for sharing ideas and learning from each other - sometimes it is like a journal club, other times it is like a practical lab meeting where one presents a particularly sticky problem to solve. It can even function as a research collective as it has done in the case of the workshop. The meeting is open to all people with an interest in bioinformatics science and an inclination for clear, honest, and kind communication about the practice and profession of computational research. We welcome new participants, regardless of experience level.

2. Session: Technologies for Genomes

Long read nanopore sequencing in the laboratory

Simone Cree

University of Otago, Christchurch
simone.macmil@otago.ac.nz
Simone Cree[(1)], Simran Maggo[(1)], Sandy Slow [(1)], Kim Ton[(1)], Lucy de Jong[(1)], Trevor Anderson[(1)], John Pearson[(2)], Allison Miller[(1)], Logan Walker[(1)], David Murdoch[(1)] and Martin Kennedy[(1)]

[(1)] Department of Pathology, University of Otago, Christchurch, NZ
[(2)] Biostatistics and Computational Biology Unit, University of Otago, Christchurch, New Zealand

The MinION device, developed by Oxford Nanopore Technologies Ltd (UK), has gained its reputation as the world's smallest DNA sequencing machine. We trialled the MinION in our laboratory for three years; mainly in the area of pharmacogenetics, as well as cDNA analysis and bacterial genome assembly. DNA barcoding was used to detect polymorphisms in long (7-8 Kb) and short (200-300 bp) PCR amplicons of important pharmacogenes including full length CYP2D6, CYP1A2 and CYP2A6 genes. Our attempt at genotyping the Human Leukocyte Antigen (HLA-B) locus led to high resolution assignment of alleles for 49 DNA samples[1]. We also successfully detected novel isoforms of BRCA1 transcripts using long range PCR amplification[2] and de novo assembled the bacterial genome of a *Legionella sainthelensi* isolate obtained from a patient at the CDHB.

Capture-based resequencing in polyploid potato

Samantha Baldwin

New Zealand Institute for Plant and Food Research
samantha.baldwin@plantandfood.co.nz

Samantha Baldwin [1], Katrina Monaghan [1], John Anderson [2], Steve Lewthwaite [2], Mark Paget [1], Marcus Davey [3], Linley Jesson [4], John McCallum [1] and Susan Thomson [1]

[1] NZ Institute for Plant and Food Research Ltd, Private Bag 4704, Christchurch Mail Centre, Christchurch, 8140
[2] NZ Institute for Plant and Food Research Ltd, Cronin Rd, RD 1, Pukekohe 2676
[3] NZ Institute for Plant and Food Research Ltd, 412 No 1 Road, RD 2, Te Puke 3182
[4] NZ Institute for Plant and Food Research Ltd, Private Bag 1401, Havelock North, 4157

Cultivated potato (*Solanum tuberosum*) is an outcrossing autotetraploid (2n=4x) with a haploid genome size of 840 Mb. Our breeding programme has been developing populations and cultivars adapted for various applications from fresh to processing types, with common founder parents often used. Identification of the alleles and haplotypes present in the programme and their association with traits under selection is an important part of the genomic breeding process. The techniques used require accurate assessment of what alleles are present and preferably at what dosage. Autopolyploids require higher sequence coverage than diploids to estimate allele and dosages confidently. Current sequence-based genotyping in autopolyploid potato makes this uneconomical within a breeding programme. Therefore a reduced representation technique was developed using capture-based resequencing to allow consistent resampling over genomic regions. Tetraploid individuals and pooled families were genotyped using ~20,000 baits designed on unique sequences present across the diploid potato reference genome. These data were used to detect changes in allele frequencies after seedling culling.

Sharing Improvements in Genotyping-by-Sequencing

Rob Elshire
The Elshire Group Limited
rob@elshiregroup.co.nz
Rob Elshire[(1)], Spencer Doo[(1)] and Robyn Johnston[(1)]

[(1)]The Elshire Group Limited, Palmerston North, New Zealand

Genotyping-by-Sequencing (GBS) is a robust, cost effective, high-throughput method for detecting differences in DNA sequence between individuals. The original method is comprised of laboratory protocols and software analysis tools that were developed by Rob Elshire and colleagues at Cornell University in a fully Open Source way. Subsequent to publication in PloSONE in 2011, many other researchers have modified GBS to tune it for particular applications or genetic architectures. GBS was originally developed to underpin genomic selection approaches to maize breeding. It has since been adopted around the world by plant and animal breeders, as well as ecologists and conservation biologists. Similarly in New Zealand, GBS is being used to great effect across the primary industries, and ecological and conservation genomics. While classic GBS is a powerful and cost effective technique, the many modifications that have been made clearly illustrate that there is plenty of room for ongoing improvement. In 2016, Tokamaneh et al compared seven GBS analysis pipelines and two sequencing technologies. One key finding is that genotyping accuracy improves with increased sequence length. The paper also made apparent that there is a need to be able to evaluate the many different pipelines that have sprung up to analyse GBS data. The Biospectra-by-Sequencing (BBS) project is a New Zealand born multi-institutional effort to develop reproducible analysis pipelines and ancillary information for genomics scientists. The BBS project presented a poster on the many GBS wet lab protocols and analysis pipelines at the 2017 Plant Animal Genome Meeting (Davy et al). Feedback from conference attendees indicated a clear need for ways to evaluate both the wet lab methods and the analysis pipelines for fitness-for-purpose for a given research question. In this talk we share the improvements that The Elshire Group has made in adapting the GBS method to the Illumina HiSeq XTen platform. These improvements increase accuracy by providing longer (and paired end) reads and reduce the cost of GBS experiments substantially. In keeping with our Open Source model, we share the know how so that you can take advantage of these improvements in your own research right away. We also demonstrate the utility of an automated analysis and reporting framework for GBS pipelines. This work builds on the efforts of the BBS project by utilising a modern data science approach with Dragonfly Data Science.

Out of the freezer and onto the farm: the genomes of white clover and its progenitors as an example of allopolyploid adaptation.

Andrew Griffiths

AgResearch

andrew.griffiths@agresearch.co.nz

Andrew Griffiths[1], Roger Moraga[1], Vikas Gupta[2], Timothy Bilton[3], Marni Tausen[2][4], Matthew A. Campbell[5][6], Rachael Ashby[3], Istvan Nagy[2], Anar Khan[3][7], Craig Anderson[1], Benjamin Franzmayr[1], Kerry Hancock[1], Niraj Shah[2], Murray P. Cox[5], Torben Asp[2], Thomas Mailund[4], Mikkel H. Schierup[4] and Stig Uggerhøj Andersen[2]

[1]AgResearch, Grasslands Research Centre, Private Bag 11008, Palmerston North 4442, New Zealand
[2]Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark
[3]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand
[4]Bioinformatics Research Centre, Aarhus University, C. F. Møllers Allé 8, 8000 Aarhus C, Denmark
[5]Institute of Fundamental Sciences, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand
[6]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA (current address)
[7]Dunedin School of Medicine, University of Otago, PO Box 56, Dunedin 9054, New Zealand (current address)

The merging of distinct genomes, allopolyploidisation, generates adaptive potential through increased genetic diversity and access to 'genomic toolboxes' from the contributing genomes. White clover, *Trifolium repens*, is an allotetraploid (2n = 4x = 16) forage crop found throughout temperate grasslands, and is derived from two diploid progenitors: *T. occidentale* and *T. pallescens*, each confined to markedly different coastal and montane niches, respectively. Genome and transcriptome sequencing and subsequent assembly of this species complex, has provided a wealth of data to gain insight into the genesis and evolution of white clover. We have confirmed the progenitors, and shown that the progenitor subgenomes within white clover have largely retained their integrity and gene expression activity following allopolyploidisation. Furthermore, we show that this hybridisation event occurred ~45,000 years ago during the depths of the last glaciation at a time when the European progenitor ranges (coastal and montane) likely overlapped. White clover, therefore, represents a clear example of allopolyploidy-facilitated niche expansion, where the two progenitor genomes expanded from disparate and highly specialised European habitats to a ubiquitous global presence. Perhaps underpinning this evolutionary success, we found high polymorphism levels in white clover, demonstrating diversity carry-over from its progenitors. Furthermore, we have also found evidence of tissue-specific expression switching between subgenome copies of genes involved in flavonoid biosynthesis, a key pathway involved in adaptive traits such as plant/microbial interactions.

Manual annotation of the Kiwifruit genome

Dr Robert Schaffer

Plant & Food Research
robert.schaffer@plantandfood.co.nz

Sarah Pilkington[1], Ross Crowhurst[1], Elena Hilario[1], Simona Nardozza[1], Lena Fraser[1], Yongyan Peng[1][2], Kularajathevan Gunaseelan[1], Robert Simpson[3], Jibran Tahir[3], Simon C. Deroles[3], Kerry Templeton[1], Zhiwei Luo[1], Marcus Davy[4], Canhong Cheng[1], Mark McNeilage[1], Davide Scaglione[5], Yifei Liu[6], Qiong Zhang[7], Paul Datson[1], Nihal De Silva[1], Susan E. Gardiner[3], Heather Bassett[3], David Chagné[3], John McCallum[8], Helge Dzierzon[3], Cecilia Deng[1], Yen-Yi Wang[1], Lorna Barron[1], Kelvina Manako[1], Judith Bowen[1], Toshi M. Foster[3], Zoe A. Erridge[3], Heather Tiffin[3], Chethi N. Waite[3], Kevin M. Davies[3], Ella P. Grierson[3], William A. Laing[3], Rebecca Kirk[1], Xiuyin Chen[1], Marion Wood[1], Mirco Montefiori[1], David A. Brummell[3], Kathy E. Schwinn[3], Andrew Catanach[8], Christina Fullerton[1], Dawei Li[7], Sathiyamoorthy Meiyalaghan[8], Niels Nieuwenhuizen[1], Nicola Read[2], Roneel Prakash[1], Don Hunter[3], Huaibi Zhang[3], Marian McKenzie[3], Mareike Knäbel[3], Alastair Harris[2], Andrew C. Allan[1][2], Andrew Gleave[1], Angela Chen[2], Bart J Janssen[1], Blue Plunkett[1], Charles Dwamena[1], Charlotte Voogd[1], Davin Leif[1][2], Declan Lafferty[2], Edwige Souleyre[1], Erika Varkonyi-Gasic[1], Francesco Gambi[1], Jenny Hanley[2], Jia-Long Yao[1], Joey Cheung[2], Karine M. David[2], Ben Warren[1], Ken Marsh[1], Kimberley C. Snowden[1], Kui Lin-Wang[1], Lara Brian[1], Marcela Martinez-Sanchez[1], Mindy Wang[1], Nadeesha Ileperuma[1], Nikolai Macnee[1], Robert Campin[1], Peter McAtee[1], Revel S M Drummond[1], Richard V. Espley[1], Hilary S. Ireland[1], Rongmei Wu[1], Ross G. Atkinson[1], Sakuntala Karunairetnam[1], Sean Bulley[4], Shayhan Chunkath[2], Zac Hanley[1], Roy Storey[4], Amali H. Thrimawithana[1], Susan Thomson[8], Charles David[8], Raffaele Testolin[5][9], Hongwen Huang[6][7], Roger P. Hellens[10] and Robert J. Schaffer[1][2]

[1] The New Zealand Institute for Plant & Food Research Ltd [PFR], Private Bag 92169, Auckland 1142, New Zealand.
[2] School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
[3] PFR, Private Bag 11600, Palmerston North 4442, New Zealand.
[4] PFR, 412 No1 Road, Te Puke 3182, Bay of Plenty, New Zealand.
[5] IGA Technology Services, Parco Scientifico e Tecnologico, Udine, Italy
[6] South China Botanic Gardens, Chinese Academy of Sciences, Guangzhou 510650, Guangdong, China
[7] Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China
[8] PFR, Private Bag 4704, Christchurch 8140, New Zealand
[9] Department of Agricultural and Environmental Sciences, University of Udine, Via delle Scienze 208, 33100 Udine, Italy
[10] Queensland University of Technology [QUT], Brisbane 4001, Australia

With the advent of large numbers of low quality “draft” genomes the research communities in non-model organisms rely heavily on computationally annotated genes. The huge variation in the quality of these draft genomes and the predicted gene lists affects research and development. The draft Actinidia (Kiwifruit) genome was a big step forward in our understanding of this important horticultural crop. Typical of draft genomes, a large proportion of the sequence was not assigned to chromosomes, and the gene list missed key

published genes. To address this we developed an international consortium firstly to improve the genome construction and secondly to manually annotate the whole genome. Manual annotation was performed using the WebApollo software and employed tracks of computationally predicted genes, EST sequences and mRNA-seq data from a variety of different tissues. In total 98 researchers from around the world produced evidence-based gene models for over 33000 genes. Our subsequent analysis showed that the majority of computational gene predictions in kiwifruit do not align to the manually edited genes. In particular, the computationally generated genes did not accurately predict intron-exon structures or translation start and stop regions. Non canonical splice sites were particularly poorly annotated. From this exercise we now have a second generation genome to start to understand the biological processes associated with this plant and a method for the scientific community to improve the accuracy and utility of draft genomes.

Use of a Selfcompatible Diploid Potato Genome Assembly and a Mutant Collection for Forward Genetic Studies

Elena Lopez-Girona

Plant and Food Research Institute

elena.lopez-girona@plantandfood.co.nz

Elena Lopez-Girona[1], Pirita Paajanen[2], Georger Kettleborough[2] , Matthew Clark[2] and Glenn Bryan[3]

[1]Molecular and Digital Breeding, Plant & Food Research,Fitzherbert Science Centre, Batchelar Road, Palmerston North 4442, New Zealand

[2]Earlham Institute, Norwich Research Park Innovation Centre, Colney Ln, Norwich NR4 7UZ,Norwich, United Kingdom

[3]Cell and Molecular Sciences, The James Hutton Institute, Invergowire, Dundee, DD2 5DA, Scotland, United Kingdom

We have sequenced the genome of the self-compatible tuber-bearing wild Mexican diploid potato species *Solanum verrucosum* (VER54). The genome assembly is based on several sequencing and physical mapping technologies, such as Illumina short reads, PacBio long reads, BioNano Genomics optical maps and Dovetail artificial Hi-C data to allow accurate assembly, and to permit comparisons of genome assembly methods. The assembled genome was evaluated for completeness by checking the presence of CEGMA core eukaryotic genes and validated locally by the alignment of the 96 BAC-end sequences. Furthermore, Genotyping by Sequencing (GBS) of a backcross *S. verrucosum* population (VER54xVER3939) has been used to anchor the assembly into pseudomolecules. The sequenced genotype is also being used as a model for forward genetic studies. Mutant collections are being produced using ethyl methanesulfonate (EMS) and gamma irradiation. The objective is to combine the two elements of the project (i.e de novo assembly and a mutant population) together to study mutants carrying interesting traits that can be physically located by sequencing and mapping them in the genome assembly. The generation of a genome sequence for *S. verrucosum* and a mutant collection are highly valuable genetic resources. The VER54 inbred genotype can serve as a useful genetic model for further studies about comparative genome structure and adaptive within the *Solanaceae*.*S.verrucosum* is a likely genome donor to the wild Mexican polyploid species and therefore very useful genome in further studies of this important species group.

Chromosome-scale scaffolding of the black raspberry (Rubus occidentalis L.) genome based on chromatin interaction data

Dr Helge Dzierzon and Rubina Jibran
Plant and Food research
rubina.jibran@plantandfood.co.nz
Rubina Jibran, Helge Dzierzon, Kevin Davies, Toshi Foster and David Chagné

The New Zealand Institute for Plant & Food Research Limited, Private Bag 11600, Palmerston North 4474, New Zealand

Black raspberry (Rubus occidentalis L.) is a niche fruit crop valued for its flavour and potential health benefits. The improvement of fruit and cane characteristics via molecular breeding technologies has been hindered by the lack of a high-quality reference genome. The recently released draft genome for black raspberry (ORUS 4115-3) lacks assembly of scaffolds to chromosome scale. We used high-throughput chromatin conformation capture (Hi-C) and Proximity-Guided Assembly (PGA) to cluster and order 9,650 out of 11,936 contigs of this draft genome assembly into seven pseudo-chromosomes. The seven pseudo-chromosomes cover ~97.2% of the total contig length (~223.8 Mb). Locating existing genetic markers on the physical map resolved multiple discrepancies in marker order on the genetic map. Centromeric regions were inferred from recombination frequencies of genetic markers, alignment of 303 bp centromeric sequence with the PGA, and heat map showing the physical contact matrix over the entire genome. We demonstrate a high degree of synteny between each of the seven chromosomes of black raspberry and a high-quality reference genome for strawberry (Fragaria vesca L.) assembled using only PacBio long-read sequences. We conclude that that PGA is a cost-effective and rapid method of generating chromosome-scale assemblies from Illumina short-read sequencing data.

Development of Genome Analysis Resources in Tetraploid *Actinidia chinensis*

John McCallum

Plant & Food Research

John.McCallum@plantandfood.co.nz

John McCallum[1], Luis Gea[3], Russell Lowe[3], Amardeep Nath[3], Susan Thomson[1], Meeghan Pither-Joyce[1], Jibran Tahir[2], Tim Millar[1], Samantha Baldwin[1]

[1] The New Zealand Institute for Plant & Food Research Limited, Lincoln, New Zealand
[2] The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand
[3] The New Zealand Institute for Plant & Food Research Limited, Te Puke, New Zealand

Gold-fleshed kiwifruit are among New Zealands most profitable and distinctive horticultural exports. Diploid A. chinensis such as 'Hort16A' proved highly susceptible to the vine-killing disease PSA-V but survival of the industry was ensured by selection of the more tolerant tetraploid variety 'Gold3'. We describe development of resources to enable genomic analysis to understand domestication and historical selection targets in these pedigrees, and enable forward selection strategies. A haplotype reference for these families is being developed by whole genome sequencing of founders and of trios from more advanced selections, using haplotyping strategies based on assemblies of the diploid variety 'Red5'. Variant data from whole-genome sequencing of diploid A. chinensis was used to select a 10k set of baits for exome sequencing. We will describe preliminary whole-genome population genetics analyses based on targeted and whole-genome sequencing of founders and three advanced families.

3. Session: Environmental Genomics

Conservation genomics of tuturuatu, a nationally critical shorebird

Ilina Cubrinovska
University of Canterbury
ilina.cubrinovska@pg.canterbury.ac.nz
Ilina Cubrinovska[1], Tammy Steeves[1] and Dave Houston[2]

[1]School of Biological Sciences, University of Canterbury, New Zealand
[2]Department of Conservation, New Zealand

The ultimate goal of captive breeding for translocation programmes for threatened species is to prevent extinction by ensuring these species have sufficient genetic diversity to adapt to environmental change. A common challenge for these programmes is determining how best to supplement captive populations with wild individuals so that genetic diversity is maximised in captivity and in the wild. One such threatened species currently facing this challenge is the nationally critical tuturuatu (shore plover, *Thinornis novaeseelandiae*). Once widespread across New Zealand, this endemic bird is now confined to a single self-sustaining wild population on Rangatira/South East Island in the Chatham Islands, two small translocated populations on predator-free islands (Motutapu and Waikawa), and a captive breeding for translocation population held jointly at the Isaac Conservation Park and Pukaha Mount Bruce National Wildlife Centre. The extent of genetic differentiation between, and the levels of genetic diversity within and between, captive and wild tuturuatu populations are unknown. Therefore it is currently difficult to assess conservation genetic management strategies to reduce extinction risk. Here, I will present preliminary results from my PhD research based on a relatively large set of genome-wide single nucleotide polymorphisms (SNPs) that will ultimately be used to inform the conservation genetic management of tuturuatu.

Wading into conservation genomics for kakī

Ms Natalie Forsdick

University of Otago
natalie.forsdick@postgrad.otago.ac.nz
Natalie Forsdick[1], Richard Maloney[2], Tammy Steeves[3] and Michael Knapp[1]

[1]Department of Anatomy, School of Biomedical Science, University of Otago, Dunedin, New Zealand
[2]Department of Conservation, Wellington, New Zealand
[3]School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

Accurate information is essential for species conservation, allowing appropriate decision-making to not just prevent extinctions, but to assist species recovery. New Zealand's endemic kakī (*Himantopus novaezelandiae*) is considered one of the world's rarest wading birds, regarded as Nationally Critical under New Zealand's Threat Classification System. Through intensive management by the Kakī Recovery Team, involving captive breeding and rearing for release, predator control, and habitat management, the wild population has recovered to over 100 adults today. When kakī numbers have been low, they have interbred with closely-related poaka (*H. himantopus leucocephalus*) that self-introduced from Australia. This interbreeding results in fertile hybrids of intermediate plumage, but the reproductive success of kakī-poaka pairs is about half that of pure kakī pairs. Hybridisation can have negative effects on population recovery, and may compromise the genetic purity of the threatened species. Previous research indicated that hybridisation has had no significant impact on the kakī genome, but used a small number of genetic markers with low power to detect genome-wide admixture. Advanced DNA sequencing techniques facilitate high-throughput sequencing of kakī, providing greater resolution to detect admixture at a genome-wide level. Sequencing of kakī and poaka was performed on an Illumina HiSeq platform, producing > 60 Gb of sequence data per individual (approximately 50X coverage based on an estimated genome size of 1.2 Gb). We first assessed four assembly programmes and then assembled draft genomes for kakī and poaka, to be used as references allowing identification of variant sites between species. We next aim to identify regions of interspecific variation that will be used to determine any effects of hybridisation on the kakī genome. This genomic assessment will be used to establish best-practice conservation management for kakī, and to evaluate impacts of hybridisation in other threatened species.

Estimating relatedness in captive breeding for translocation programmes to enhance species recovery

Stephanie Galla

University of Canterbury

stephanie.galla@pg.canterbury.ac.nz

Stephanie Joy Galla [(1)], Marie Hale [(1)], Anna Santure [(2)], Richard Maloney [(3)] and Tammy Steeves [(1)]

[(1)] School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

[(2)] School of Biological Sciences, University of Auckland, Auckland, New Zealand

[(3)] Science and Policy Group, Department of Conservation, Christchurch, New Zealand

Threatened species recovery programmes may utilise captive breeding for translocation as a technique to prevent extinction and enhance recovery. Captive pairing decisions are generally based on available pedigree data to minimise inbreeding and maximise genetic diversity in an effort to maintain the ability to adapt to environmental change. However, pedigrees available to captive breeding for translocation programmes are often shallow (<5 generations deep), incomplete or error-prone. Furthermore, pedigree-based relatedness estimates are probability-based, and therefore, may not reflect true relatedness. While genetic-based techniques (microsatellites) offer programmes a way to estimate genetic relatedness among individuals without a pedigree, emerging evidence indicates microsatellite markers may be insufficient for accurately estimating relatedness, particularly in genetically impoverished species. More accurate measures of relatedness should be obtained from thousands of independent genome-wide single nucleotide polymorphisms (SNPs), as they provide greater genetic resolution across the genome. Here, we compare genetic and genomic-based estimates of relatedness in captive kakī (*Himantopus novaezelandiae*), a critically endangered New Zealand bird, where parent-offspring and sibling relationships are known. This research is part of a larger effort to determine the best approach for making captive pairing decisions in threatened species, which can be used to inform ~400 captive breeding for translocation programmes worldwide.

Characterisation of *Serratia proteamaculans* strain AGR96X encoding an anti-feeding prophage (tailocin) with broad activity against grass grub (*Costelytra giveni*) and manuka beetle (*Pyronota spp.*) larvae

Ms Aurelie Laugraud

AgResearch

aurelie.laugraud@agresearch.co.nz

Mark Hurst [(1)(2)], Amy Beattie[(1)], Sandra Jones[(1)], Aurelie Laugraud[(3)] and Lincoln Harper[(4)]

[(1)]Forage Science, AgResearch, Lincoln Research Centre, Private Bag 4749, Christchurch 8140, New Zealand
[(2)]Bio-Protection Research Centre, Lincoln, Christchurch, New Zealand
[(3)]Knowledge & Analytics AgResearch, Lincoln Research Centre, Private Bag 4749, Christchurch 8140, New Zealand
[(4)]Curtin University; Department of Environment and Agriculture; Centre for crop and disease management; Kent St, Bentley, WA 6102 Australia

A highly virulent *Serratia proteamaculans* strain, AGR96X, exhibiting specific pathogenicity against larvae of the New Zealand grass grub (*Costelytra giveni*), formerly *C. zealandica* and the New Zealand manuka beetle (*Pyronota festiva* and *P. setosa*), was isolated from a diseased grass grub larva. Following ingestion of a neat dose of AGR96X, death occurred within 5–12 days. Pot trials assessing AGR96X applied as a seed coat found the bacterium to be effective at controlling larvae of both grass grub and manuka beetle, giving comparable, if not greater protection than chemical insecticides. The rapid kill of AGR96X suggests a mode of action more similar to an insecticide than to the slower disease progression of the better known commercialized grass grub specific bacterium *Serratia entomophila*, which can take 3-4 months to kill the larvae. In contrast to *S. entomophila*, AGR96X rapidly multiplied within the insect host, leading to a bacterial load of approximately 2×10^8 cells per larvae within 3 days of ingestion. Genome sequencing of AGR96X revealed a plasmid encoding a variant of the *S. entomophila* anti-feeding prophage (Afp), a tailocin designated AfpX. Unlike the Afp, AfpX contains two Afp16 tail-length termination protein orthologues and two putative toxin components. A 37-kb DNA fragment encoding the AfpX-associated region was cloned, transformed into *Escherichia coli*, and fed to *C. giveni* and *Pyronota larvae*, causing mortality. In addition, deletion of the *afpX15* putative chaperone component abolished the virulence of AGR96X. We talk here about the comparative analysis of the AfpX cluster with other bacteria of different genus and also species in order to understand its mechanism and their relationship. The rapid lethality and broader host range of AGR96X make this bacterium a viable alternative to *S. entomophila* for pest control.

Genomics Aotearoa

Genomics Aotearoa is a partnership of AgResearch, Auckland University, Environmental and Scientific Research, Plant and Food Research, Landcare Research, Massey University and Otago University.

genomics.aotearoa@otago.ac.nz

Genomics Aotearoa is a new investment from the Ministry of Business Innovation and Employment that aims to improve the use of genomic sciences in New Zealand. New Zealand has unique peoples, environments and primary production systems. Genomics Aotearoa aims to develop the tools and technologies that will support the New Zealand Health, Agriculture and Environment. As much of our work will focus on the key differences between New Zealand and the rest of the world, Genomics Aotearoa will work in partnership with Maori, ensuring benefits and knowledge are generated together and shared.

4. Session: Statistical genetics

The use of family information to investigate genotyping-by-sequencing properties

Ken G. Dodds
AgResearch
ken.dodds@agresearch.co.nz
Ken G. Dodds[(1)], John C. McEwan[(1)], Rudiger Brauning[(1)], Andrew S. Hess[(1)], Shannon M. Clarke[(1)]

[(1)]AgResearch Invermay Agricultural Centre, Puddle Alley, Private Bag 50034, Mosgiel 9053, New Zealand

Genotyping-by-sequencing (GBS) produces data suitable for genomic analysis. Analyses are improved when (1) models which incorporate facets of the genotyping process (e.g. the read depths of each genotype) are taken in to account, or (2) when data which does not approximately follow the analysis model (e.g. variants in duplicated regions rather than single nucleotide polymorphisms) are excluded. We investigate models for non-random allele sampling by using excess mismatch rates (the difference between observed and expected rates) in parent-offspring sets. We also investigate whether excess mismatch rates might be useful for excluding SNPs that do not behave in a Mendelian manner.

Construction of genomic relationship matrices to account for differences between *Cervus elaphus* and *Cervus canadensis*

Dr Andrew Hess
AgResearch
andrew.hess@agresearch.co.nz
Andrew Hess[(1)], Ken Dodds[(1)], Rudiger Brauning[(1)], John McEwan[(1)] and Suzanne Rowe[(1)]
[(1)]AgResearch Limited, Invermay Agricultural Centre, Puddle Alley, Mosgiel, New Zealand

The New Zealand deer industry consists of two species of deer: *Cervus elaphus* (Red deer) and *Cervus canadensis* (Wapiti or Elk) and the crossbreds thereof. These two species display easily distinguishable morphological differences and are genetically distinct. Methods that include Red deer, Wapiti and their crossbreds in the same genetic evaluation are desirable given that crossbreeding between the species is common; however, not accounting for genomic differences between these species may result in biased genetic parameter estimates and less accurate estimated breeding values. Constructing genomic relationship matrices that use either breed-specific allele frequencies of SNPs or haplotypes have the potential to account for breed differences and improve genomic prediction accuracy. Thus, we investigate the ability of these genomic relationship matrices to account for differences between Red, Wapiti, and Red x Wapiti deer and the subsequent impact this had on breeding value estimation.

The forgotten side of the coin: phenotyping in association studies

Mr David Ashton
Plant & Food Research
david.ashton@plantandfood.co.nz
David Ashton, Peter Ritchie and Maren Wellenreuther

Seafood Research Unit, 293 Akersten St, Port Nelson, Nelson, 7010, New Zealand

The goal of genetic association studies is to identify regions of the genome that are responsible for phenotypic variation. To achieve this, detailed genetic and phenotypic data is needed. While recent developments now allow high-throughput collection of genomic data, collection of phenotype data is often more time consuming. Further development is needed for phenotyping to reach the scale of throughput and consistency now being achieved for genomic data. In this presentation, I will discuss the application of automated image-based phenotyping to genetic association research in the marine finfish species snapper (Chrysophrys auratus). Firstly, I will discuss the differences in speed and error rate between manual and automated fish length measurements. Second, I will investigate how these error rates can affect QTL detection using simulated datasets. Finally, I look at the advantages of collecting more complete phenotyping data, by contrasting the genetic association results for a single measure of fish height with results for multiple measures of the same trait.

Constructing linkage maps using low coverage sequencing data for multi-family full-sib populations

Timothy Bilton

AgResearch/University of Otago

timothy.bilton@agresearch.co.nz

Timothy P Bilton[1][2], Matthew R Schofield[2], Michael A Black[3], David Chagne[4], Phillip Wilcox[2] and Ken G Dodds[1]

[1] Invermay Agricultural Centre, AgResearch, Mosgiel, New Zealand

[2] Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand

[3] Department of Biochemistry, University of Otago, Dunedin, New Zealand

[4] Palmerston North Research Centre, New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand

Genetic linkage mapping is an important tool in the era of next generation sequencing, as it facilitates the investigation of non-model species' genes and their genomic assemblies. Frequently, linkage mapping is performed using full-sib family populations derived from the outcrossing of two individuals. However, in these populations, linkage mapping is complicated by unknown parental phase and varying segregation types. In addition, mapping using data generated from high-throughout sequencing methods that multiplex a large number of individuals (e.g., genotyping-by-sequencing) is complicated by the fact that one or more parental alleles may not be seen. Missing parental alleles results in some heterozygous genotypes being called as homozygous which, if unaccounted for, leads to inflated genetic maps. We present a new methodology, implemented in a package called GusMap, for modelling low coverage sequencing data in full-sib populations that accounts for under-called heterozygous genotypes without requiring filtering with respect to read depth. Results suggest that GusMap was able to accurately estimate the recombination fractions and overall map distance, while existing mapping packages were unable to provide reasonable estimates when the sequencing depth was low.

Bayesian genome-wide selection of apples: Validating cross-validation

Dr Peter Jaksons

Plant & Food Research

peter.jaksons@plantandfood.co.nz

Peter Jaksons [1], Linley Jesson [2], Richard Volz [2], Satish Kumar [2] and David Chagné [3]

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

[2] The New Zealand Institute for Plant & Food Research Limited, Hawke's Bay, New Zealand

[3] The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand

Genomic predictions have become an important tool for many animal and plant breeding programs to assist with the selection of superior individuals and families. Successful genomic selection can reduce the costs and time needed to select and further develop new varieties. Various statistical models can be fitted to the genetic marker and phenotypic training data to obtain predicted values for the test data. However, many of these models assume different distributions for the marker effects. This can range from a uniform marker effect (Ridge regression) to very specific marker effect sizes (BayesC). Even though the predicted values (BLUPs) based on these different markers are often only marginally different, it is important to understand their biological differences and to understand how we can evaluate and compare these models. We present the results of a genome-wide selection of apple varieties analysed in a Bayesian setting. We discuss the differences in statistical models and highlight the importance of choosing and understanding cross-validation evaluation methods for model selection.

Activity based estrus traits... New phenotypes to improve dairy cow fertility

Dr Ahmed Ismael

SCION

ahmed.ismael@scionresearch.com

Ahmed Ismael[1][3] , Erling Strandberg[2], Peter Løvendahl[1] and Guosheng Su[1]

[1]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, PO Box 50, DK-8830 Tjele, Denmark.

[2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

[3]Scion, Private Bag 3020, Rotorua 3046, New Zealand.

Improving fertility in dairy cows is becoming increasingly important, as reduced fertility has a substantial effect on overall profitability. The genetic improvement of traditional fertility traits in dairy cows is hampered by low heritabilities (< 0.1). The low heritability is caused by firstly the nature of the trait itself, and secondly, a bias induced by farmer interventions. Taken together, this results in a large unexplained residual variance that can mask the genetic effect and thereby decrease selection efficiency. Activity tags are widely used to help detect estrus, and may also provide a way to improve the accuracy of selection for fertility through more precise phenotyping, without farmer imposed bias. A study was performed on 11,363 cows to evaluate the feasibility of using activity tags to develop fertility traits with lower bias. Physical activity data were used to determine the interval from calving to the first high activity (CFHA) as a measure of the ability of cows to return to reproductive cycling after calving. Heritability of CFHA was double the heritability estimate for the interval from calving to first insemination (CFI) with 0.16 for CFHA vs 0.07 for CFI. Furthermore, the genetic correlation between CFI and CFHA was 0.96 indicating that including CFHA in the selection index, in addition to CFI, would increase the genetic gain for a rapid return to cyclicity. The accuracy of EBVs for CFHA, however, was affected by the limited numbers of the phenotypic records available. To increase the reliability of EBVs for CFHA with small numbers of phenotypic records, a multi-trait model to use additional information of the correlated traits such as CFI could be a valuable approach. Moreover, genomic selection could further improve the accuracy of fertility trait EBVs. Using a multi-trait BLUP model doubled the reliability of EBV for CFHA compared with single-trait BLUP model. Furthermore, the reliability of EBV from single-step genomic BLUP was 40% higher than the conventional BLUP model. This study indicated the importance of increased recording of activity based CFHA to improve the accuracy of EBVs for fertility traits. Activity data also allows an opportunity to develop new traits such as estrus duration and estrus strength. These traits were found to be heritable (0.03-0.06) but are not measured in current breeding programs. Such traits could possibly improve the ability of cows to display estrus behavior which improve to possibility of estrus detection.

5. Session: Maori Engagement

More than ticking a box – working with Māori and Pacific communities

Lisa Matisoo-Smith
University of Otago
lisa.matisoo-smith@otago.ac.nz

Partnering with Māori in research; whose table we sit at, whose invited to the meal and what's on the menu. Lessons learnt from the dining experience.

Irene Kereama Royal
Unitec
anewa.royal@gmail.com

Working with Maori Communities – Key Lessons from the Rakaipaaka Health and Ancestry Study

Phillip Wilcox
University of Otago
phillip.wilcox@otago.ac.nz

PL Wilcox [1,2], J Symes [1], G Symes [1], P Symes [1], N Pomana [1], J. Fitzpatrick [1], J.P. Wilson [2]

[1] Te Iwi o Ngāti Rakaipaaka
[2] Department of Mathematics and Statistics, University of Otago

In 2006 a study was initiated by ESR Ltd researchers focusing on Ngāti Rakiapaaka, a small iwi whose rohe (traditional tribal boundaries) is located in northern Hawkes Bay. This study - the Rakaipaaka Health and Ancestry (RHAS) – consisted of collecting of health, lifestyle and genetic information – including candidate-gene based SNPs - from over 400 iwi members, to understand the contributions of lifestyle and genetic factors to health status of iwi members. The study began well, and has resulted in enduring improvements in diet and physical exercise. However, to date, the genetic component of the study has been unsuccessful, largely because the iwi objectives for participating in this study – i.e., to understand genetic contributions to extant iwi health - have not been addressed. Instead, the original research team members focused on academic goals, which have included publishing a mitochondrial genome sequence and developing ancestry-informative markers, as part of two university-based post-graduate research projects. More disturbingly to the iwi however, were (a) difficulties in retrieving study data from the original research team – even though the data are legally owned by Te Iwi o Rakaipaaka Incorporated; and (b) public release of results from genetic analyses in direct defiance of an explicit instruction from the iwi to *not* release results. In addition, a complaint had been made to the Ministry of Health by a University of Otago researcher in regard to breach of the original ethics consent by original team members. However, despite these issues the iwi has restated the goal of completing the genetic analyses, and has engaged the University of Otago, who have secured HRC postdoctoral funding to do so. In this talk I will present a summary of important lessons from these experiences, and outline recommendations for those considering working with Māori communities.

6. Session: Genetics and genomics

Metabarcoding profiles: a source of quantitative traits for disease resistance?

Dr Dan Jones

Plant and Food Research
dan.jones@plantandfood.co.nz
Daniel Jones [1], Janet Yu [2], Toan Hong [2] and Joel Vanneste [2]

[1] Plant and Food Research (Mt Albert), 120 Mt Albert Road, Sandringham, Auckland 1025
[2] Plant and Food Research (Ruakura), Bisley Road, Hamilton 3214

Metabarcoding is a high-throughput method of assessing the presence and abundance of bacterial species within a population. Metabarcoding involves amplification of a phylogenetically informative region of DNA using highly conserved flanking sequences such as 16s rRNA. The relative abundance of bacterial species is inferred from the sequence of the PCR product.

We present a defined workflow for using metabarcoding data using QIIME2. Our preliminary results suggest a clear host genetic component in the abundance of bacterial species on leaves of a population of kiwifruit vines.

From this, we propose that metabarcoding can be used more widely to determine whether there is a host genetic component to abundance of bacterial species within plant populations. If this is the case, the abundance of bacterial species can be measured as a quantitative trait that has broader applications for GxE trials. As such, metabarcoding may be an information-rich method of phenotyping for disease resistance and detecting quantitative differences in bacterial abundance with the potential to identify QTLs that are likely involved in the host-pathogen response.

Genetic determinants of *Pseudomonas syringae* pv. *actinidiae* (Psa) Biovar 3 tolerance in diploid *Actinidia chinensis* (kiwifruit)

Jibran Tahir

Plant and Food Research
Jibran.tahir@plantandfood.co.nz
Jibran Tahir [1], Luis Gea [2], Stephen Hoyte [3], Cyril Brendolise [4], Heather Bassett [1], Cecilia Deng [4], Ross Crowhurst [4], Elena Hilario [4], Paul Datson [4], Ivan Liachko [5], Shawn Sullivan [5], Zac Hanley [4], Helge Dzierzon [1], David Chagné [1], and Sue Gardiner [1]

[1] The New Zealand Institute for Plant & Food Research Limited, Private Bag 11030, Manawatu Mail Centre, Palmerston North, 4442, New Zealand.
[2] Plant & Food Research, Te Puke Research Centre, 412 No. 1 Road, RD 2, Te Puke, 3182 New Zealand.
[3] Plant & Food Research Limited (PFR), Private Bag 3230, Waikato Mail Centre, Hamilton, New Zealand
[4] Plant & Food Research Limited, Private Bag 92–169, Auckland, 1025, New Zealand
[5] Phase Genomics, 4000 Mason Road, Suite 225, Seattle, WA 98195 USA

Pseudomonas syringae is a detrimental biotrophic pathogen of plant species that causes a range of diseases in its host system. Most of the globally cultivated *Actinidia chinensis* var. *chinensis* (Yellow, Gold and Red) and *Actinidia chinensis* var. *deliciosa* (Green) cultivars of kiwifruit are natural hosts of diverse *Pseudomonas syringae* pv. *actinidiae* (Psa) strains, indicating a classical host-pathogen evolutionary relationship. Multiple and geographically localised waves of Psa infections have been reported during the past thirty years and the economic losses in kiwifruit production have been enormous. The genetic and molecular mechanism of tolerance to the virulent strain of Psa, biovar 3 in kiwifruit species, however, remains largely unknown. In this study we explore the genetic regions responsible for Psa tolerance in a male diploid *A. chinensis*, a Psa biovar 3-tolerant parent (P1). Utilizing a genetic map constructed using Genotyping-by-sequencing (GBS) in a population derived from the cross of a highly susceptible gold-fleshed cultivar 'Hort16A' to P1 and detailed phenotyping of responses to natural and artificial Psa biovar 3 infection, we identified multiple quantitative trait loci (QTLs) in both P1 and 'Hort16A' that contributed to Psa tolerance and susceptibility in controlled bioassays and field conditions. To unravel the genetic architecture and candidate genes underlying these putative QTLs further, we performed whole genome assembly of the P1 male genome and employed chromosome conformation capture technology (Hi-C) to assemble ~32,000 contigs in an initial draft genome of ~632 Mbp. Using the genome assembly and candidate QTLs, we present a network of multiple loci which represent the first layer of innate immunity in diploid *A. chinensis* against Psa biovar 3.

Using genotyping-by-sequencing for high-throughput rumen microbial profiling

Dr Melanie Hess

AgResearch

melanie.hess@agresearch.co.nz

Melanie Hess[1], Suzanne Rowe[1], Tracey Van Stijn[1], Rudiger Brauning[1], Andrew Hess[1], Michelle Kirk[2], Graeme Attwood[2], Peter Janssen[2] and John McEwan[1]

[1]AgResearch Limited, Invermay Agricultural Centre, Puddle Alley, Mosgiel, New Zealand

[2]AgResearch Limited, Grasslands Research Centre, Tennent Dr, Palmerston North, New Zealand

Ruminants have a symbiotic relationship with the microbes in their rumen, and the composition of the rumen microbial community has been associated with host traits such as feed efficiency and methane production. Gaining an understanding of the rumen microbial profile of a large number of animals may help to select more efficient and environmentally-friendly animals. Traditional sequencing methods are too expensive or time-consuming to obtain rumen microbial profiles on large numbers of animals. We therefore explored the use of genotyping-by-sequencing (GBS) to generate cost-effective and high-throughput rumen microbial profiles. A pipeline was developed to BLAST reads against the Hungate 1000 genomes database and assign taxonomies to the reads. This pipeline was then tested on GBS data generated from 236 rumen samples from high- or low-methane sheep. Our results showed significant differences between samples from high- and low-methane individuals, suggesting that GBS is a promising technique for low-cost, high-throughput rumen microbial profiling.

Exploring the hypomethylome of a tired nurse (somatic embryogenesis cell line)

Natalie Graham

Scion

natalie.graham@scionresearch.com

Natalie Graham [(1)], Tancred Frickey [(2)] and Catherine Reeves [(1)]

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

[(2)] Forest Industry Informatics, Scion, 49 Sala Street, Rotorua, New Zealand

Radiata pine somatic embryogenesis (SE) is an essential component of clonal forestry and a valuable research tool. However, these SE tissue culture lines are prone to both genetic and epigenetic instability which can induce unwanted somaclonal variants. One of the mechanisms by which this occurs is through dysregulation of methylation and reactivation of transposable elements. While many conifer *in-vitro*propagation systems take precautions to minimise somaclonal mutations, other epigenetic modifications are still likely to accumulate over extended periods in culture. Methylation of cytosine residues is one of the most common and well-studied epigenetic modifications and has been shown to increase with repeated subculturing of embryogenic cell cultures. We explored methylation differences in two putatively genetically identical *Pinus radiata* cell lines with different culturing histories and divergent phenotypic characteristics. One line had been in cryopreservation with limited opportunity for accumulation of epigenetic modifications. The second line had been serially maintained for more than a decade as a nurse culture to produce somatic embryos, which now displayed a loss of maturation potential. Previous work in these cell lines has confirmed significant global methylation differences between these lines, with a higher methylation status observed in the nurse line. Higher resolution methylation studies in this species must consider alternatives to the gold standard of whole genome bisulphite sequencing due to the massive size of the radiata pine genome (~25 Gb) and only a partially assembled reference available. There are many reduced representation approaches published which enrich for the methylated portions of the genome. However, with a repetitive sequence content of approximately 80% in the radiata pine genome and a higher likelihood of methylation in these regions, we opted to focus on the unmethylated portion of the genome using methylation-sensitive restriction enzyme-based sequencing (MRE-Seq). Some significant differences in the elements represented in the hypomethylomes of these two cell lines have been shown. We report on further exploration of the differences between these hypomethylomes and their respective transcriptomes, and assess the usefulness of this approach to evaluate differential methylation in this species.

Assessing genomic selection in perennial ryegrass using GBS and a multi-population training set

Marty Faville

AgResearch
marty.faville@agresearch.co.nz
Marty Faville [(1)], Siva Ganesh [(1)], Mingshu Cao [(1)], Zulfi Jahufer [(1)], Timothy Bilton[(2)] , Andrew Griffiths [(1)], Roger Moraga[(1)], Brent Barrett [(1)]

[(1)]AgResearch Grasslands Research Centre, Private Bag 11008, Palmerston North 4442, New Zealand
[(2)]AgResearch Invermay Agricultural Centre, Puddle Alley, Private Bag 50034, Mosgiel 9053, New Zealand

Perennial ryegrass (*Lolium perenne* L.) is a key source of nutrition for ruminant livestock in NZ as well as in other temperate environments around the world. Improving the yield of herbage dry matter (DMY), seasonally and annually, is a high priority for breeding but the realised rate of genetic gain for DMY has been moderate at best. The potential for genomic selection to improve rate of genetic gain for this trait was investigated. Because the number of phenotypic records for DMY are typically limiting (e.g. <100 for a given population), the use of a multi-population training set, representative of the genetic variation within a breeding programme, was investigated. A training approach based on a half-sibling (HS) family breeding system was employed. Maternal parents of 517 HS families, sourced from five advanced breeding population, underwent genotyping-by-sequencing (GBS) to generate SNP marker information. DMY phenotypic information was acquired by assessing the HS families of each maternal parent, with each sown as replicated plots in multiple NZ environments and evaluated for two years. GBS using the enzyme ApeKI yielded 1.02M single nucleotide polymorphism (SNP) markers from the training set, after filtering. Cross-validation prediction accuracy for average DMY performance within- or across environments ranged from 0.07 to 0.43. Flowering time was also assessed, as a model trait with contrasting genetic architecture, and yielded a prediction accuracy of 0.51. For all traits Best Linear Unbiased Predictor (BLUP)-based genomic prediction methods were superior or equal to other more computationally-demanding approaches, including Ridge Regression and Random Forest. Prediction accuracy of the genomic prediction models was likely principally due to the capture of genetic relationships between training and validation sets, which may limit successful application over multiple cycles of selection due to decay of model accuracy. However, modelling indicated enhancement of genetic gain for DMY is possible within a single cycle of selection, by combining among-family selection based on phenotype with within-family selection using genomic selection (ApWgsF-HS).

Implementation of genomic selection in an advanced breeding population of *Eucalyptus nitens* using a multi-species SNP chip

Dr Jaroslav Klápště
Scion
Jaroslav.Klapste@scionresearch.com
Jaroslav JK Klápště [(1)], Mari MS Suontama [(1)], Emily ET Telfer [(1)], Natalie NG Graham [(1)] and Heidi HD Dungey [(1)]

Forest Genetics, Scion, 49 Sala Street, Rotorua 3046, New Zealand

Eucalyptus nitens is the most important commercial eucalypt species in New Zealand with an advanced open-pollination based breeding programme. *Eucalyptus nitens* has been grown predominantly for pulp wood production with a rotation age of 15 years in Southland, the southernmost region of New Zealand and the major *E. nitens* plantation area. Wood quality traits have not been the focus of breeding until recently, with increasing interest in the use of *E. nitens* for higher value solid wood products. Measurements for solid wood properties such as wood density, growth strain, wood shrinkage and internal checking at age seven were undertaken at one site. Genomic selection has been proved as a valuable tool to increase efficiency of breeding programmes and deliver genetically improved material in animal and agricultural crop breeding. Currently, the forest tree breeding community is implementing this genomic tool to accelerate the rate of genetic progress and shorten generation intervals. We applied the multi-species *Eucalyptus* SNP chip in an advanced breeding population of *Eucalyptus nitens* comprised of two samples derived from different parental populations: 1) a clonal archive incorporating a large amount of genetic diversity and a large number of parents, 2) seed orchard representing a population under higher selection intensity with reduced effective population size and a smaller number of parents. We report on the investigation of changes in linkage disequilibrium patterns due to parental population development history and its effect on prediction accuracy including transferability of SNP effects among these populations.

Poster

WRKY transcription factors (TF) and pathogenesis-related (PR) proteins involved in apple defence following challenge by *Venturia inaequalis*

Cecilia H. Deng

The New Zealand Institute for Plant & Food Research
Cecilia.Deng@plantandfood.co.nz
Cecilia H. Deng[1] , Kui Lin-Wang[1], Scott Wang[1][2], Luigi Falginella[3] , Lindy Guo[1] , Guido Cipriani[3] , Vincent Bus[4] , Andrew C. Allan[1][2] , Richard Espley[1] and Joanna Bowen[1]

[1]The New Zealand Institute for Plant & Food Research, Auckland, New Zealand
[2]The University of Auckland, Auckland, New Zealand
[3]University of Udine, Udine, Italy
[4]The New Zealand Institute for Plant & Food Research Ltd, Havelock North, New Zealand

The Ascomycete fungus *Venturia inaequalis (Vi)* can cause serious scab disease on apple, which if left untreated could lead up to a 70% reduction in apple production (MacHardy et al, 1996). WRKY transcription factors (TF) are common repressors or activators in regulating plant resistance. Pathogenesis-related (PR) proteins are associated with resistance responses in many plant-pathogen interactions. To understand the molecular mechanisms of the apple response to Vi through investigation of the roles of WRKYs and PRs, total RNA profiling experiments were designed across multiple time points (0, 24, 48 and 120 hpi - hours post inoculation) on the susceptible accession 'Royal Gala' (RG), and two resistant accessions: A248R04T010 (*Rvi5*) and GMAL3631-W193B (*Rvi8*). The latter two were descendants from host(5) and host(8), bearing resistance genes *Rvi5* and *Rvi8*, respectively. The resistance reaction conditioned by *Rvi5* typifies a hypersensitive response (HR), while *Rvi8* characterises a delayed stellate necrosis (SN) response. Clustering transcriptome assemblies from RG, h(5) and h(8) to the newly published 'Golden Delicious' double haploid (GDDH13 v1.1) genome (Daccord et al., 2017), we constructed a *Malus* pan-genome of 61,962 gene models, which served as the reference for analysis of differentially expressed genes (DEGs). More DEGs were detected in h(5) at the early infection stage compared with h(8) and RG, which indicated the response to Vi was stronger and plant defence immunity was triggered earlier in *Rvi5*. This agrees with their phenotypic resistance responses, with h(5) showing HR within two days, while it takes five days for h(8) to show the SN reaction. The interactions between differentially expressed PRs and WRKYs were verified with transient dual luciferase assays on *Nicotiana benthamiana*. Our study showed that in response to fungal pathogen challenge by *V. inaequalis*, some WRKYs may act upstream of PR proteins in the molecular defence mechanisms in the apple host.

One of these things is not like others: a simple robust PCR tool for identifying clover interspecific hybrids from the field

Ms Anna Larking
AgResearch
anna.larking@agresearch.co.nz
Anna Larking[1], Roger Moraga[1] and Andrew Griffiths[1]

[1]AgResearch, Grasslands Research Centre, Private Bag 11008, Palmerston North 4442, New Zealand

White clover (*Trifolium repens* L) is the main legume in New Zealand's pastoral agriculture systems. It provides quality forage and, through its symbiosis with Rhizobia, converts atmospheric nitrogen into plant available forms. As a crop, white clover provides some challenges as it is susceptible to biotic stressors such as nematode infection, and abiotic stressors such as drought, and poor performance in phosphorus-deficient soil. While there is untapped genetic potential within white clover that may address these issues, another avenue is to widen the gene pool to include closely related clover species with known resistance to various stressors. Developing inter-specific hybridisation methods provides a fast-track to introduce new traits to white clover. During the process of incorporating new traits into white clover, the initial interspecific hybrid is back-crossed repeatedly to white clover to regain the required agronomic performance while retaining the trait of interest. With each back-cross, the resulting progeny gain more white clover physical characteristics. This can be an issue in field trials where interspecific plant material is being assessed, as it can be difficult to differentiate between the hybrids and white clover 'volunteer' plants that may have infiltrated the trial from dormant seed in field. Our aim was to develop a simple PCR-based tool for breeders to confirm whether a clover plant was an interspecific hybrid, or not. Focusing on *T. uniflorum*/*Trifolium repens* hybrids, we screened reference assemblies of each contributing species for genome-wide repeats, designed species-specific primers within these repeats, and optimized PCR conditions for greatest amplification discrimination. We have tested the set of species-primers across a wide range of *T. uniflorum* and *Trifolium repens*accessions as well as interspecific hybrid germplasm of different back-cross levels with known pedigree, as well as samples from field trials. The PCR primers successfully discriminated between white clover and interspecific hybrid germplasm, and are, therefore, a simple robust tool for rapid identification of *T. uniflorum*/*Trifolium repens* hybrids.

Breeding kiwifruit for resistance to latania scale (Hemiberlesia lataniae).

Casey Flay

Plant and Food Research
Casey.Flay@plantandfood.co.nz
Casey Flay [1], Paul Datson [2],

[1] Plant & Food Research, Te Puke Research Centre, 412 No. 1 Road, RD 2, Te Puke, 3182 New Zealand.
[2] Plant & Food Research Limited, Private Bag 92–169, Auckland, 1025, New Zealand

Kiwifruit is a billion dollar industry which is highly susceptible to various pests and diseases. Breeding plants resistant to these biotic threats has been shown to be the most cost effective and environmentally friendly route for control of biotic threats. This year I started a PhD aimed at improving the speed and cost efficiency of breeding kiwifruit resistant to biotic threats using molecular breeding tools. This research initially focuses on resistance of diploid (*A. chinensis*) Kiwifruit to latania scale insect *(Hemiberlesia lataniae)*. It involves aspects of phenotyping and bioassays to identify resistance phenotypes. Genotyping by sequencing and genetic mapping to identify regions of the genome that contain resistance alleles. And fine mapping to narrow the gene region associated with an identified marker.

An inexpensive, versatile, high-throughput plant genomic DNA extraction method suitable for genotyping-by-sequencing.

Mr Craig Anderson

AgResearch

craig.anderson@agresearch.co.nz

Craig B. Anderson[1], Andrew G. Griffiths[1] , Benjamin K. Franzmayr[1][2] , Soon Won Hong[1] , Anna C. Larking[1], Rachel Tan[1] , Roger Moraga[1] , Marty J. Faville[1] and Tracey C. van Stijn[3]

[1]AgResearch Grasslands Research Centre, Private Bag 11008, Palmerston North 4442, New Zealand
[2]Slipstream Automation, Bachelor Centre, Module 4, Tennent Drive, Palmerston North 4442, New Zealand
[3]AgResearch Invermay Agricultural Centre, Puddle Alley, Private Bag 50034, Mosgiel 9053, New Zealand

High-throughput, inexpensive methods for extracting high-quality plant genomic DNA (gDNA) underpin deployment of next-generation sequencing marker technologies such as genotyping-by-sequencing (GBS). Several high-throughput DNA extraction methods are available, but typically provide low yield or quality gDNA, or are costly (US$6-$9/sample). We modified a non-organic solvent protocol to extract microgram quantities (1–13 µg) of sequencing-quality gDNA inexpensively in 96-well plates from fresh, freeze-dried or silica gel-dried plant tissue. The method was effective for several difficult-to-extract forage, crop and horticultural species. Tissue ground in a 96-well plate format is mixed with a homogenisation buffer, followed by precipitation to remove impurities by centrifugation. The supernatant is applied to a 96-well silica filter plate, washed, then the gDNA eluted. This scalable, readily-automated protocol requires approximately three hours to process 192 samples (384-570 samples/day) at US$0.62/sample for consumables. This method produces sequencing-quality gDNA for GBS, enabling downstream applications such as genomic prediction and genome-wide association studies.