

MapNet 2019

Victoria University of Wellington, 18–19 November

mapnet2019.nz

mapnet2019@vuw.ac.nz

**VICTORIA
UNIVERSITY
WELLINGTON**

TE WHARE WĀNANGA
O TE ŪPOKO O TE IKA A MĀUI



Sponsors

Gold sponsor



dnature

At dnature we recognise that we're different to other molecular diagnostic suppliers in that we also have our own laboratory for DNA diagnostic work. This means that the instruments that we distribute need to be fit for purpose (yours and ours) – are easy to use, handle low volumes, and are suitable for small and large project work.

We are proud to distribute a range of liquid handling instruments that are great for improving laboratory workflow and efficiency. These instruments range from our Viaflo Electronic pipettes through to the Myra and Assist Plus.

Our Viaflo pipettes speed up lab work but are also perfect for avoiding OOS. The Myra, (which can work in conjunction with the Mic qPCR machine) and the Assist Plus, are options for rapid plate filling with reagents and the reformatting of 8-16 samples.

For more information, or to book a demo contact us at info@dnature.co.nz or 0800DNATURE.

Sponsors

Silver sponsor



Custom Science is a specialist supplier of services, reagents and tools for Next-Gen Sequencing. We are New Zealand's foremost supplier of NGS services, including Illumina, PacBio, Nanopore PromethION platforms. We also specialize in POC and field-based extraction and testing systems, including for MinION and qPCR.

Invited speakers sponsor



ESR

ESR (the Institute of Environmental Science and Research) is New Zealand's Crown Research Institute that specialises in science relating to people and communities. It's our science that helps safeguard people's health, protect food-based economies, improve the safety of freshwater and groundwater resources, and contributes expert forensic science to justice systems.

MapNet Programme 2019

Mane / Monday - 18 November

8:30 – 9:30	Whakaurunga / Registration	Te Toki a Rata (TTR) Foyer
9:30 - 10:00	Paramanawa / Morning tea	TTR Foyer
10:00 - 10:15	Mihi Whakatau / Welcome	TTR L1

Wāhanga tuatahi / Session 1: Primary sector genomics (Chair: Jeanne Jacobs)

10:15 – 10:30	Sara Montanari	Tools for molecular breeding in pear
10:30 – 10:45	Rachael Ashby	From discovery to application: implementing Genotyping-by-Sequencing for genomic breeding of New Zealand Greenshell™ Mussel
10:45 – 11:00	Noemie Valenza-Troubat	Genomics of New Zealand Trevally: Exploring the Genetic Basis of Quantitative Traits to Inform the Breeding of a New Species for Aquaculture
11:15 – 11:30	Natalie Graham	Opening Pandora's Box: The Pedigree of New Zealand's Radiata Pine
11:30 – 11:45	Seoljong Kim	A population-genomic and taxonomic study of <i>Eucalyptus argophloia</i> and <i>E. bosistoana</i>
11:45 – 12:00	Hymmi Kong	Integrated information system for Kiwifruit molecular breeding

12:00 – 13:15	Tina / Lunch	TTR Foyer
---------------	---------------------	-----------

Wāhanga tuarua / Session 2: Ecological and evolutionary genomics (Chair: Peter Ritchie)

13:15 – 13:45	Libby Liggins	Ira Moana – Genes of the Sea – Network and Database
13:45 – 14:00	Yvan Papa	Draft genome assembly and whole genome-level analysis of New Zealand tarakihi stock structure
14:00 – 14:15	Annabel Whibley	Assembling the common myna genome: establishing a foundation for invasion genomics
14:15 – 14:30	Gancho Slavov	1000 genomes projects – what is their value for tree improvement and conservation?
14:30 – 14:45	Ann McCartney	Utilising Oxford Nanopore Data for the Genome Assembly of Endemic New Zealand Species
14:45 – 15:00	Dafni Anastasiadi	Human self-domestication: can animal models help to explore the role of epigenetics in vertebrate domestication?

15:00 – 15:30	Paramanawa / Afternoon tea	TTR Foyer
---------------	-----------------------------------	-----------

Wāhanga tuatoru / Session 3: Methodological Advances (Chair: David Ashton)

15:30 – 15:45	Ken Dodds	Identity matching using low-depth sequencing data
15:45 – 16:00	Tim Millar	Local haplotype assembly in Autopolyploid Actinidia
16:00 – 16:15	Yilei Zhang	Identification of key drivers from complex network modelling
16:15 – 16:30	Eliatan Niktab	Network-based Nonparametric Tests to Identify Genetic Modifiers of Niemann-Pick type C disease
16:30 – 16:45		
16:15 – 17:00		

18:30	Hapa / Dinner	<i>The Backbencher, 34 Molesworth Street</i>
-------	----------------------	--

Useful information

Tūrei / Tuesday - 19 November

Wāhanga tuawhā / Session 4: Invited speaker (Chair: Miles Benton)		
09:00 – 10:00	Joep de Ligt (ESR)	WGS to study patterns of genetic variation in human health and disease
10:00 – 10:30	Paramanawa / Morning tea	TTR Foyer
Wāhanga tuarima / Session 5: Human and health genomics (Chair: Donia Macartney-Coxson)		
10:30 – 10:45	Wenting Liu	Exploring Combinatorial Genetic Contributions to a Serious Adverse Drug Reaction
10:45 – 11:00	Cintya Del Rio Hernandez	Genetic interaction networks within genomes mediate individual drug response
Wāhanga tuaono / Session 6: Microbial genomics (Chair: Una Ren)		
11:00 – 11:30	Tim Stinear	Genomics to understand the emergence of Buruli ulcer (<i>Mycobacterium ulcerans</i> infection) around Melbourne
11:30 – 11:45	Melanie Hess	Metagenome profiling of thousands of New Zealand sheep rumen samples
11:45 – 12:15	Xochitl Morgan	Persistence of a poultry-associated lineage of Vancomycin-resistant Enterococci
12:15 – 13:15	Tina / Lunch	TTR Foyer
Wāhanga tuawhiti / Session 7: Māori kaupapa/Te Ao Māori in genetics teaching and research (Chair: Phil Wilcox)		
13:15 – 13:45	Phil Wilcox	Hei Whakamohio ai i nga Tauira Māori me Pākehā o te Mātauranga o te Ira: Raising understanding of Te Ao Māori in genetics education in Aotearoa/New Zealand
13:45 – 14:00	Levi Collier-Robinson	
14:00 – 14:15	Tia Haira	Ngā hua o te kānuka: the fruits of kānuka
14:15 – 14:30		
14:30 – 15:00	Lisa Warbrick	
15:00 – 15:30	Whakawhitiwhiti korero / Final discussion and farewell	

Website: mapnet2019.nz

Contact: mapnet2019@vuw.ac.nz

Useful information

Conference dates

18-19 November, 2019

Start time 9:30 am Monday 18 Nov

Finish time 3:30 pm Tuesday 19 Nov

Location

MapNet 2019 will be held in the Te Toki a Rata building lecture theatre (TTR L1) on the Kelburn Campus at Victoria University of Wellington.

Click [here](#) to view Google Maps location or download the campus map at the following address:

<https://www.victoria.ac.nz/about/explore-victoria/campuses/kelburn/kelburn-campus-map.pdf>

Social Media

If you are using twitter, tag your conference posts #mapn19. Please respect the requests of speakers and conference attendees that ask or suggest not to be included in social media posts.

Travel

Wellington airport (WLG) is about 9 km to Victoria University. Taxis from the airport cost \$40+ and take half an hour or so. Ask to be dropped off at Gate 7, Kelburn Parade. There are also shared shuttle services which can drop you in the centre of town or at your accommodation for about \$18–25, and take around an hour. The airport bus will take you to the center city, where you can transfer to a bus to VUW.

Accommodation

The best low cost and convenient accommodation option for MapNet 2019 is Te Puni Village. Click on “Book now”. Use the code MapNet2019 in the promo field after selecting the dates for your stay.

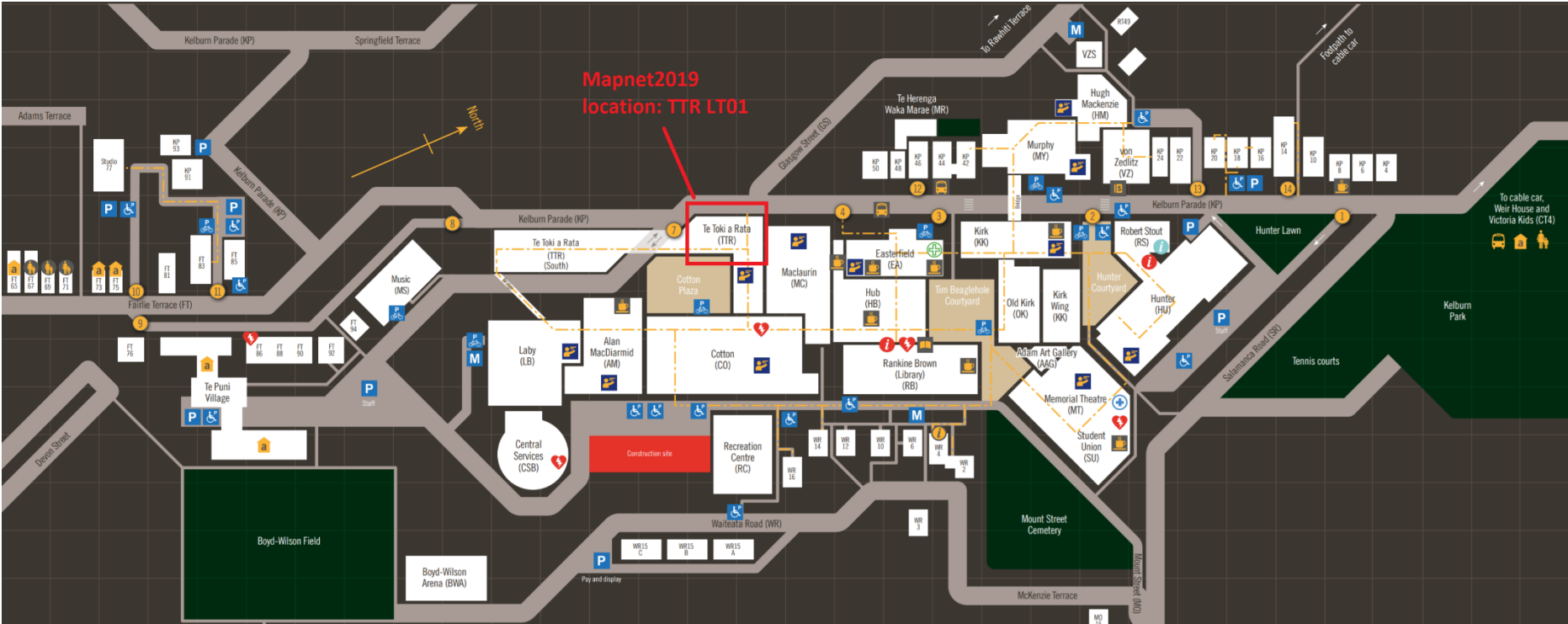
Conference organising committee:

Peter Ritchie, Mel Dohner, Yvan Papa, and Tom Oosting

Programme organising committee:

Peter Ritchie, Donia Macartney-Coxson, Una Ren, Miles Benton, and Phil Wilcox

Useful information



1: Primary sector genomics

Abstracts

Session 1: Primary sector genomics

Tools for molecular breeding in pear

Sara Montanari¹

¹Plant and Food Research, Motueka

The development of genetic tools for high-density and large-scale genotyping have enabled the dissection of important traits and the evaluation of genetic diversity in several plant species. With the final objective of enhancing the implementation of marker assisted breeding in pear, we designed a highly efficient Affymetrix Axiom Pear 70K Genotyping Array and we used it to screen the entire *Pyrus* collection held at USDA National Clonal Germplasm Repository (NCGR) in Corvallis, OR. This large collection includes more than 2,000 clonal pear accessions, encompassing nearly every known *Pyrus* species and providing a valuable source of diversity to be exploited in pear breeding programs. We have then used this large dataset to reconstruct the pedigree of several of these accessions and to evaluate the genetic diversity of the germplasm collection, which will provide useful information for pear breeding.

1: Primary sector genomics

From discovery to application: implementing Genotyping-by-Sequencing for genomic breeding of New Zealand Greenshell™ Mussel

Rachael Louise Ashby¹, Andrew Hess¹, Hayley Baird¹, Rudiger Brauning¹, Ken Dodds¹, Nick King², Rodney Roberts³, Neil Gemmell⁴ and Shannon Clarke¹

¹AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

²Cawthron Institute, Nelson, New Zealand

³SpatNZ, Nelson, New Zealand

⁴Department of Anatomy, University of Otago, New Zealand

The New Zealand Greenshell™ Mussel (GSM) is a taonga species of economic importance to the New Zealand Aquaculture industry as the biggest export by both value and volume. Despite this, the underlying genomics of GSM and the Mollusca phylum are inherently understudied. They have highly complex genomes containing large levels of repeats, single nucleotide polymorphisms (SNPs) and gain SNPs as they age. In addition, current breeding methods utilise recorded pedigree and phenotype information to generate breeding values, which can be error prone and labour intensive. Over the last seven years, AgResearch has been working in collaboration with SpatNZ and Cawthron to develop a high-throughput cost-effective method of implementing genomics for uses including parentage assignment and genomic selection. However, the interesting characteristics of the species' genomics has meant current Genotyping-by-Sequencing (GBS) pipelines already used for farmed species are unable to work efficiently with GSM. We have overcome these limitations and have developed, tested, benchmarked and implemented a high-throughput GBS pipeline and have now genotyped over 7,500 mussels. This pipeline has been utilised for downstream analyses of interest to the aquaculture industry, including parentage assignment/verification, genomic selection for breeding and genome wide association studies (GWAS). We will discuss the challenges faced during the development of our pipeline, including considerations for scalability from hundreds to thousands of samples, and present how the tools developed have been used to advance the New Zealand GSM breeding program.

1: Primary sector genomics

Genomics of New Zealand Trevally: Exploring the Genetic Basis of Quantitative Traits to Inform the Breeding of a New Species for Aquaculture

Valenza-Troubat, N.^{1,3}, Montanari, S.², Morrison-Whittle, P.¹, Ashton, D.¹, Ritchie, P.A.³, Wellenreuther, M.^{1,4}

¹The New Zealand Institute for Plant and Food Research, Nelson, NZ,

²The New Zealand Institute for Plant and Food Research, Nelson, NZ,

³School of Biological Sciences, Victoria University of Wellington, Wellington, NZ,

⁴Faculty of Science, University of Auckland, Auckland, NZ.

Understanding the complex network of genes underlying phenotypic variation and their external modulation has been a major and longstanding challenge in genetics. In animal breeding, locating and characterising quantitative trait loci (QTL) is of great importance to localise genetic variation involved in commercially relevant traits. High density linkage maps are useful tools for fine-scale mapping of quantitative trait loci as they allow the relative positioning of different marker loci along the genome. Our long term research programme on the New Zealand trevally (*Pseudocaranx georgianus*) aims to inform the selective breeding of this species for aquaculture. As part of this, we have developed a suite of genomic resources to support breeding decisions, including a well-assembled reference genome, and a high density single nucleotide polymorphism (SNP) dataset of the broodstock and F1 cohort based on whole-genome resequencing data from the parents and Genotyping by Sequencing data from the offspring. Here we present the initial efforts to construct the first high density linkage map for the New Zealand trevally, and to align it with the reference genome assembly. Over 60,000 SNPs were mapped and ordered on 24 trevally linkage groups, which is akin to the number of expected chromosomes, using a pedigreed population comprising 231 fish from 4 full-sib families. We discuss the insights that can be gained from this linkage map and how it will be used in the future to map QTLs in this breeding population.

1: Primary sector genomics

Opening Pandora's Box: The Pedigree of New Zealand's Radiata Pine

Natalie Graham¹, Jaroslav Klapste¹, Emily Telfer¹, Ahmed Ismael¹, Gancho Slavov¹, David Pont², Mark Paget³, Heidi Dungey¹

¹Forest Genetics, Scion, 49 Sala Street, Rotorua, New Zealand

²Forest Industry Informatics, Scion, 49 Sala Street, Rotorua, New Zealand

³Radiata Pine Breeding Company, 99 Sala Street, Rotorua, New Zealand

Accurate pedigrees are essential when estimating breeding values, directly impacting selection accuracy and the delivery of genetic gain. Errors, however, can happen at any stage in the breeding cycle but without the tools to detect and address them, errors gradually accumulate in documented pedigrees. With the newly developed radiata pine 36,285 SNP array, NZPRAD02, we now have access to robust and affordable genotyping, with approximately 12,000 trees within the New Zealand breeding programme already genotyped. Pedigree reconstruction in these population will, therefore, be the first application of this SNP array for the Radiata Pine Breeding Company. It will also enable the recovery of hidden relatedness, which has been shown to improve predictions accuracies in another forestry species, *Eucalyptus nitens*. Recreating pedigrees has also been a goal of Scion's Growing Confidence in Forestry Futures programme - remote sensing approaches are identifying elite performers in commercial forests, but pedigrees are not usually available for these individuals. The ability to resolve the genetics of these individuals allows site-specific performance of these breeds to feed back into the breeding programme, further supporting the move towards precision forestry.

1: Primary sector genomics

A population-genomic and taxonomic study of *Eucalyptus argophloia* and *E. bosistoana*

Seol-Jong Kim^{1,2}, Clemens Altaner², Luis Apiolaza², Tammy Steeves¹ and Pieter B Pelser¹

¹School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

²School of Forestry, University of Canterbury, Christchurch, New Zealand

The New Zealand Dryland Forests Initiative (NZDFI) aims to create plantations of high-value *Eucalyptus* timber species in dry environments on the east coast of New Zealand. This would enable the sustainable production of naturally-durable hardwood in New Zealand as a substitute for CCA-treated pine and unsustainably harvested tropical hardwoods. For this purpose, Australian seed collections of five promising *Eucalyptus* species have been used since 2009 to establish progeny trials in New Zealand. These trials are used to select and breed plant lines with growth and wood properties that are desirable for the New Zealand environment. As part of this effort, NZDFI is interested in understanding how genomic and environmental variation interact to influence commercially important traits in the NZDFI progeny trials. My PhD research project is a component of this project. Its specific research questions are: 1) what is the taxonomic identity of a morphologically deviating population of *E. bosistoana*, 2) what are the patterns of genetic diversity and structure of *E. argophloia* and *E. bosistoana*, and 3) what is the mating system of *E. bosistoana*. I aim to address these questions using morphological, DNA sequence and Single Nucleotide Polymorphism (SNP) data. To be able to compile the latter data set, we joined the 'Eucalyptus 65kSNP Axiom array production and deployment initiative'.

1: Primary sector genomics

Integrated information system for Kiwifruit molecular breeding

Hymmi Kong¹, Chetan Baadkar¹, Gaurav Chandani¹, Brett Davis¹, Lee O'Grady¹, Elizabeth Jones², Barry Peralta¹, John McCallum¹, Tim Millar¹, Amardeep Nath¹, Yaw Nti-Addae², Jean Sabado¹, Susan Thomson¹, Deb Weigand² and Guy Davenport¹

¹Plant & Food Research, 412 No 1 Road, RD2, Te Puke, 3182, New Zealand

²Genomic Open-source Breeding Informatics Initiative, Cornell University, Institute of Biotechnology, 606 Frank H. T. Rhodes Hall, Ithaca, NY 14853

Plant breeders require a combination of pedigree, phenotypic and genotypic information to make decisions on parental and progeny selection for new crosses. The challenge is ensuring these data are fully integrated to allow these decisions to be made in an accurate and timely manner. In the Kiwifruit programme at Plant and Food Research (PFR), data are collected across multidisciplinary teams. The size and shape of these data have led to different database solutions for their capture and reporting. Observations such as flowering time, fruit colour and soluble solid content are recorded in the orchard and laboratories directly into our information system. Tissue samples for genotyping are registered in separate databases and results are imported back into the information system to align with the pedigree and phenotypic information.

PFR have started a collaboration with the Genomic Open-source Breeding informatics initiative (GOBii; <http://gobiiproject.org>) to extend their Genomics Data Manager (GDM) for use in Kiwifruit breeding and integrate it with the PFR in-house breeding information system. As part of this collaboration support for polyploids will be added to GDM and a new web-based data loader will be developed. The software from this collaboration will be made freely available under an open source license. We will make use of the Breeding API (BrAPI) (<https://brapi.org/>) to provide the integration between the two systems and to provide a platform for connection to other BrAPI apps, such as Pedigree Viewer. Full integration of these systems will ensure good data quality, timely analysis and reduce operational costs.

2: Ecological and evolutionary genomics

Session 2: Ecological and evolutionary genomics

Ira Moana – Genes of the Sea – Network and Database

Libby Liggins^{1,2} and Cory Noble¹

¹School of Natural and Computational Sciences, Massey University, Auckland; and ²Natural Sciences, Auckland Museum, Tāmaki Paenga Hira

Ira Moana Network (<https://docs.google.com/spreadsheets/d/1dn--SHEbKQVcCnYrng76Fj7B8d4Gflsww9NtQE7wl54/edit?usp=sharing>)

The Ira Moana Project (www.massey.ac.nz/iramoana) is a Catalyst Seeding Funded initiative to build a collaborative research network of New Zealand molecular ecologists, and to develop a metadata database for the genetic and genomic data of our natural populations. The Ira Moana Network has worked to enable New Zealand's molecular ecology community to practice the principles of F.A.I.R. (Findable, Accessible, Interoperable, and Reusable) and C.A.R.E. (Collective benefit, Authority to control, Responsibility, and Ethics); leveraging the value of both open and collaborative science, but also culturally aware science. Over 85 researchers, from over 25 institutions have participated as part of the network via several workshops and datathons. The network has established a metadata database infrastructure suited to the needs of New Zealand researchers and communities and meets the latest international standards for biodiversity and genomic data. The metadata database aims to ensure the stewardship of our data, linking sequences with sample information – such as location, habitat, and indigenous community (māori) context – creating opportunities for data synthesis, to inform our future research, and to operationalise appropriate data re-use. To date, metadata for over 2,000 genetic resources have been uploaded to the Ira Moana Project Database hosted by the Genomics Observatory Metadata Database (GEOME, www.geome-db.org). I will present an overview of the infrastructure and data resource that has been created by the Ira Moana Project, I will share a consensus statement regarding community best-practice developed by early career researchers who have participated in the network, and extend an invitation to everyone to be involved and participate.

2: Ecological and evolutionary genomics

Draft genome assembly and whole genome-level analysis of New Zealand tarakihi stock structure

Yvan Papa¹, Maren Wellenreuther^{2,3}, Mark A. Morrison⁴, Peter A. Ritchie¹

¹ School of Biological Sciences, Victoria University of Wellington, New Zealand

² The New Zealand Institute for Plant and Food Research, Nelson, New Zealand

³ Faculty of Science, University of Auckland, New Zealand

⁴ National Institute of Water and Atmospheric Research, Auckland, New Zealand

Tarakihi (*Nemadactylus macropterus*) is a widely distributed fish around the inshore areas of New Zealand and South of Australia. It supports an important commercial fishery with annual landings in New Zealand averaging over 5,000 t for the past 40 years. Very little is known about its stock structure and previous low-resolution genetic studies have sampled around Australia but only analysed one sample site from New Zealand. The aim of this ongoing research project is to use whole-genome sequencing to better understand the population genetics of this species. We used 200x coverage Illumina short-read and nanopore Promethion long-read whole genome sequencing of one harvested specimen to assemble a *de novo* draft genome of *N. macropterus*. This resulted in an assembly of 2,696 scaffolds with a genome size of 609 MB (N50=69, L50=1.87 MB, maximum scaffold length=19 MB). Detection and annotation of repetitive elements and coding regions of this genome are currently underway. We also aim to obtain 10x coverage whole genome Illumina data of 200 wild tarakihi specimens from around New Zealand and South Australia. Of these, the 47 specimens from New Zealand sequenced so far have yielded a dataset of 4,651 SNPs filtered for quality, allele frequency and neutrality. This will be one of the first whole genome-level studies of a New Zealand fishery species and will enable precise population genetic testing for differentiation and the discovery of adaptive variation. The results of this study will be compared to parallel studies that use mitochondrial DNA and otolith microchemistry.

2: Ecological and evolutionary genomics

Assembling the common myna genome: establishing a foundation for invasion genomics

Annabel Whibley¹, Richard Major², Rebecca Johnson², Anna Santure¹

¹School of Biological Sciences, University of Auckland, Auckland, New Zealand

²Australian Museum, Sydney, Australia

Common myna (*Acridotheres tristis*) is one of only three birds to feature on the IUCN Species Survival Commission blacklist of the 100 worst invasive alien species. Introduced to New Zealand in the 1870s to control vineyard, grain and market garden pests, common myna now inhabit much of the North Island, where they thrive in urbanised and disturbed environments. A similar and widespread history of introduction (or escape) and invasion has occurred from its native range in Central and Southern Asia across much of the globe. Where considered a pest, this is generally due to the threat that the common myna poses, or is perceived to pose, to native birds and other wildlife or to agricultural production. I will present data on our experiences of generating a draft assembly of the 1.05Gb genome. Our mixed sequence datasets combine Illumina short reads, Chromium 10x sequencing, PacBio and Oxford Nanopore long reads and Hi-C datasets and emphasis will be given to the relative performance of different assembly strategies. This draft genome will underpin our efforts to understand the demographics and functional biology of the common myna invasion at a range of spatial scales. One focus will be how bottlenecked populations seem able to adapt despite their restricted genetic diversity. We are particularly keen to assess whether novel genetic diversity generated by transposable element (TE) movement might offer one solution to the so-called genetic invasive species paradox and will present preliminary data on the TE landscape of the common myna genome.

2: Ecological and evolutionary genomics

1000 genomes projects – what is their value for tree improvement and conservation?

Gancho T. Slavov^{1,2}, Luke M. Evans^{3,4}, David Macaya-Sanz³, Stephen P. DiFazio³, Glenn T. Howe⁵

¹Forest Genetics, Scion, Rotorua, New Zealand

²Computational and Analytical Sciences, Rothamsted Research, Harpenden, UK

³Department of Biology, West Virginia University, Morgantown, WV, USA

⁴Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA

⁵Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR, USA

Accurate prediction of individual- and population-level phenotypic traits from genomic data can help accelerate tree improvement and mitigate the effects of climate change. We used phenotypes (bud flush, bud set and height growth) and whole-genome re-sequencing data for 840 *Populus trichocarpa* trees sampled from natural stands along 16 rivers in the Pacific Northwest of North America to compare the predictive abilities (PAs, the correlations between predicted and observed phenotypic values) of models based on single-nucleotide polymorphisms (SNPs) to those of traditional genealogical models. The overall ability to predict individual-tree phenotypes (i.e., incorporating variation across stands and rivers) was moderate to high ($PA > 0.5$) using geography, climate, or SNPs (i.e., GBLUP) as predictors. On average, GBLUP PAs (0.735) were only slightly higher than those based on climate (0.683) or geography (0.659), and the relative advantage of GBLUP was largest (i.e., ca. 10%) for bud flush. However, partitioning of PAs into hierarchical levels revealed that (1) whereas GBLUP models were best for predicting the phenotypes of clonal genotypes within stands, none of the models performed well ($PA < 0.2$), (2) models based on SNPs were consistently best at predicting stand-level phenotypes; and (3) all models were able to predict river-level phenotypes very well for each trait (mean $PA = 0.946$, range = 0.804–0.989). We illustrate the implications of these results by delineating “seed zones” based on genomic versus phenotypic, climatic and geographic data and then discuss the relative utility of genome re-sequencing projects performed at this scale (i.e., $N \sim 1,000$).

2: Ecological and evolutionary genomics

Utilising Oxford Nanopore Data for the Genome Assembly of Endemic New Zealand Species

Ann Mc Cartney¹, Chen Wu³, Ross Crowhurst³, Joseph Guhlin², Chris Smith¹, David Chagne⁵, Thomas Buckley¹

¹Systematics Team, Manaaki Whenua, 231, Morrin Road, Saint Johns Auckland.

²Department of Biochemistry, School of Biomedical Sciences, University of Otago, Cumberland Street, Dunedin.

³Plant and Food Research, 120 Mt Albert Road, Sandringham, Auckland.

⁴Commerce A, Room 113, University of Auckland, Symonds Street, Auckland.

⁵Plant and Food Research, Batchelar Road, Fitzherbert, Palmerston North.

As part of Genomics Aotearoa, a high-quality genomes project has been established to generate pipelines for the assembly of species across New Zealand. These pipelines are specifically targeted at key species that are on the verge of extinction, treasured by Māori, key players in the primary production industry, a significant threat to biosecurity within New Zealand or are considered genomically complex i.e. are abnormally large, have higher ploidy levels, are highly repetitive or heterozygous. These species have been sequenced using a variety of NGS platforms, namely Illumina, Oxford Nanopore (ONT), PacBio, Chromium10X and Hi-C. The utilisation and hybridization of data from these technologies has already resulted in the generation of high quality genomes for invasive wasp species, Polynesian rat, manuka, blueberry and 125+ kakapo genomes to name but a few. Here, genome assembly strategies will be outlined specifically using ONT and Illumina datatypes in order to highlight the impact of read depth and coverage on genome assembly quality. This study deals with the optimisation of genome assembly construction for projects with a limited budget or those that are confined to certain locations/sequencing platforms. It also addresses optimal assembly strategies for species with unique genome architectures. In order to investigate this five species with unique genome characteristics were selected, namely; *Hericium novae-zealandiae* a small diploid fungi, *Clitarchus hookeri* a species containing a large, repetitive and highly heterozygous genome, *Knightia excelsa* a plant species with a medium size, non-repetitive genome with low heterozygosity, and kiwifruit another plant species with a smaller and more repetitive genome structure.

2: Ecological and evolutionary genomics

Human self-domestication: can animal models help to explore the role of epigenetics in vertebrate domestication?

Dafni Anastasiadi¹, Francesc Piferrer², Maren Wellenreuther^{1,3}, Antonio Benítez-Burraco⁴

¹The New Zealand Institute for Plant and Food Research, Nelson, New Zealand ²Institute of Marine Sciences, Spanish National Research Council, Barcelona, Spain ³School of Biological Sciences, The University of Auckland, Auckland, New Zealand ⁴Department of Spanish, Linguistics, and Theory of Literature, Faculty of Philology, University of Seville, Seville, Spain

Human evolution is characterized by the acquisition of features found in domesticated animals referred to as the self-domestication hypothesis. Features of self-domestication appear altered in people with cognitive diseases, including autism spectrum disorders, schizophrenia and Williams syndrome. The genetic basis of self-domestication has been revealed by non-random associations of animal domestication candidate genes with genes positively selected in recent humans and involved in cognitive disorders. Epigenetic modifications may have also participated in human self-domestication. However, their role is difficult to disentangle unless using animal models because epigenetics have a partly genetic basis and epigenetic patterns are tissue-specific, while only fossil bones are available. Here, we used our published dataset on the early domestication effects on DNA methylation of European sea bass (*Dicentrarchus labrax*). We mapped 2114 genes related to mammalian domestication to their orthologues in sea bass. We overlapped these with genes differentially methylated between wild and domesticated fish (n = 3, 4 tissues). A set of ~90 orthologues exhibited changes at the early domestication stages in fish. Moreover, we found overlaps with genes selected in anatomically-modern humans and with candidates for human-specific cognitive disorders. Thus, alterations of their methylation status may account in part for some of the human self-domestication phenotypes, and because of the links between self-domestication and cognitive disorders, also for aspects of the etiology of these disorders. Our results suggest that epigenetic modifications can be a general feature involved in vertebrate domestication, and highlight the need to catalogue methylation differences to gain insights into human evolution.

Session 3: Methodological Advances

Identity matching using low-depth sequencing data

Ken G. Dodds, Rudiger Brauning, John C. McEwan and Shannon M. Clarke

AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand

Some situations require the comparison of one or more genotyping results to see if they belong to the same individual. Examples are the comparison of results from positive controls, in forensics and more generally to check sample tracking when results are supposedly from the same individual. For diploid genotypes derived from (low-depth) DNA sequencing it is likely that only one of the two alleles might have been observed which can result in conflicting observed genotypes even when the individuals are the same. We show how between-result relatedness estimates and a new measure, the identity excess mismatch rate, can be used to allow statistical comparison of results.

Local haplotype assembly in Autopolyploid *Actinidia*

Tim Millar^{1,2}, Susan Thomson¹, John McCallum¹, Sam Baldwin¹, Phil Wilcox^{2,3}, Mik Black^{2,3}

¹Plant and Food Research, Lincoln

²Department of Biochemistry, Otago University

³Department of Mathematics and Statistics, Otago University.

Actinidia is a genetically complex group of interbreeding taxa with variable ploidy levels. The natural mechanism of polyploidisation in *Actinidia* is typically autopolyploidy, arising from the full duplication of a single genome. The vast majority of available statistical methods and software packages available for genetics and genomics are designed for diploid organisms assuming disomic inheritance and are not applicable to autopolyploids. We aim to develop and apply software tools that will build local haplotypes to enable better statistical analyses and prediction in *Actinidia* and other polyploids. Our local haplotype assembly method exploits variants that co-occur on Illumina short reads using Bayesian inference and can be applied to individuals or populations.

3: Methodological Advances

Identification of key drivers from complex network modelling

Wenting Liu¹, Yilei Zhang²

¹University of Otago, Christchurch, NZ

²University of Canterbury, Christchurch, NZ

It is always challenging to identify key drivers of the underlying mechanisms from multiple-level large-scale noisy omics data in human genetics. Here we developed a data analysis framework to identify the key drivers by fusion the complex relationships of various data via networks, which was successfully applied in multiomic data analysis. Firstly, a coherent network is built from deriving the coherence underlying multiple various data. Then the systematic modularity of the comprehensive coherent network is analysed to extract the best module of interest. The causality network such as Bayesian Network is then built within the interest module from the relevant data. Finally, the network analysis is implemented in the causal networks to derive the most important driving factors. This framework was successfully applied in fibrosis mechanisms modelling and identify novel and successfully validated key driver in the biological process of extracellular matrix. The same framework could also be applied to integrate heterogeneous relationships of the data such as temporal, spatial patterns, various features and contexts in complex systems i.e., ecology systems, social network, knowledge generation, etc.

3: Methodological Advances

Network-based Nonparametric Tests to Identify Genetic Modifiers of Niemann-Pick type C disease

Eliatan Niktab¹, Stephen Sturley², Ingrid Winship³, Mark Walterfang³, Paul Atkinson¹, Andrew Munkacsi¹

¹School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand

²Department of Biology, Barnard College at Columbia University, New York, New York, USA

³Melbourne Neuropsychiatry Centre, Royal Melbourne Hospital, Melbourne, Australia

Genome and exome sequencing has been extensively successful in identifying inherited disease-causing gene mutations, but there has been little success in deducing the variants that significantly modify disease progression. The reason is that current GWAS analysis tests only for linear genetic effects of single SNPs and explains only trivial fractions of heritability. Here we will investigate genetic variants that modify the causal gene of monogenic diseases and ultimately regulate the onset and progression of these diseases in individuals. Niemann-Pick type C (NP-C) disease, an example of a rare monogenic Mendelian disease, is one of more than 6,000 Mendelian diseases for which there is no cure. Most NP-C patients with the NPC1 mutation are diagnosed as late infants and die before or during adolescence, yet survival of some to adulthood provides a testbed for elucidating missing heritability in epistatic networks of disease-modifying variants. Therefore, we collected whole-genome sequences of pediatric and adult-onset patients. We then developed a novel pipeline that integrates an innovative mathematical model of genetic polymorphisms, augmented biological networks, clinical records, and semantic ontologies of GWAS data. Our approach mitigates the statistical challenge of sample sizes inherent to current GWAS methodology. Here we identify disease-modifying loci worthy of further characterization in cell and animal models of NPC disease.

4: Invited speaker

Session 4: Invited speaker

WGS to study patterns of genetic variation in human health and disease

Joep de Ligt¹

¹Institute of Environmental Science and Research Ltd, Porirua, New Zealand

We have used Whole Genome Sequencing (WGS) to study genetic variation in both healthy human adult stem cells (ASCs) and cancer cells. Our findings indicate that ASCs in different tissues can have very similar numbers of variants but can exhibit vastly different patterns and types of variation throughout life. Some of these patterns can be found back as driver mutations in cancers originating from these tissues but in many cases they are only part of the story. In addition to providing insights into the origins of cancer these mutational patterns can also be used to tailor cancer treatment and study the effects of different treatment regimes.

We studied 45 ACS derived from 20 healthy individuals from 3 different organs (small intestine, colon, liver) with ages ranging from 3 to 87 years old. Clonal expansion of single cells using the organoid culturing technology allowed high quality WGS of single cells to be generated without the need for whole genome amplification. The organoid technology was subsequently applied to the establishment of a breast cancer biobank of >100 primary and metastatic cancers, 33 of these were characterised by WGS, RNA sequencing and drug testing to study the potential impact on treatment by genomics data.

Session 5: Human and health genomics

Exploring Combinatorial Genetic Contributions to a Serious Adverse Drug Reaction

Wenting Liu, Yusmiati Liao, Simran Maggo, Matt Doogue, Martin Kennedy

Department of Pathology & Biomedical Science, University of Otago, Christchurch

Drug-induced angioedema has been reported to occur in response to a wide range of drugs and vaccines. However, the genetics of this serious adverse drug reaction (ADR) remains unknown. We have used whole genome sequencing (WGS) on a cohort of patients who have suffered angioedema, but failed to detect obvious, shared gene variants that may account for predisposition to this ADR. Instead of single variant pathogenicity identification, we aimed to examine more complex genetic pathogenicity and disease causing genetic mechanisms. Variant Combinations Pathogenicity Predictor (VarCoPP) is a recently published machine learning tool to identify pathogenic variant combinations in gene pairs from genomic features such as combined annotation dependent depletion (CADD), haploinsufficiency, recessiveness and biological distance, etc. We employed VarCoPP to explore possible gene pairs that may underlie risk of angioedema.

The WGS data were obtained from a cohort of 15 angioedema samples and 22 controls. 44 variants on 39 candidate protein coding genes and with minor allele frequency less than 0.05 were considered in combinations. 946 ($=44 \times 43/2$) combinations of variants were investigated using VarCoPP, where 893 were predicted as disease causing at 95% confidence, and 626 at 99% confidence. Out of the 626 disease causing gene pairs, 197 occurred in at least one angioedema sample, and 16 occurred in at least two angioedema samples. Interestingly, the top enriched pathways of the 16 validated gene pairs include circulatory system process (FDR=2.02E-6), renin angiotensin system (FDR=6.32E-6), peptide catabolic process (FDR=3.34E-5), multicellular organismal macromolecule metabolic process (FDR=3.96E-5), regulation of systemic arterial blood pressure by circulatory renin angiotensin (FDR=0.001), vascular process in circulatory system (FDR=0.002), endothelial cell proliferation (FDR=0.007), and vasodilation (FDR=0.009). These biological processes are very relevant to angioedema, which shows the power to explore complex genetic mechanisms by combining the genetic features in complex models as in VarCoPP.

5: Human and health genomics

Genetic interaction networks within genomes mediate individual drug response

Cintya Del Rio Hernandez, Eliatan Niktab, Bede Busby, Andrew Munkacsi, Paul Atkinson

Centre for Biodiscovery, Victoria University of Wellington.

Chemical genetic interaction network (CGIN) analysis of drug-gene interactions enables identification of new targets for combination therapies. CGINs are derived from genome-wide deletion libraries in yeast and, to a lesser extent, human cell lines. They are functional networks that may be analysed by functional enrichment and topology centrality metrics to identify network key bottleneck regulator genes and cellular processes. CGINs vary with individual genetic background and our laboratory has developed gene deletion libraries from several genetic backgrounds that are resistant to statin treatment allowing identification of strain-specific unfolded protein responses. Statins have anticancer activity (about 60 active clinical trials) but of unknown mechanism. To address this, we have created double-mutant networks in statin-susceptible and statin-resistant genetic backgrounds utilising query gene deletions of known relationship to the anticancer activity of statins. Specific CGINs we have elucidated vary in the three genetic backgrounds, information that allows us to pinpoint mechanism commonalities and also genes of individual response for further investigation. These CGINs will be discussed in this presentation for statins as well as the potential for any other compound.

Session 6: Microbial genomics

Genomics to understand the emergence of Buruli ulcer (*Mycobacterium ulcerans* infection) around Melbourne

Tim Stinear

Department of Microbiology and Immunology, University of Melbourne, Australia

Buruli ulcer (BU) causes a terrible, destructive skin and soft tissue infection that can lead to permanent deformity and disability. Buruli ulcer is caused by infection with the bacterium, *Mycobacterium ulcerans*. Cases of the disease are reported in more than 30 countries but have been exponentially increasing locally around Melbourne, with potential for a sustained epidemic to affect thousands of Victorians. Investigations over the past 15 years have shown that Australian native possums are wildlife reservoirs of *M. ulcerans* and that mosquitoes are likely vectors that spread the disease to humans. Detailed analysis of hundreds of *M. ulcerans* isolates obtained from patients over the last sixty years has revealed the likely origins and pattern of spread of *M. ulcerans* in Victoria. These new insights from research have created the first opportunities to test public health intervention strategies that will stop the spread of this disease.

6: Microbial genomics

Metagenome profiling of thousands of New Zealand sheep rumen samples

M.K. Hess¹, L. Zetouni¹, J. Budel¹, T. Van Stijn¹, H. Henry¹, R. Brauning¹, A. McCulloch¹, S. Hickey², A. Hess¹, M. Kirk³, G. Wood¹, P. Janssen³, J.C. McEwan¹ and S.J. Rowe¹

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

²AgResearch Limited, Ruakura Agricultural Centre, Bisley Rd, Hamilton, New Zealand

³AgResearch Limited, Grasslands Research Centre, Tennent Dr, Palmerston North, New Zealand

The rumen microbiome plays an important role in feed digestion and is associated with environmentally and economically important traits such as feed efficiency and methane production. We have developed a low-cost, high-throughput method for metagenome sequencing using restriction-enzyme-reduced representation sequencing, followed by either a reference-based or a reference-free approach for profiling the reads. The reference-based approach compared reads against the Hungate1000 Collection of genomes using BLAST, followed by taxonomic assignment of reads; while the reference-free approach counts the occurrence of a set of common reads within each sample. This approach has been used to generate metagenomic profiles from thousands of sheep and cattle rumen samples over the past year, which has allowed us to explore differences in rumen metagenomic profiles across breeds, ages, cohorts, diets and genetics. The goal of this project is to develop an approach that will utilize rumen metagenomics for selection purposes in a practical, agricultural setting.

Persistence of a poultry-associated lineage of Vancomycin-resistant Enterococci

Rushton-Green, R.¹, Darnell, R.L.¹, Tairoa G.¹, Carter, G.P.², Cook, G.M.^{1,3}, Morgan, X.C.¹

¹Department of Microbiology and Immunology, University of Otago, Dunedin, NZ

²Doherty Institute, University of Melbourne, Parkland, Australia

³Maurice Wilkins Centre for Biomolecular Discovery, Auckland, NZ

Enterococci are human and animal gut commensals. When vancomycin-resistant, they are also important opportunistic pathogens. Historically, agricultural use of the glycopeptide antibiotic avoparcin selected for vancomycin resistance in poultry, resulting in its discontinuation in agriculture in 2000. To better understand the phylogenetic relationships and antibiotic resistance patterns of human and animal-associated VRE strains in post-avoparcin New Zealand, we sequenced the genomes of 233 NZ VRE isolates (77 human clinical, 156 poultry) collected between 1998 and 2017. A clinically-relevant *E. faecalis* lineage (ST 108) was highly prevalent among both poultry and human isolates in the three years following avoparcin discontinuation, and has persisted for more than fifteen years since then.

Session 7: Māori kaupapa/Te Ao Māori in genetics teaching and research

Hei Whakamohio ai i nga Tauira Māori me Pākehā o te Mātauranga o te Ira: Raising understanding of Te Ao Māori in genetics education in Aotearoa/New Zealand

Phillip L. Wilcox¹, M. Husdon², K. Ruckstuhl³.

¹Senior Lecturer/Pukenga Mātua and Kaiwahina Māori, Department of Mathematics and Statistics/Te Tari Pāngarau me Tātauranga, University of Otago/Te Whare Wānaka o Otākou

²Associate Professor, Faculty of Maori and Indigenous Studies, University of Waikato

³Associate Dean Māori/Manutaki Tuarua, Otago Business School/Te Kura Pakahi, University of Otago/ Te Whare Wānaka o Otākou

There are significant gaps in secondary and tertiary education regarding the interface between modern gene technologies and Te Ao Māori. Despite Māori having long-held concepts of inheritance - many of which are still in use in contemporary Māori society - to date, very few have been incorporated into undergraduate genetics teaching in New Zealand universities. In addition, there is a marked underrepresentation of Māori relative to non-Māori in contemporary genetics research, despite the potential importance of modern genetic technologies for Maori health and for the Māori economy, and well as tools for enhancing kaitiakitanga. Moreover, the increasing importance of Vision Mātauranga and use of Māori ethical frameworks to guide genetics-based research with Māori communities and taonga species, has had a similarly low profile in both undergraduate and graduate genetics education. In the last four years, various efforts have been made to change this, including the Summer Internship of Native Peoples in Genomics Aotearoa, genetics modules taught in University of Otago's marae-based Science Wānanga programme, and inclusion of Māori-related content and assessment in genetics, biochemistry and statistics courses at both undergraduate and graduate level at the University of Otago. This presentation will describe these initiatives, including a range of encouraging outcomes that collectively are increasing Māori participation and responsiveness in genetics research and teaching in Aotearoa/New Zealand.

7: Māori kaupapa / Te Ao Māori in genetics teaching and research

Levi Collier-Robinson

School of Biological Sciences, University of Canterbury

7: Māori kaupapa / Te Ao Māori in genetics teaching and research

Ngā hua o te kānuka: the fruits of kānuka

Tia Haira¹, Storm Blockley-Powell¹, Michael Jackson¹, Irene Lopez-Ubiria², Alvaro Vidiella², David Monika³, Bella Paenga³, Manu Caddie³, Robert Keyzers^{4,5}, Andrew Munkacs^{1,5}

¹School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand.

²VLU Science, Gisborne, New Zealand.

³Hikurangi Bioactives Limited Partnership, Ruatoria, New Zealand.

⁴School of Chemistry and Physical Sciences, Victoria University of Wellington, Wellington, New Zealand.

⁵Centre for Biodiscovery, Victoria University of Wellington, Wellington, New Zealand.

Natural products are a robust source of drug leads in modern medicine. *Kunzea ericoides* (kānuka) is a taonga species with strong cultural significance and importance in Māori culture. Kānuka leaf extracts are used in rongoā (traditional medicine) to treat various skin ailments. However, the pharmaceutical potential of kānuka is unknown. To scientifically validate traditional knowledge and also identify biomedical potential not suggested by traditional knowledge, we conducted unbiased genome- and proteome-wide analyses using the genetic model *Saccharomyces cerevisiae*. These analyses, in the presence and absence of natural products, measure sensitivity of strains within the gene deletion library as well as the abundance and localization of proteins in the GFP library. The kānuka chemical genomic profile provides molecular bases for traditional and unsuspected activities, and these profiles are currently being monitored in fractionation experiments to identify the bioactive components. Together, these analyses provide molecular insight of taonga species, in this case the sustainable resource of kānuka as a source of pharmaceutically interesting bioactives. Overall, our goal is to develop a new bioactives industry to transform the East Coast region by creating high-value intellectual property to generate new jobs and income for a region that has had increasing unemployment and a declining population for decades.