

"I was privileged to have been among the first to embark on the adventure that has led to the structure of the human genome. It has been fascinating, every step of the way."

Emeritus Professor George Bouet Petersen (1933 - 2021)

Professor Petersen is considered the "father of DNA research in New Zealand" for bringing sequencing technology and techniques into the country and training a generation of scientists in DNA research.



Genomics Aotearoa Annual Report

Year 4

1 July 2020 - 30 June 2021

Table of contents

| Report from the Director | 4 |
|--|----|
| Report from the Chair | 5 |
| Who we are and what we do | 6 |
| Genomics Aotearoa in a nutshell | 7 |
| Vision Mātauranga A new co-governance model elevates Māori leadership within Genomics Aotearoa | 8 |
| Our Partners and Associates | 10 |
| 12. Health | |
| The Aotearoa Variome Capturing genomic diversity for better disease diagnosis | 14 |
| Epigenome-Wide Association Study A pipeline to identify personalised health indicators | 16 |
| Clinical Genomics Improving diagnostics of developmental disorders | 18 |
| CASE STUDY: Upskilling our clinical work force | 20 |
| case study: Syndactyly | 21 |
| Genomic Translational Oncology Towards the everyday use of genomics in cancer research and treatment | 22 |
| Rakeiora A pathfinder for genomic medicine in Aotearoa New Zealand | 24 |
| 26. Environment | |
| Environmental Metagenomics Revealing the tangled web of life in soils and water | 28 |
| CASE STUDY: Aquatic ecosystem function and diversity along a freshwater-to-marine gradient | 30 |
| case study: Blooming cyanobacteria | 31 |
| High Quality Genomes and Population Genomics for conservation Controlling invasive rats and wasps and protecting taonga and evolutionary icons | 32 |
| CASE STUDY: Sequencing the genome of wasps as a first step to novel biological controls | 36 |
| CASE STUDY: Genomic adaptations to a precarious environment: hot-spring cosmopolitanism | 37 |
| | |

38. Primary production

| High Quality Genomes and Population Genomics for primary production New technology, insights, and collaborations | 40 |
|---|----|
| сая втиру: High quality sheep genomes capitalise on existing knowledge | 43 |
| сая втиру: A high quality Bilberry genome for the benefit of Blueberry growers | 44 |
| CASE STUDY: Rewarewa becomes our first native tree with a high quality genome – and Māori honey producers could benefit | 45 |
| 46. Infrastructure | |
| Bioinformatics Genomics Aotearoa's core discipline | 48 |
| NeSI provides the computing power to support genomics projects | 49 |
| The Aotearoa Genomic Data Repository Culturally appropriate secure data storage, and management | 50 |
| CASE STUDY: ensuring indigenous people have authority over how their genomic data is used | 51 |
| Training Building capability of the next generation of genomics researchers and end-user | 52 |
| Summer Internship for Indigenous Genomics (SING Aotearoa) Mentoring emerging Māori genomics researchers | 54 |
| Te Nohonga Kaitiaki Guidelines for genomic research on taonga species | 56 |
| Public engagement | 56 |
| 60. New projects | |
| Genome Graphs to unravel pangenomes | 62 |
| Identifying the drivers of Streptococcus pyogenes to better understand disease causation | 62 |
| Invasomics for better biosecurity and invasive species management | 63 |
| Ruatau: connecting Māori genomic scientists and communities | 63 |
| Environmental Microbiomes, from virus to eukaryote | 64 |
| Scion | 64 |
| Contributing researchers | 65 |
| Glossary of genomic terms | 66 |

Report from the Director

I am delighted to present this year's Genomics Aotearoa report; the first we have made public since we began three and a half years ago. Over this time, the Genomics Aotearoa platform has worked to fill gaps, identify needs, and provide solutions that will ensure New Zealand makes the best use of genomic and bioinformatic techniques and technologies. I am delighted to present the emerging fruits of that labour, and to highlight how Genomics will transform our future.

I am writing this in my office/laundry room/spare bedroom at home. Not the best of places to write, but I am driven here by lockdown, the Delta wave, and the need to protect my family, friends, and colleagues from this generation's plague; COVID. I am, however, proud that Genomics, as a field, has stepped up to the challenge of the pandemic. Genomics produced the first genome sequences of the virus incredibly rapidly, provided the information on which the miraculously fast development of vaccines was based, and uses whole genome sequence to track and trace outbreaks and find new variants.

Genomics, as a way to track the virus, is now discussed by our Prime Minister, but it can do far more if we get the tools and technologies in place to support it. Genomics Aotearoa is committed to ensure those tools, technologies, techniques, ideas, and outcomes are developed,

are equitable in their use, and are available to and trusted by all New Zealanders. I think genomics will transform our world, we just need to ensure we use it to help make our world one that is fairer and more equitable.

In the past year Genomics
Aotearoa has gone through a
series of changes, including
staff in our management group. I
would like to extend my gratitude
to Angela Phillips, Ben Te Aika
and Lisa Early for their service,
and welcome Jayashree Panjabi,
Jess McLean and Tracey
Godfery to the management
team. Together with our
Communications Manager,
Claire Grant, our new staff have
been instrumental in developing
the report before you.

With our new staff on board, and new projects underway, I see a strong future for Genomics Aotearoa. I hope you enjoy reading what we have done and are inspired by our research and our researchers.

Professor Peter Dearden *Director of Genomics Aotearoa*





Report from the Chair

Genomics Aotearoa has reached the halfway point of its initial funding programme. Three and a half years is a good opportunity to take stock and reflect on the progress we have made in terms of delivery and benefit to New Zealand.

The COVID pandemic has brought public attention to the benefits of genomic sequencing. This report highlights the broad and diverse impact the work of Genomics Aotearoa is having on Aotearoa New Zealand and our drive to ensure its relevance and utility.

Like many, we have faced the challenges and disruptions to the normal interactions between science, end-users, and community created by the pandemic. Despite this, we have made good progress for which the team can be proud.

Aotearoa's unique population, ecosystems, biodiversity, and our primary industries which operate within and beside them are the focus and breadth of our activities. Central is Te Ao Māori. I am pleased to see the imminent strengthening of Māori representation at the board table – designed to give effect to the principles of Te Tiriti o Waitangi. While this has replaced the advisory role of the Kāhui, I would like to sincerely thank those Kāhui members for their service and erudite advice.

Use of genomics and bioinformatics internationally is accelerating apace. The highlights and case studies in this report exemplify the focus our team has on developing the tools and technologies that will enable Aotearoa New Zealand to keep in the vanguard of these game-changing technologies.

In reading this report you will appreciate the breadth of activities described is significant. It spans our themes of health, primary production, and the environment, and ranges from fundamental research to develop new tools, to actual applied outcomes of use in New Zealand. I would like to thank the board for their guidance and our Director Professor Peter Dearden and his management team for their hard work and expertise.

In particular I would like to give my warm congratulations to the scientists and researchers in Genomics Aotearoa for their contributions to our successes and encourage them to keep asking the questions and addressing the needs that drive genomics and bioinformatics research.

Dr William Rolleston

Chair of the Genomics Aotearoa Governance Board



Genomics Aotearoa was established in 2017 as an agile, leading-edge, and collaborative platform to facilitate and promote the international participation and leadership of Aotearoa's science community in the rapidly developing fields that involve genomic research. Genomics (the study of the complete genetic information in living things) and bioinformatics (the computational methods and software that enables analysis of genomic data) are the key scientific fields we facilitate.

Genomics Aotearoa's nine partners represent the genomics expertise of the entire country. Our partners are five universities - Auckland, Massey, Otago, Waikato, and Te Herenga Waka- Victoria University of Wellington, and four Crown Research Institutes - AgResearch, Environmental Science and Research (ESR), Manaaki Whenua - Landcare Research, and Plant & Food Research. We also have 32 associate organisations that represent researchers and endusers of genomics and bioinformatics across sectors that intersect with life science research. The inclusive nature of this consortium places Genomics Aotearoa in a prime position to lead Aotearoa's genomics research into the future.

Sector inclusiveness is vital to our success. Genomics Aotearoa has established national collaborations among genomics and bioinformatics researchers and end-users across key themes of health, environment, and primary production, which represent all the applied life sciences relevant to New Zealand's economic, environmental, and social wellbeing. All our projects align with at least one, and sometimes several, of these themes.

The connections of Genomics Aotearoa partners and associates with practitioners and end-users - including iwi and community groups, government

departments and Councils, District Health Boards and hospitals, breeders and primary producers, conservation practitioners - are designed to ensure that the capabilities and knowledge we develop translate to benefits throughout Aotearoa.

Underpinning all our research is the enabling infrastructure we develop and provide. Over 100 researchers and other specialists work with Genomics Aotearoa. We have developed a national genomics data repository and bioinformatics analytical platform, hosted by the New Zealand eScience Infrastructure (NeSI), and are actively growing the skills of researchers. To date, the platform has funded 17 postdoctoral fellowships to attract and retain worldclass early-career scientists. Genomics Aotearoa also offers extensive genomics and bioinformatics training to students, scientists, and clinicians, disseminating new methods and developing capability across the science community. Strong linkages worldwide ensure that we capitalise on and contribute to international best practice.

Principal funding for Genomics Aotearoa for 2017-2024 comes from the Strategic Science Investment Fund, following an application and assessment process managed by the Ministry of Business, Innovation and Employment.

Genomics Aotearoa in a nutshell

RESEARCH + RESEARCHERS

17 Postdoctoral fellowships



★ 3 new this year

170 Non-Genomics Aotearoa collaborators



156 International collaborations

9 Alliance Partners

◆ 4 Crown Research Institutes (AgResearch, Environmental Science and Research (ESR), Manaaki Whenua - Landcare Research, and Plant & Food Research) ◆ 5 Universities (Auckland, Massey, Otago, Waikato, and Te Herenga Waka - Victoria University of Wellington)

Research themes Health, Environment and Primary Production

- ◆ 131 Researchers and specialists from 21 institutions− 16 new this year
- 32 Associate organisations

35 Courses delivered to date



15 tauira and19 researchers

attended the **Summer Internship for Indigenous Genomics (SING)** this year

\$0

The **cost** of attending a Genomics Aotearoa **training course**

780

Trained
researchers
from 21
Universities
and Research
Institutes to date

TRAINED THIS YEAR:

445 participants in 17 events

13 Publications this year



12 Case studies

UTPUTS





51 Peer-reviewed publications

Supported by international-standard bioinformatics and data infrastructure.



Te Ao Māori is at the heart of Genomics Aotearoa's activities. We promote research undertaken by and for Māori, engage with and respond to community needs and mentor future Māori genomics leaders.

Our research practice and data curation are informed by tikanga. As a result, Vision Mātauranga is embedded throughout the successes of the projects summarised in this report. These are just a few of the highlights.

A new co-governance model elevates Māori leadership within Genomics Aotearoa

Over the last year, the actualisation of Vision Mātauranga within Genomics Aotearoa has been moving from a centralised single sub-group to a more diversified model. In May 2021, the partners to the Consortium agreed that the Governance Board would be increased to as many as six members (excluding the chair), with two to three being Māori, thus establishing a co-governance model.

It was agreed that the new Māori members will:

- Include one representative of mana whenua Ngāi Tahu
- Be experts in tikanga, or in connections between mātauranga and science research, or
- Have subject-specific knowledge, which may include genomics and bioinformatics, or may include end-user expertise in the theme areas of health, environment or primary production.

Vision Mātauranga is also embedded at the project level. This will be enhanced by having Board members with subject matter expertise who will approve projects. This two-tier approach elevates the status of Vision

Mātauranga, while ensuring its application across projects and themes as a whole.

Embedding Vision Mātauranga throughout our work: some examples

In the Health theme Rakeiora - a pathfinder for genomic medicine in Aotearoa/New Zealand embodies the approach that Genomics Aotearoa wishes to embed. Rakeiora seeks to unlock the potential of genomic and health information to benefit Māori, supported by tikanga-based ethical frameworks and principles for rangatiratanga - Māori governance over Māori genomic health resources and data.

While the Environment theme is looking for solutions for Predator-free 2050, cultural aspects are also considered. Kiore (*Rattus exulans*) can be a pest rat, but in Aotearoa is an important taonga species for some Māori.

If genomic-based technologies are to be used to control or eradicate other rats the potential impacts on kiore need to be understood. The project team is consulting with Māori on a kiore genome proposal whose aim is to minimise harm to kiore from genomics tools used to target other rat species.

In our Primary Production theme, the rewarewa genome has been sequenced to enable the burgeoning rewarewa honey industry to commercialise their product based on its whakapapa (provenance), as has been done for mānuka. Establishing a high-quality genome for this taonga species has specific cultural importance to Māori. This tree has many traditional uses such as medicines, for spiritual needs and in construction.

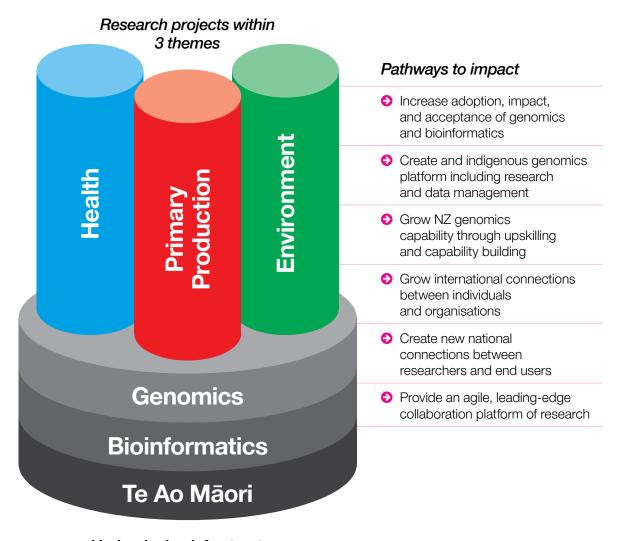
Appropriate protocols and engagement are imperative to Vision Mātauranga, and Genomics Aotearoa researchers are contributing knowledge on better ways to engage. For example, a paper published in the prestigious journal Nature Reviews Genetics by Maui Hudson and colleagues outlined kaupapa for working with Indigenous communities on genomic research around the world. In addition, Te Nohonga Kaitiaki (Place of the Guardians) was completed this year, laying the foundations for

tikanga-informed data curation by providing guidelines for genomic research in a way that honours taonga, kaitiaki and mātauranga Māori. The guidelines can also help hapū and iwi formulate protocols for research relevant to their own tikanga and mātauranga. As Vision Mātauranga is incorporated into the design of every project, examples of how it has been implemented can be found throughout the report.

As well as mentoring activities within projects, SING Aotearoa (the Summer Internship for Indigenous Genomics) has continued this year, with 15 tauira and 19 researchers discussing aspects of all our themes as well as genomics techniques. The interns were overwhelmingly positive about their experience. We are excited by the prospect of the new Ruatau project, which connects Māori genomic scientists who have grown up in Te Ao Pākeha to Māori communities for mutual benefit.

Mā te mōhio ka mārama, Mā te mārama ka mātau, Mā te mātau ka ora.

Through awareness comes understanding, through understanding comes knowledge, and through knowledge comes wellbeing.



Underpinning infrastructure

Our Partners and Associates

Genomics Aotearoa is an alliance of nine partners:

- University of Otago
- AgResearch
- University of Auckland
- Manaaki Whenua Landcare Research
- Massey University
- ESR
- University of Waikato
- Plant & Food Research
- Te Herenga Waka Victoria University of Wellington



















Our associates include:

- AbacusBio www.abacusbio.com
- AngusPure www.anguspure.co.nz
- AsureQuality
 www.asurequality.com
- Auckland City Hospital www.adhb.health.nz/auckland-city-hospital/
- Auckland University of Technology www.aut.ac.nz
- Beef + Lamb New Zealand Genetics www.blnzgenetics.com/
- Biomatters www.geneious.com/
- Bio-Protection Aotearoa bioprotection.org.nz/
- Bragato Research Institute bri.co.nz/
- Brain Research New Zealand www.brnz.ac.nz/
- Cardiac Inherited Disease Group www.cidg.org.nz/
- Cawthron Institute www.cawthron.org.nz/
- Garvan Institute of Medical Research www.garvan.org.au/
- Growing Up in New Zealand www.growingup.co.nz/en.html
- Hill Laboratories
 www.hill-laboratories.com/
- LabPLUS www.labplus.co.nz/
- Livestock Improvement Corporation www.lic.co.nz/

- Maurice Wilkins Centre for Molecular Biodiscovery
 www.mauricewilkinscentre.org/
- Multi-Ethnic New Zealand Study of Acute Coronary Syndromes
 www.fmhs.auckland.ac.nz/en/som/ about/our-departments/medicine/
 - www.fmhs.auckland.ac.nz/en/som/about/our-departments/medicine/ Heart-Health-Research/heart-health-research-menzacs-.html
- NIWA www.niwa.co.nz
- New Zealand eScience Infrastructure www.nesi.org.nz
- New Zealand Merino www.nzmerino.co.nz
- New Zealand's Biological Heritage National Science Challenge
 www.biologicalheritage.nz
- Orion Health orionhealth.com/nz
- Pastoral Genomics
 www.pastoralgenomics.com
- Prevarprevar.co.nz
- Real Time Genomics www.realtimegenomics.com
- Royal College of Pathologists of Australasia <u>www.rcpa.edu.au</u>
- Scion www.scionresearch.com
- Sophia Genetics
 www.sophiagenetics.com
- University of Canterbury www.canterbury.ac.nz
- Waitemata District Health Board www.waitematadhb.govt.nz

Our goal is to develop and apply genomics technologies to improve the health of all in Aotearoa.

The theme's current initiatives include understanding the variation in genomes among Māori, epigenome-wide associations, clinical genomics, and genomic translational oncology for better disease diagnosis and health outcomes. All our projects have strong links to communities, families, and clinicians to ensure broad adoption of the technologies we are developing.





Variomes describe the unique genetic signatures of populations. Without a comprehensive variome for Aotearoa, shaped as it is by our unique human history, genomic tools based on overseas non-Māori populations have limited use in genomic medicine for the increasing number of people with Māori ancestry.

The Aotearoa Variome project categorises Māori genomic variation, to improve disease diagnosis and to support healthcare research driven by Māori, for Māori. Problems with genomic data sourced from elsewhere relate to the complexity of genetic interactions. Sometimes variants associated with disease susceptibility in non-Māori populations have no relationship in Māori, and sometimes associations are completely unknown. Understanding the Māori variome is critical for effective, accurate, and equitable genomic diagnosis in Aotearoa. The Aotearoa Variome project seeks to address this health inequity.

Our planned outcomes include:

• A catalogue of genetic variants and their observed frequency among Māori.

- Analytical capability in human variation to support future genomics-based diagnosis.
- Integration of the Aotearoa Variome with genomedriven precision medicine through linkages with diagnostic services.
- Awareness and use of the Aotearoa Variome by healthcare researchers in the country.
- Increased trust in the application of genomic technologies by Māori communities.

Te Ao Māori is at the heart of the Aotearoa Variome project. An eight-person leadership rōpū ensures that research priorities are co-developed for the benefit of communities and that benefits arising from the data and kaitiakitanga are consistent with Te Ao Māori. The

kaupapa of Vision Mātauranga is firmly embedded throughout the project. The project's effective leadership rōpū is a highlight of this kaupapa.

Engagement and collaboration are keys to successes this year

The Aotearoa Variome team continues to engage with communities. Recruitment momentum is building with significant excitement from communities. Partnerships are embedded or are developing with several groups, including Ngāti Porou Hauora (Tairawhiti), Tuiora Health (Taranaki New Plymouth), the Centre for Health (Tauranga Moana), Poutini Waiora (Te Tai Poutini), and Hāpai te Hauora (Opanuku). Flexibility in the mode by which different communities wish to engage with their people (e.g., whānau consent with Tauranga Moana; community-based protocols with Hāpai te Hauora) reflects that diversity of practice across this kaupapa. The model of developing partnerships with Māori entities which already have established trusted relationships with the communities they serve has increased enthusiasm, acceptance of the kaupapa, and engagement.

International collaborations continue to flourish. Māori are not the only Indigenous group lacking population-specific genomic information. This global issue results from poor practices around engagement between researchers and Indigenous groups and the use and storage of Indigenous peoples' genetic data. The Aotearoa Variome team partnered with Indigenous researchers from Canada to compare issues and identify critical success factors for Indigenous genomic research. In addition to cataloguing variomes, the collaboration identified key requirements for the success of Indigenous genomic projects:

- Genomic health literacy and capacity of Indigenous peoples working in genome science to address future needs.
- Culturally safe guidelines, procedures, and policies for engaging communities and bridging the genomic divide.
- The rights, interests, and views of Indigenous participants to their data are protected and respected.

Other partnership work has captured cross-cultural insights into Indigenous perspectives on access to genomic data (see case study)², as well as identified barriers to effective diagnosis of rare diseases in Indigenous populations³.

On the technical side, the project is on track to sequence and analyse the first 200 genomes by the end of 2021, working with the Garvan Institute in Sydney for sequencing and NeSI for alignment and analysis. Pipelines for alignment, variant calling, storage, and cataloguing of a first phase variome are in place and continue to evolve as tools become more precise and efficient. Key organisations involved in protocol development and tool sharing include ESR, NeSI, Garvan, and Silent Genomes Canada.

The agreement with Garvan aligns with the project's kaupapa, including the return of residual samples and all data to the communities. Garvan have engaged in this approach and been very willing to adopt new ways of functioning to align with our requirements. The sequencing at Garvan in Sydney is the first large cohort of Māori DNA samples shipped overseas for wholegenome health research. Working with Garvan in this way will set in place tikanga that will serve genomics into the future.

Despite the achievements over the last year, the project has not entirely escaped the effects of COVID. One key initiative has been hampered. Four long-read genomes have been sequenced and the data repatriated from the Australian Genome Research Facility to develop a long-read sequence pipeline to detect variation not discoverable by the short-read method. The postdoctoral researcher is unable to come to Aotearoa, so the project plans to recruit a scientist based in New Zealand to develop this pipeline.

The in-depth period of relationship-building, progress in sequencing, and pipeline development are all major milestones for the Aotearoa Variome project team.

Our research team has been recognised both nationally and internationally for this project, with invited (online) presentations to the annual MapNet workshop and the Human Genetics Society of America annual conference⁴, and one of the project leads was profiled in Nature Communications⁵. Collectively these demonstrate that the approach used in the Aotearoa Variome is world leading.

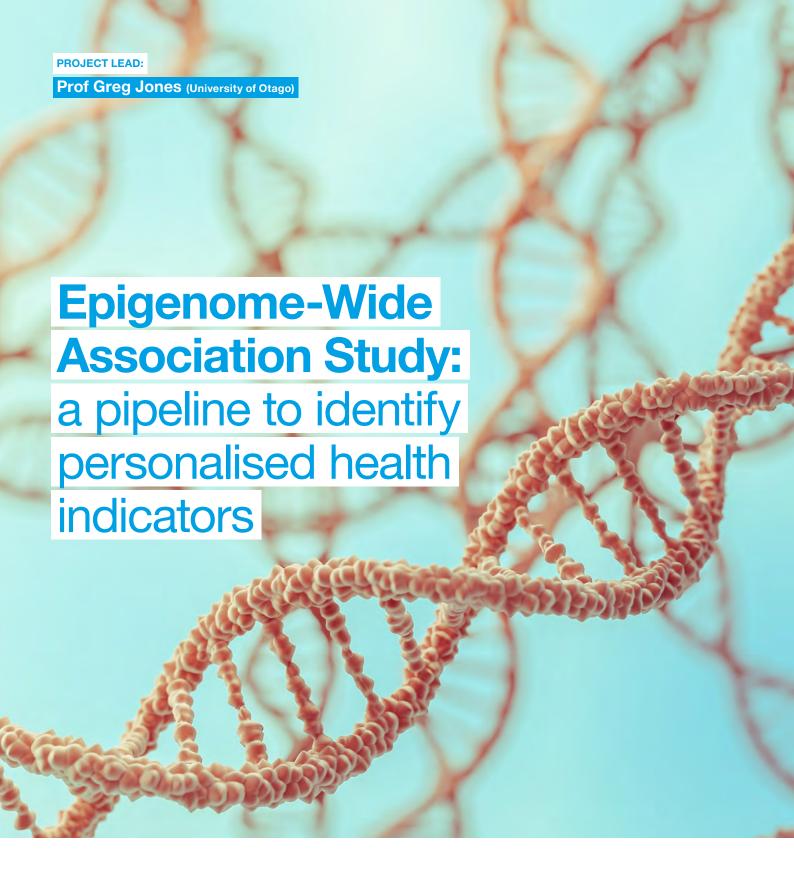
¹ Caron NR, Chongo M, Hudson M, Arbour L, Wasserman WW, Robertson S, Correard S, Wilcox P. 2020. Indigenous Genomic Databases: Pragmatic Considerations and Cultural Contexts. Frontiers in Public Health. https://pubmed.ncbi.nlm.nih.gov/32391301/

² Hudson M, Garrison NA, Sterling R, Caron NR, Fox K, Yracheta J, Anderson J, Wilcox P, Arbour L, Brown A, Taualii M, Kukutai T, Haring R, Te Aika B, Baynam GS, Dearden PK, Chagné D, Malhi RS, Garba I, Tiffin N, Bolnick D, Stott M, Rolleston AK, Ballantyne LL, Lovett R, David-Chavez D, Martinez A, Sporle A, Walter M, Reading J, Carroll SR. Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. Nat Rev Genet. 2020. 21(6):377-384.

³ DAngelo CS, Hermes A, McMaster CR, Prichep E, Richer É, van der Westhuizen FH, Repetto GM, Mengchun G, Malherbe H, Reichardt JKV, Arbour L, Hudson M, du Plessis K, Haendel M, Wilcox P, Lynch SA, Rind S, Easteal S, Estivill X, Thomas Y, Baynam G. Barriers and Considerations for Diagnosing Rare Diseases in Indigenous Populations. Front Pediatr. 2020. 8:579924. doi: 10.3389/fped.2020.579924. PMID: 33381478; PMCID: PMC7767925

⁴ American Society of Human Genetics (ASHG) meeting in 2020 (online) – Session 46 Indigenous Biobanking: Global Perspectives on Keeping Our Genomic Data Local: see https://www.abstractsonline.com/pp8/#!/9070/session/46

⁵ see https://natureecoevocommunity.nature.com/posts/celebrating-national-dna-day-population-genomics



Epigenome-wide association studies (EWAS) aim to understand the molecular basis of disease risk. While genome-wide associations (GWAS) define the inherited risk, EWAS investigates how this risk is modified by changes in gene regulation, i.e., how, and when those genes are switched on (or off). These epigenetic switches may be due to environmental exposure or lifestyle factors.

Epigenomics can provide insights into a wide range of health-related problems including cardiovascular disease, cancer, diabetes, obesity, ethnicity related differences, very low birthweight, and placental insufficiency. Epigenetic modifications (or marks) may be useful predictors of health and disease. DNA methylation is an epigenetic mechanism that modifies the expression of genes and therefore their effects. Analysis of these marks promises accurate, personalised indicators of lifetime exposure to environmental stressors and may make a useful contribution to precision disease risk prediction. While potentially very useful, identifying these marks is challenging as they change from cell to cell, tissue to tissue, and over time.

Aotearoa has several very well characterised cohort studies of disease which could be usefully analysed by EWAS. However, these analyses are not yet well established and so new pipelines and tools are needed. This is the aim of our EWAS project. Epigenetic marks may be useful in prediction of many traits, such as reproduction potential, and so this project has potential links into the Primary Production and Environment themes.

The goal of our EWAS project has been to develop effective, accessible, and standardised pipelines to assess epigenetic data in large cohort studies, that incorporates best practice for data quality assurance, analysis, meta-analysis, and multivariate modelling.

A new pipeline for EWAS in Aotearoa

The Epigenome Wide Association Study Pipeline (EWASP) web-based tool has been developed and updated with further enhancements over the course of this project. This tool is being hosted on a New Zealand eScience Infrastructure (NeSI) server. A new component helps identify polymorphic influences across a large number (800,000) of methylation sites. This feature is particularly relevant when examining multi-ethnic populations - such is the case in Aotearoa New Zealand. The EWASP will therefore be an important contribution to our collaboration with the Multi-Ethnic New Zealand Study of Acute Coronary Syndromes (MENZACS) consortium, with an initial analysis including participants of Māori, Pasifika, Indian, and European ancestry.

The project is an excellent demonstration of Genomics Aotearoa's efforts to build the genomic and bioinformatic capability Aotearoa needs. Although initially developed to help interpret information from Principal Investigator Greg Jones' unique cohort of vascular studies⁶, the pipeline has the potential for use in any genome – not just in human health but also in farmed animals and potentially in conservation

too. The EWASP tool has already been tailored to process DNA methylation results from mouse DNA. Other species options, including sheep and cattle, are being developed.

Knowledge sharing with the genomics community of Aotearoa

The EWASP tool is being promoted through the Aotearoa genomics community to enable the best use of information from the genomes of people, plants, animals, and bacteria specific to our contexts. Events have included *Epigenome-Wide Association Study Pipeline Training* for 27 participants and the *Otago Epigenetics Symposium* with 60 participants in March 2020, together with a Genomics Aotearoa seminar on *Developing an online epigenome-wide analysis tool:* Lessons learnt along the way in May 2020.

Monthly meetings have been held with the Multi-Ethnic New Zealand Study of Acute Coronary Syndromes (MENZACS) Epigenetics Analysis Group on using the EWASP tool. The analysis principles underpinning the EWASP tool are also the subject of a book chapter.

Looking to the future: equity, diversity, and new initiatives

Several projects have been initiated that use the EWASP tool for disease research, with a number of these working with Māori and Pasifika peoples. For example, we are currently involved with the development of a Māori researcher project (PI Melanie Cheung, entitled He Ara Hauora). We have also secured Healthier Lives National Science Challenge funding for a project which will include a Māori-specific investigation of vascular disease epigenetic markers. In both cases the studies are being co-designed. The EWAS project aligns strongly to Genomics Aotearoa's kaupapa Māori centred approach.

Just a few of the other epigenetics initiatives already underway include: The Christchurch Heart Institutes' Pasifika Heart Study, which investigates DNA methylation associations with cardiometabolic and dietary traits; The MENZACS consortium uses EWASP to investigate coronary disease outcome-related markers; The Christchurch Health and Development Study uses EWASP to assess epigenetic changes associated with low birth weight; DNA methylation profiling in meningioma brain cancer is being assessed with EWASP to identify potential prognostic (predictive) biomarkers; and DNA methylation in Lymphangioleiomyomatosis, a form of breast cancer, is also being assessed using EWASP.

⁶ Jones GT, Marsman J, Bhat B, Phillips LV, Chatterjee A, Rodgers EJ, Williams MJA, van Rij AM, McCormick SPA, DNA methylation profiling identifies a high effect genetic variant for lipoprotein(a) levels. Epigenetics. 2020. doi.org/10.1080/15592294.2020.1739797



Genomic diagnosis in clinical medicine is well established elsewhere, but Aotearoa lacks diagnostic capability and infrastructure, which forces clinicians to look overseas for sequencing and analysis. Building infrastructure and capability in genomics diagnosis is critical to the future of genomically informed healthcare in Aotearoa.

The Clinical Genomics project was completed this year, and aimed to enhance skills, tools, and pipelines to support genomic diagnostics in healthcare delivery through collaborative relationships between laboratories and clinicians.

The project took the approach of applying genomic diagnostics to a group of families with a child (or children) with an undiagnosed developmental disorder. Ensuring equitable access was a major focus, and communities with a high proportion of Māori and Pacific families were encouraged to participate. Laboratory scientists from District Health Boards (Auckland, Capital & Coast, and Canterbury DHBs) were offered the opportunity to analyse genomic datasets from patients with real clinical problems who present to genetic services, with the ability to deliver results back to patients.

Transferring the datasets to DHB-located laboratories provides diagnostic laboratory scientists with opportunities to expand their analytical skills and capability in a collaborative way. In turn, physicians and genetic counsellors involved in clinical genetics services can draw on this analytical expertise to then develop their own skills in clinical genomics, replicating the multidisciplinary nature of international genomic diagnostic teams. Building familiarity and capability in genomic testing represents a step-change for national laboratory services and will hopefully be a catalyst for further health funding investment.

The importance of equity, collaboration, and engagement

From the outset, the project team sought to increase equity in clinical genomics participation. The project team engaged the support of Genomics Aotearoa's Māori advisors, so that the clinical team could recruit Māori families appropriately and effectively. The project was completed in April 2021, with participation from 51 families, 22% of whom are Māori and 6% Pasifika. Sequencing for 60 of the 150 genomes shared with the project, which took place this year, was generously funded by Cure Kids. A qualitative study with three families who received a positive diagnosis of syndactyly through this programme has recorded their perceptions of the impact of the project on them.

Interactions with DHB scientists and the formation of cohesive national collaborations has been a project highlight, particularly the growing enthusiasm and capability for full genome sequencing within DHB labs, and their formalisation of methods to analyse full genome

datasets. A full team of clinicians, laboratory scientists, and the researchers in Dunedin participated in the project. Together with the DHBs, participants included Canterbury Health Labs, Lab Plus, and Genetic Health Services NZ (Auckland, Wellington, and Christchurch).

Capability development for clinicians has included enabling full genome data transfers and genome analysis within the ADHB and CDHB infrastructures, assembly of pipelines for interpretive analysis in ADHB, CDHB, and at ESR (who assisted CCDHB with their analyses) and the ability for scientists and pathologists at CDHB to soon offer a genome sequence interpretive service.

A high rate of diagnosis

The project researchers arrived at a diagnosis for 16 out of the 51 (31%) whānau who participated, which, given the nature of the study, is a high rate. A further six families (12%) had potential diagnoses and 57% were without a diagnosis at the end of the project. The project detected a new variant associated with an intellectual disability, speech delay, and dysmorphism syndrome⁷. Time-to-diagnosis for all referred cases progressively shortened over the course of the project, with some centres establishing shorter timelines than others.

Development of protocols, tools, and pipelines

Protocols and tools have been publicly shared⁸ with national and international groups with experience in medical genomics to ensure pathways, analytics, and formats are aligned with best practice. All participating labs are establishing practice using this approach, one in close collaboration with Genomics Aotearoa associate, Sophia Genetics, another with an Agilent based analysis programme.

The project has established pipelines for alignment, variant calling, storage, variant curation, and diagnostic validation protocols including optimising and comparing bioinformatics tools and pathways with national and international experts. All participating diagnostic labs have established their own pipelines, with some heterogeneity in practice.

Overall, the outcomes expected at the start of the project were exceeded. Next steps will include extending the analysis to singleton cases where one or both parents are not able to offer samples, linking whānau who have diagnoses to appropriate support services to maximise the dividend of a diagnosis and piloting the integration genomics analysis of this type into medical school curricula.

⁷ Chopra M, McEntagart M, Clayton-Smith J, Platzer K, Shukla A, Girisha KM, Kaur A, Kaur P, Pfundt R, Veenstra-Knol H, Mancini GMS, Cappuccio G, Brunetti-Pierri N, Kortüm F, Hempel M, Denecke J, Lehman A; CAUSES Study, Kleefstra T, Stuurman KE, Wilke M, Thompson ML, Bebin EM, Bijlsma EK, Hoffer MJV, Peeters-Scholte C, et al. Heterozygous ANKRD17 loss-of-function variants cause a syndrome with intellectual disability, speech delay, and dysmorphism. Am J Hum Genet. 2021. 108(6):1138-1150.

⁸ https://www.ga4gh.org/genomic-data-toolkit/



A lack of both capability and infrastructure here means clinicians have typically been getting their sequencing and analysis done overseas. Fostering local capability in genomics diagnostics is therefore critical for Aotearoa to exploit the ever-growing genetic medicine opportunities being developed globally.

Building familiarity and capability in genomic testing is a step change for national laboratory services, and one leading the charge is the Canterbury District Health Board.

The DHB has been working with Clinical Genomics lead researcher Professor Stephen Robertson, developing up existing resources to analyse some of the real-world genetic-based problems the project has been researching.

"While genomics diagnosis is already being done in well-resourced countries, it's a big process for us to get this underway here – firstly to have more people on the ground with the skills needed, and also to work out how to incorporate this into laboratories, then into the wider existing health services," he explained.

"Working with Richard King, Clinical Director of Chemical and Genetic Pathology at the Canterbury DHB, we've shown in a small way that we can do whole genome analysis here. Developing the processes and the skills needed has primed the pump, now it's over to these laboratories to work out how to take this forward from small scale."

Building capacity will result in quicker and cheaper diagnostics, which will mean Aotearoa's researchers and clinicians will not need to send sequencing and analysis work offshore.

And that paves the way for more and faster genetic and genomic testing, with long-term and lasting impacts on health care in Aotearoa.

"We're on the cusp of something very big for Aotearoa; all credit to the DHB professionals we have been working with to help get this off the ground," Professor Robertson said.

Led by Prof Stephen Robertson (University of Otago)



Rarer, more severe forms of syndactyly can significantly impair hand and foot function and have more far-reaching implications and often complex surgery is required. Syndactyly does not stop people from working and leading fulfilling lives – there are surgical procedures that help - but aesthetically it is a visible condition.

One New Zealand family who is fully aware of the effects that syndactyly has on generations has been taking part in the Clinical Genomics project.

Three generations of the family have various forms of fusion of fingers and toes. The project participant's son has fused fingers and an extra finger, and has six toes, and had corrective surgery in Australia. Her father had two fingers on each hand amputated. Her sister had unsuccessful surgery on her hands as a baby. She shared with us her family's experiences with syndactyly and why she took part in the project.

"My father struggled with the stigma and that had an impact on the way he lived his life. I found out that my son had the condition via a scan when I was pregnant – which gave us plenty of warning. We were determined to instil a positive attitude in his upbringing, and his sense of humour has helped him through the social and practical aspects of having webbed fingers and toes.

We were already aware this genetic condition was being passed on in our family. But what we didn't know how syndactyly was inherited and what that meant for future generations. What are the risks? Who would inherit and why would some not? So, I jumped at the chance to understand more about the specific gene variant that caused this, in my family.

Now that I know, I have to say it makes a difference. It hasn't made it go away; it's just given us slightly more control.

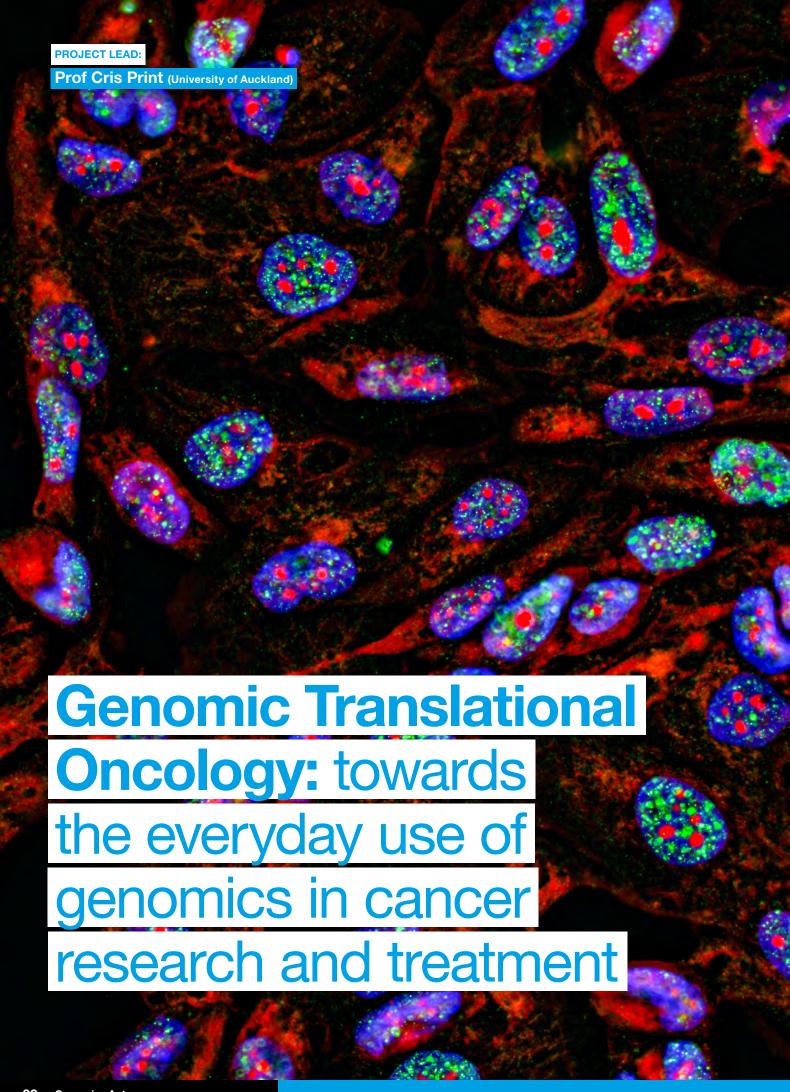
There is a test to identify the gene and my daughters, and my extended family will now have the option of having the test to see if they are a carrier. It will be their choice to find out, if they have children. Every circumstance is different, but I believe knowledge is power.

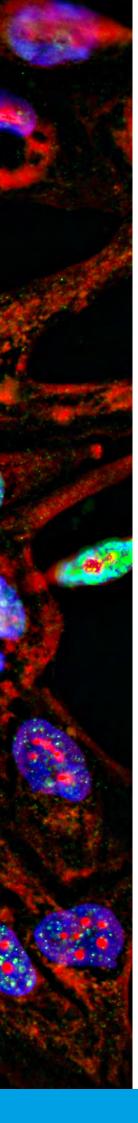
It was a positive experience and I'm pleased we went down this path – from this basic research we have more understanding of this genetic abnormality, and that will help our families, and other families also. I see knowing as control in an otherwise uncontrollable world, and that has given me some comfort.

And who knows what will happen in the long-term – knowing more about the gene may allow drugs to target it, or it may direct future researchers to use CRISPR technology to edit it. I see genetics as the future of medicine, and our family has contributed to this in a small way."

It is extremely rewarding for the project team to have such an appreciation for the work they are doing and encourages the team to look for solutions for this woman, her family, and others in the community.

Led by Prof Stephen Robertson (University of Otago)





Better understanding of the biology of tumours and potential treatments is crucial to clinical oncology, but large-scale genomic analysis is not yet benefiting people with cancer in Aotearoa as much as in other countries. This project coordinates and translates genomic technologies for precision oncology research and practice.

While a large amount of next generation sequencing data is being accumulated across the country for patients with a wide variety of cancer types, this has been occurring through several disconnected projects. The lack of coordination of the tools and technologies used makes comparisons among studies difficult. Moreover, translating research findings into consistent improvements for clinical management have been inhibited by a lack of cohesive, agreed protocols and analytical approaches. The goal of this project has been to generate and facilitate use of harmonised tools and technologies to enhance translating research findings into clinical practice. This will smooth the way for the everyday use of genomics in clinical oncology.

Rather than reinvent the wheel, the project's approach has been to adopt and tailor overseas best practice for Aotearoa, being mindful of the differences and opportunities that our health environment offers and undertaking this work as a partnership with Māori. The team developed extensive relationships with overseas clinicians and researchers to assist with this initiative. particularly colleagues in Australia and the United Kingdom. The approach also leverages data already collected in National Science Challenge and HRC-funded projects. Partnering with clinical teams ensures appropriate translation of the research into clinical practice.

Partnerships are critical to project success

The Genomic Translational Oncology project enjoys a fruitful collaboration with District Health Boards in the Auckland region. This partnership provides a rare opportunity for the research to be co-developed with, and beneficial to, clinical end-users. Although subject to the ups

and downs of the pandemic, the collaboration means the result will be a strong platform for future researchers.

The project's activities have provided a base for development of precision genomic medicine in the Rakeiora Pathfinder programme with Māori, particularly working with the Ira Tātai Whakaheke rōpū. The ground was also prepared for follow-on precision genomic medicine initiatives that would not have been possible without the techniques, training, and relationships built through this work. Other collaborations are in place with the Healthier Lives National Science Challenge ctDNA in cancer programme, the Network! Cancer programme, the MoST clinical trial, and the University of Auckland Genomics into Medicine programme.

As a result of COVID, genetic counsellor, clinical scientist, and oncology contribution to the project was delayed. To overcome this challenge, the genetic counselling roles were initially replaced with a registered clinical laboratory scientist, to bring similar benefits to the project's formulation of national pathways. Input to ethical and practical consenting issues was provided by the Principal Investigator, in consultation with the genetics service staff.

In this final year of the project, five peerreviewed articles have been published,
including use of big data to estimate
partial cancer prevalence in Aotearoa9.
The outcomes of the project's activities will
continue to evolve through the Genomics
Aotearoa Rakeiora project. This project has
generated understanding of the high value,
as well as the necessity, of partnering with
Māori in precision genomic medicine. It has
resulted in the assembly of advanced genomic
pipelines, in collaboration with our partner
organisation NeSI, to analyse tumours and
is contributing to the generation of clinically
accredited cancer genomic tests.

⁹ N Brewer, J Atkinson, P Guilford, C Print, T Blakely, A Teng. An estimate of partial cancer prevalence in New Zealand using 'big' data, New Zealand Medical Journal, 133(1514), May 2020



The goal of Rakeiora is to unlock the potential of genomic and health information for research in Aotearoa in an explicit and deliberate partnership with Māori. The team have implemented tikanga-based ethical frameworks and are generating a best-inworld culturally informed computational infrastructure.

This infrastructure will enable research while allowing individuals and whānau, as well as communities, iwi, and mātāwaka, to maintain control over their genetic information. Importantly, it embeds principles of Māori governance over Māori genomic health resources and data.

Rakeiora uses action learning to explore options for the acquisition, use, and storage of genomic datasets linked to clinical information for use in healthcare research in two healthcare projects - Primary Care (community healthcare led by a nurse or GP) and Tertiary Care (for advanced cancer treatment in a specialist facility).

Rakeiora's Primary and Tertiary Care projects have been guided by the Ngati Porou Co-design rōpū and

the Ira Tātai Whakaheke rōpū, respectively. Together, these rōpū have led transformative scientific, ethical, and infrastructural advances, which we believe can provide a community-mandated template for Aotearoa's future genomic science, ethics, and data infrastructure. Ira Tātai Whakaheke provides project governance for the Tertiary Care component, while Ngāti Porou Hauora provides governance over the Primary Care component.

Adapting the methodology based on learnings

The action learning approach has resulted in the evolution of methods and protocols as participants, stakeholders, and researchers share perspectives and ideas. Significant methodological improvements include:

- Co-governance by Te Ira Tātai Whakaheke (ITW) for the Tertiary Care project. Kawa and tikanga inform genomics research by Māori in collaboration with others, with respect for Māori data sovereignty.
- Division of the Primary Care project into two parts (linking of clinical records with pharmacogenetic data and exploring the use of genealogical and DNA data to improve diagnosis and risk stratification of patients) to facilitate action learning, community codesign, and future use of accumulated data in research.
- Adaption of a 2-layer IT infrastructure with an authorisation/permitting platform for researchers and a "secure walled garden", in which clinical and genomic data are analysed. A mixture of bespoke and internationally sourced tools will be deployed across these two environments over time.
- Creation and adoption of written and audio-visual tools for engagement with the Ngāti Porou Hauora and tertiary care patient communities to learn from their perspectives on the precision medicine research proposition.

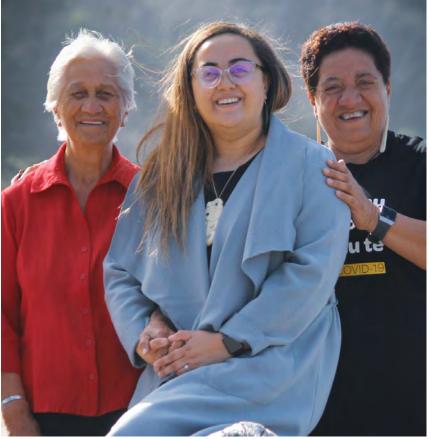
It has taken more time than anticipated to capture the nature and extent of Māori attributes for this kaupapa, and the pandemic has also contributed to delays in engaging participants. As a result, the number of participants is less than originally planned. However, this extended timeframe has resulted in the creation of tools and resources to enable the necessary social licensing for this project.

The project is highly collaborative. Our partners include ESR, University of Auckland, University of Otago, and Ngāti Porou Hauora. Sub-contractors include Auckland DHB; Animation Research Limited for effective communication with participants; Cancer Trials NZ and the National E-Science Infrastructure (NeSI) to co-develop nationally scalable computational tools. International collaborations include Global Alliance for Genomics and Health (GAG4H, https://www.ga4gh.org/), together with Melbourne Genomics; Silent Genomes

Canada; and ongoing advice from world-renowned statistical geneticist Professor Bruce Weir (University of Washington) who visited in February of this year.

As the Rakeiora project enters its second year we will complete recruitment of initial participants and have begun consultation to engage two additional iwi/ hauora organisations. With significant technical work planned, including genome sequencing and generation of pharmacogenomic data, an IT platform for clinical metrics and related analysis, it will be a very busy year for the project. At the end of the project this collaboration will emerge with clear guidelines (including case studies) for how a scalable environment that facilitates research and joins clinical, genomic, and public health data can be safely constructed and deployed within a unique Aotearoa context.

The Rakeiora project has been internationally promoted via presentations by team members. Its internationally unique status has led to invited presentations at national and international conferences in 2020 and 2021¹⁰. In addition, the experiences in this project have informed additions to courses on study design (statistics) and genetics at the University of Otago to include key aspects of culturally informed codesign in Indigenous genomics context. Dr Phil Wilcox (Ngāti Rakaipaaka, Rongomaiwahine, Ngāti Kahungunu ki te Wairoa), one of the project leaders who teaches this content, was awarded the Genetics Society of Australasia Excellence in Education in August 2021.



¹⁰ For example: American Society of Human Genetics (ASHG) meeting in 2020 (online) – Session 46 Indigenous Biobanking: Global Perspectives on Keeping Our Genomic Data Local: see https://www.abstractsonline.com/pp8/#!/9070/session/46

Our Environment theme unravels the genomic secrets of the smallest and largest of Aotearoa's species.

The projects that fall under the Environment theme have sequenced the metagenomes of multiple species in aquatic environments along environmental gradients, large, complex insect and plant genomes, and key invasive species such as wasps and rats. Along the way a significant number of new tools, technologies and pipelines have been explored, refined, and shared among the genomics community of Aotearoa.





From precarious environments like hot springs to flourishing algal blooms, a myriad of tiny organisms affects the functioning of our waterways and soils. Environmental Metagenomics focuses on these species, their diversity, interactions in nature and evolutionary adaptations. Although their genomes may be small, disentangling who is who in the microbial soup (and what they do) is the challenge this project tackles.

Micro-organisms are a major component of the environment – determining their role reveals their impact on environmental health. A better understanding of Aotearoa's microbial ecosystem has major opportunities for our environmental, primary production, and health research.

A powerful way to study microorganisms is by direct sampling of their DNA and RNA from the environment. Metagenomics, the study of the genetic material recovered from environmental samples, allows sets of complete or partial genomes to be disentangled from complex mixtures of the variety of microbes inhabiting it.

While metagenomics approaches are rapidly growing internationally, Aotearoa has limited capacity to apply them to environmental management. This project fills a technology and capability gap in genomics to contribute to improved environmental monitoring. Our goal is to reconstruct high-quality genomes from environmental samples to study genome adaptation across chemical and physical gradients and microbial ecosystem functioning; and develop sharable workflows and training resources.

Three key initiatives have been to: determine aquatic ecosystem function and diversity on a freshwater-to-marine gradient; categorise bacteria responsible for "algal" blooms; and identify genomic adaptations in hotspring organisms. Our case studies describe this work in more detail.

The project has also sequenced a river-to-sea virome - a unique set of viral genomes along the gradient. Viromes are the characterisation of all the DNA and RNA viruses within a microbial community. These genomes will give us a better understanding of how viral populations affect the productivity of microbial communities in our water and soils.

New biological knowledge leads to new technologies for the genomics community

To generate their discoveries, the Environmental Metagenomics team developed a range of protocols to improve environmental genomics research and its accessibility.

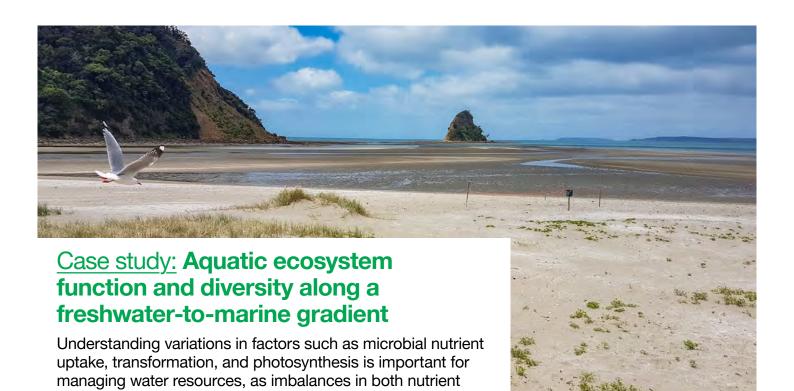
New workflows that have been developed are all publicly accessible via the Genomics Aotearoa GitHub webpages. The workflows include: prokaryotic genome binning, metagenome annotation, and transcriptomics (https://github.com/GenomicsAotearoa/

environmental_metagenomics); Metagenomics Summer School workflows from data quality control to genome analysis (https://github.com/GenomicsAotearoa/methods-and-musings).

The knowledge gained through the development of these workflows has been shared widely in partnership with NeSI. More than 70 Aotearoa-based researchers from 14 institutions (all seven universities and four CRIs [GNS, AgResearch, Plant & Food Research, Manaaki Whenua - Landcare Research], Ministry for Primary Industries, Cawthron, and Livestock Improvement Corporation) participated in Metagenomics Summer Schools from 2019 to 2020. The net was spread even more widely by providing support for participants to train other researchers within their groups via open access and access upon request NeSI-based course resources and computing. The course was ranked highly by participants, with 100% ranking it as "excellent" (76%) or "good" (24%). Several participants immediately applied the skills learnt during the course on their own projects, and the course resulted in several new collaborations.

The progress of the project identified a need for more tikanga-informed guidelines microbiological research. Preliminary work is continuing as part of the new Environmental Microbiomes project, which promises to deliver many new insights into environmental genomics.





Genomics creates new and exciting opportunities for aquatic biodiversity studies.

The Genomics Aotearoa team based at the University of Auckland has generated a detailed reconstruction of ecosystem composition and function from river-to-sea, following the Waiwera River near Auckland to its estuary and finally beach environment.

availability and photosynthesis can lead to poor water quality.

Our estuaries are a key interface between the terrestrial and marine environments, funnelling freshwater and nutrients drained from catchments offshore. To determine the function and diversity of the freshwater-to-marine aquatic ecosystem the project team generated detailed genome- and transcriptome-based reconstructions of microbial communities. The project discovered a strong spatial and taxonomic division of metabolic processes resulting in variation in community contributions to aquatic biogeochemistry for ecosystem productivity, nutrient uptake, and nitrogen removal. Extraordinarily rich aquatic communities.

An extraordinary level of detail has been discovered about the composition of aquatic life across the freshwater-to-sea gradient. Although aspects of this have been done in other studies here and overseas, it's the first-time detailed descriptions have been done on this scale.

"We've filled in a lot of gaps, gaining a good understanding of the roles performed by different groups of microorganisms, including methods for obtaining the nutrients they require in water and sediment. Our explanations include the assimilation and transformation of the essential nutrients (and sometimes problematic environmental contaminants) nitrogen and phosphorus; photosynthesis, which supplies the ecosystem with organic carbon; and strategies for tolerating salt stress," lead researcher Kim Handley said.

Because the tidal lagoon estuarine system studied is fairly typical and not severely contaminated, this provides

a benchmark, and findings can readily be applied to other New Zealand waterways. It also offers insights for similar river-estuary gradients across the globe.

Variation in biochemical processes across the gradient.

This study is one of the first to directly compare metabolic processes in waterborne free-living microorganisms and the underlying soft sediments in an aquatic system, and the first across a freshwater-to-marine gradient. This has shed a light on the different biochemical processes at work.

Freshwater microbial communities and metabolic processes are highly distinct from those in the saline (brackish and marine) environment, as are microbial communities in water versus sediment. This is due to significant differences in salt, nutrient, and oxygen availability requiring distinct strategies for osmo-regulation (control of water balance in cells), nutrient acquisition, and metabolisms reliant on oxygen or oxygen-free environments.

Genome predictions of differences in microbial communities along the gradient indicate that bacteria differentially evolved amino acids that enable tolerance of small increases in salinity, primarily increases in acidic amino acids (the building blocks of proteins). This is a feature not previously known for organisms inhabiting common, but relatively low salinity environments (e.g., brackish, and marine). Increases in acidic amino acid content increase protein stability in the presence of salt. Thus, the incremental changes to genomes and predicted proteomes across the salinity gradient help to explain the highly distributed nature of organisms between freshwater and marine habitats. Results illustrate a strong link between bacterial genome characteristics, habitat range, and lifestyle.

Led by Dr Kim Handley (University of Auckland)



Cyanobacterial blooms in rivers cause a reduction in water quality and are a threat to animal and human health.

The project (in collaboration with the Cawthron Institute) determined key genetic mechanisms (diverse nutrient acquisition strategies) that enable members of the cyanobacterial genus *Microcoleus* to proliferate in freshwater habitats.

So, how do cyanobacteria live in their natural habitat, how do they coexist with other bacteria and microbial life forms, and what makes them so good at forming blooms?

Genomics Aotearoa has been funding stream-to-ocean microbiome research to better understand microbial life. Through a collaboration between the University of Auckland School of Biological Sciences and the Cawthron Institute, Genomics Aotearoa researchers have studied the genes and proteins in *Microcoleus* proliferating in our rivers.

While globally common and problematic, *Microcoleus* forms a type of bloom that is less well-studied than other cyanobacteria – it coats the benthic environment (along

river or lake beds) in thick cohesive mats. *Microcoleus* proliferations are composed of either non-toxic species, or mixtures of co-occurring non-toxic and the toxic species that are responsible for animal deaths.

Research by the group shows that *Microcoleus* are equipped with diverse mechanisms for acquiring nitrogen and phosphorus, enabling them to proliferate and out-compete others in low-phosphorus waters, while taking advantage of nitrogen compounds likely introduced by agricultural runoff.

Further genomic research by the group has determined important differences in the abilities of toxin-producing and non-toxic *Microcoleus* species to store and acquire nutrients. These differences are associated with substantial differences in metabolic capabilities. Genetic evidence indicates that the toxic species are metabolically less versatile, less able to withstand fluctuating nutrient supply, and unable to synthesize an essential vitamin for self-sustained growth.

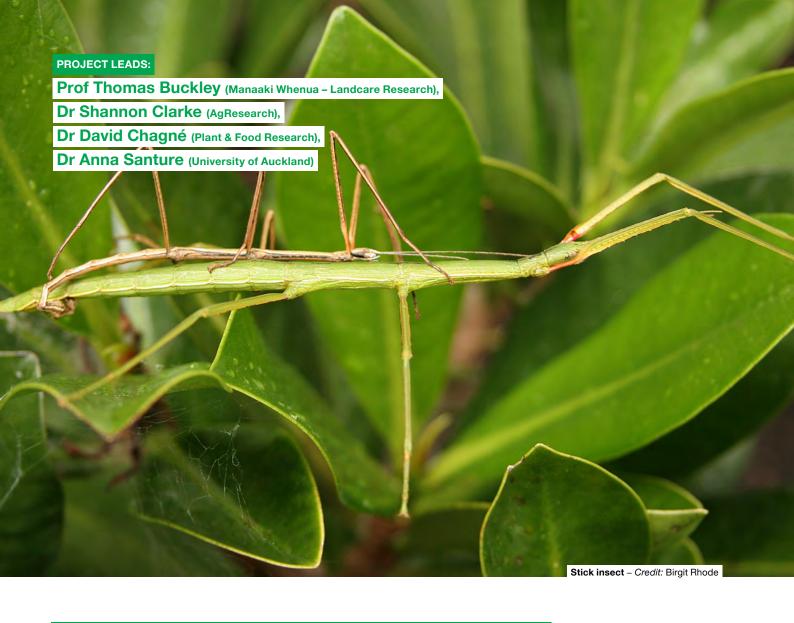
Future work will investigate the project's prediction that toxic species are dependent on non-toxic counterparts for this vitamin, which would explain why toxic

species are invariably found to co-occur with non-toxic *Microcoleus*.

Understanding the conditions needed for these bacteria to grow and thrive, and those that select for toxic species, will potentially help us predict and manage bloom formations, and better manage water quality and safety.

Led by Dr Kim Handley (University of Auckland)





High Quality Genomes and Population Genomics for conservation: controlling invasive rats and wasps and protecting taonga and evolutionary icons

High quality genomes provide conservation ecologists and practitioners with tools to predict key characteristics of threatened species (and their threats) to aid their survival and recovery and characterise resilience to environmental instability such as climate change. Naturally, partnership with Māori is critical to the success of research on taonga and other native species.

The High Quality Genomes and Population Genomics project straddles both the Primary Production and Environment themes. More on the tools and technical developments can be found in the Primary Production theme.

A huge genome in a small package: the challenge of the *Clitarchus hookeri* stick insect

Stick insects are biological enigmas. In challenging conditions, the insects become parthenogenic, meaning females can produce offspring without the contribution of a male. In Aotearoa, they have evolved to thrive in a range of temperatures and altitudes. Adaptive radiations like this can shed broader insights such as the role of adaptation to temperature variation under climate change. But the large and highly repetitive *Clitarchus hookeri* stick insect genome was incredibly challenging to assemble.

Postdoctoral researcher Ann McCartney (Manaaki Whenua – Landcare Research) worked with colleagues at NeSI, who provided the huge computing platform needed for such complexity, to assess the utility of pseudo-long or Chromium10X linked-read technology for assembling the *Clitarchus hookeri* genome.

The success of the approach provided a new pipeline for other genomes, including the threatened hihi. The analysis of hihi from Te Hauturu-o-Toi and Tiritiri Matangi islands revealed high levels of inbreeding among some individuals which is critical information for the conservation management options for the species.¹



Working relationships with Māori iwi and organisations to build trust and forge partnerships

Effective partnership with Māori is critical to the success of conservation of native species and these interactions have been rewarding for project team members. Several relationships have been advanced since the project's engagement and partnership strategy was initiated in December 2020.

Kuaka/Whenua Hou Diving Petrel (*Pelecanoides* whenuahouensis) are critically endangered, with only around 150 individuals remaining. A three-way relationship has been formed among Department of



¹ Duntsch L, Whibley A, Brekke P, Ewen JG and Santure AW. Genomic data of different resolutions reveal consistent inbreeding estimates but contrasting homozygosity landscapes for the threatened Aotearoa New Zealand hihi. Molecular Ecology. 2021. https://doi.org/10.1111/mec.16068

Conservation (DoC), Manaaki Whenua - Landcare Research, and Ngāi Tahu (Whenua Hou Komiti, Kaitiaki Rōpū) to facilitate consent from Ngāi Tahu for genome sequencing and population genomics of kuaka. There are concerns about hybridisation and *genetic swamping* of Kuaka from related petrel species. Our high quality genome and resequencing data will assess this and examine risk to Kuaka. The results will inform management of this species by Ngāi Tahu and DoC.

Postdoctoral researcher Alana Alexander (Te Hikutu (Ngāpuhi)) has won a Te Poutama Māori Advancing Research Collaborations (TPMARC) grant from the University of Otago to enable kanohi ki te kanohi interactions with iwi and hapū for the Hector's and Māui dolphin genome. As grandchildren of Tangaroa (or Tane in the pūrakau of some iwi), whales and dolphins, are whanaunga (relatives), for whom mana whenua have manaakitanga and kaitiakitanga responsibilities. This grant allows Alana to build relationships needed to ensure the genome data supports local hapori Māori aspirations.

A dialogue has begun with Te Atiawa ki te $\bar{\text{U}}$ poko o te Ika $\bar{\text{a}}$ Maui Potiki (Fisheries) Trust about genome sequencing of tarakihi, rāwaru (blue cod), and other important fisheries species. The goal is to enable the trust to contribute to decision-making, for their interests to guide our research objectives and data collection, and to brief members about our findings before we share them more widely with others. We have been asked to consider a new research objective on kākahi (freshwater mussel), which is of conservation concern, and we are in further discussions about how we can support their interests for this taonga species.

Diverse species, diverse research questions

As well as the initiatives already discussed, research projects are underway on a variety of other native species, all with important science questions.

The Myrtaceae plant family, which includes pōhutukawa, rātā, mānuka, kānuka, swamp maire, and ramarama, are all threatened by myrtle rust disease. Genomic research could identify pathways for disease resistance in these species. We are undertaking genome sequencing of swamp maire, a threatened species from the rohe of Rangitane to better understand the impact of myrtle rust on native species.

Te Awahohonu Forest Trust has consented to sequencing of the huhu beetle genome. The huhu beetle has extraordinary wood digesting abilities due is to its complex microbiome and is a potentially important protein source. We will use this species as a model to evaluate tools for genome assembly when samples are contaminated by diverse microbiomes.

As well as being important to iwi, the Hector's and Maui dolphin genome work will enable an assessment of the historic population sizes of these two subspecies, and how population size has changed in response to changes in the environment, as well as determining historic levels of interbreeding between these two subspecies. The results will inform management by helping to quantify the potential of both subspecies to adapt to future environmental change.





Domestic and international collaborations inform and showcase our research

Other domestic and international collaborations have been important to increase the profile of the project's successes and share international knowledge. Key collaborations include:

- The discovery of at least 12 new invertebrate species thriving in the high elevation, climate-threatened glaciers of the Southern Alps, through collaboration with researchers from New Jersey and Poland
- Oxford Nanopores org.one initiative, which provides no-cost MinION flow cell sequencing for Hector's/ Māui dolphins.
- Revive & Restore Wild Genomes, the Vertebrate Genomes Project, and B10K consortium support of Takahē genomics.

Genomics for Predator Free 2050

Ship rats (*Rattus rattus* - also known as black rats) have caused species extinctions and dramatic biodiversity loss on every island group around the world, including Aotearoa. The rat also spreads human and animal disease and reduces agricultural production.

For several years Genomics Aotearoa has been collaborating with Predator Free 2050 Ltd, the Biological Heritage National Science Challenge, and the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) to develop genomic resources

for rats. Last year the High Quality Genomes project delivered a new reference genome for the ship rat to empower scientists working on critical questions of conservation, evolution, and disease prevention. The partnership aims to further enhance genomic knowledge with a population genomics overview of the ship rat.

Kiore (*Rattus exulans*) in Aotearoa is an important taonga species for some Māori. If genomic-based technologies are to be used to control or eradicate ship rats (and *Rattus norvegicus* [Norway rat, the common lab rat]) potential impacts on kiore need to be understood and minimised.

The project team has now begun consultation with Māori on the kiore genome proposal. If granted permission, researchers will then develop a high-quality annotated genome for a population genomics overview of the kiore population in Aotearoa. This next step will provide essential data to guide control approaches and support new technology development.





In Aotearoa they are one of our biggest predator problems – we have some of the highest densities of German and common wasps in the world thanks to an abundance of food, no natural enemies, and a favourable climate. They are a massive threat to our biodiversity, consuming honeydew in our beech forests, eating native insects, and competing for food with our birds. They are also affecting our honey industry.

The wasp genomes study was highly collaborative, involving international researchers from the University of Otago, Te Herenga Waka-Victoria University of Wellington, and NeSI, University of Leeds, UK, University of Sydney, Australia, University of California Riverside, and Phase Genomics Inc. Seattle, USA.

A triple hit - genomes from three invasive wasps

In a world first, Genomics Aotearoa researchers, together with a group from the United States have produced the genomes of three wasps - the common wasp (Vespula vulgaris), German wasp (Vespula germanica), and western yellowjacket (Vespula pensylvanica).²

Sequencing and interpreting the genomes is a major milestone in understanding the biology of *Vespula* wasps.

The researchers used short-read sequencing and Hi-C scaffolding with tissues from male wasps, which allowed them to produce high quality assemblies quickly

and cheaply. Male wasps are *haploid* – they have a single copy of the genome, which makes analysing their genome more straightforward.

Genome data suggest novel solutions for invasive species control

Researchers will now be able to use this genomic information to research and develop next-generation control strategies as well as informing future chemical and genetic control of the pests. It will also be essential for monitoring control methods.

The researchers have already identified genes related to specific traits that may be suitable for targeted pest control. The information could lead to novel control methods, such as RNA interference (RNA-i) and gene drives. The data are already being used to research a potential gene drive targeting spermatogenesis to control the common wasp (*Vespula vulgaris*) as part of the country's Biological Heritage National Science Challenge.

The chance to use leading edge technology for much more targeted and effective wasp control than has ever been possible, paves the way for new control methods in New Zealand's fight against invasive predators. This potentially could make all the difference to a major environmental issue.

Led by Prof Peter Dearden (University of Otago)

² Harrop TWR, Guhlin J, McLaughlin GM, Permina E, Stockwell P, Gilligan J, Le Lec MF, Gruber MAM, Quinn O, Lovegrove M, Duncan EJ, Remnant EJ, Van Eeckhoven J, Graham B, Knapp RA, Langford KW, Kronenberg Z, Press MO, Eacker SM, Wilson-Rankin EE, Purcell J, Lester PJ, Dearden PK. High-Quality Assemblies for Three Invasive Social Wasps from the *Vespula* Genus, Genetics Society of America journal G3: Genes, Genomes and Genetics. 2021.



<u>Case study:</u> **Genomic adaptations to a precarious environment - hot-spring cosmopolitanism**

In Aotearoa, novel species of the bacterial genus *Acidithiobacillus* are exceptionally prevalent in hot springs, including at temperatures far above those previously known for the genus.

Not only does that provide a fascinating insight into the hidden lives of hot pools, it also very importantly demonstrates some of the profound genome changes required for life at high temperature. These extreme genomic adaptations are a new and significant discovery in cosmopolitan hot spring bacteria, found in both ambient and high temperature settings.

Genomics Aotearoa metagenomics studies by university of Auckland researcher Kim Handley and her team have shown that novel groups of the bacterial genus *Acidithiobacillus* are exceptionally widespread and abundant in hot spring settings and are present at temperatures far above those previously known for this bacterium (up to 40 °C higher).

These Acidithiobacillus have reduced genome sizes, and a smaller proportion of duplicated genes (paralogs) and non-coding RNA. These features indicate genome streamlining to reduce cell complexity, which is important for survival at high temperatures, and so are predicted to have higher optimal growth rates and temperatures.

The genomes have a significantly higher proportion of genes encoding for proline – an amino acid that increases thermostability – than *Acidithiobacillus* from other environments. Moreover, DNA composition was contrary to expectations for streamlined genomes. Instead, the novel *Acidithiobacillus* have significantly higher proportions of specific DNA bases that may also further improve thermotolerance. The team also found

similar features in a related genome derived from a Taiwanese hot spring.

Such extreme genomic adaptations by these previously unrecognised groups have helped them to become widespread in diverse geothermal environments. These adaptations were characteristic of all members of the genus derived from hot springs, indicating substantial divergence from *Acidithiobacillus* species inhabiting other environments.

The study was co-supported by Genomics Aotearoa and an Australian Research Council grant awarded to the team's collaborator at the University of New South Wales, Sydney.

The project's results have revealed previously unrecognised extreme genomic adaptations in this group of bacteria, that facilitate their wide dissemination in geothermal environments, and provide insights into the genome changes required for life at high temperature.

Led by Dr Kim Handley (University of Auckland)

Genomics Aotearoa's Primary Production theme has a broad focus on providing genomic tools and technologies for improved outcomes in agricultural applications.

Our diverse set of initiatives optimise genomic technologies for many purposes, including but not limited to animal welfare, improving yields in honey production, and understanding the evolutionary relationships among important fruits.

Broadly the techniques developed can be used for a wide range of applications in horticulture, agriculture, aquaculture, and fisheries. Even more broadly, the methods developed from primary industry genomic data also have the potential to be refined for conservation genomics.





High quality genomes provide researchers in the primary production sector with platforms to better predict biological traits such as the reproduction, appearance, behaviour, or disease susceptibility of a variant, and better understand the genetic diversity of different populations. Understanding this variation can be used to improve economic yields and increase resilience to environmental change.

Genomics Aotearoa's largest project, High Quality Genomes and Population Genetics (HQG+PG) builds upon our earlier High Quality Genome (HQG1) and Better Breeding Values (BBV) projects and includes eight partner institutions. The HQG1 and BBV projects implemented new platforms for long read sequencing such as Oxford Nanopore technology, chromatin-

conformation interaction-based scaffolding (Hi-C), and linked-reads technologies (e.g., 10X Chromium).

While we demonstrated the utility of these methods, sequencing technologies, and bioinformatics methods are continually evolving and must be evaluated and tailored to Aotearoa's needs in both sectors.

The quality of genome assemblies and the accurate characterisation of variation among individuals is critical to support downstream applications. We selected a set of model species of biological, ecological, agronomical, and cultural importance. A special focus was placed on species with large genomes, species that are highly heterozygous, auto and allo-polyploid species, taonga species, species contaminated with commensal organisms, and species with a range of effective population sizes. These characteristics bring several challenges. The project is evaluating tools and developing methodologies to resolving the technical issues that prevented the adoption of genomics solutions.

Engagement with end-users is key to the success of the project. The project team works with industry (Abacus Bio Limited, AgResearch Limited, and Beef+Lamb New Zealand) to provide genetic evaluation and breeding scheme design expertise across a range of economically important species. Native species in the project are under active management by the Department of Conservation (DoC), iwi, and restoration groups. Our research team are actively engaged with these groups, in many cases through formal partnerships with involvement in fundamental research.

For native species, and taonga species in particular,

Māori participation is critical. Engagement with the principles of Te Nohanga Kaitiaki is a major feature of the project, as is the contribution of Māori leadership, researchers, and organisations. One of the model species (rewarewa) was selected in response to conversation with partners in the Māori honey industry.

Mentoring and capability development of future leaders in the genomic field in Aotearoa is another strong focus. Workshops, employment of post-doctoral researchers, and funding of emerging researchers in partner organisations all contribute to this goal.

New technology, insights, and collaborations

In the last year, the project has added three high quality genomes to its catalogue for primary production: bilberry (a relative of blueberry), rewarewa, and *Gillenia* (an evolutionarily ancestral member of the Rose plant family). The rewarewa genome data were used to benchmark a suite of bioinformatic tools for long read assembly and analysis¹. This pipeline is being applied to more species and is available to other researchers. The rewarewa genome is the first large native tree to have a high-quality genome assembled to contribute to improved productivity in the honey industry.



Gillenia trifoliata (Bowman's root) is significant within the Rosaceae plant family as the sister species to the Malae plant group of around 1000 species that include apple and pear. These fruits have duplicated (polyploid) genomes that Gillenia does not. Gillenia also has many ancestral characteristics which makes it ideal for comparative genomics across the entire plant family.

As well as a high-quality genome assembly and annotation, and a benchmark for our assembly pipelines, our team's research characterised the phenotypic development of the plant across life stages. This research has helped verify the genome evolution model for the entire Rosaceae family and provides a reference to unravel the mechanisms behind the evolution of fruits such as apple, pear, peach, strawberry, and raspberry.²



¹ McCartney AM, Hilario E, Choi S-S, Guhlin J, Prebble JM, Houliston G, Buckley TR, Chagné D. An exploration of assembly strategies and quality metrics on the accuracy of the rewarewa (*Knightia excelsa*) genome. Molecular Ecology Resources. 2021. DOI: 10.1111/1755-0998.13406

² Ireland H, Wu C, Deng C, Hilario E, Saei A, Erasmuson S, Crowhurst R, David K, Schaffer R, Chagné D. The Gillenia trifoliata genome reveals dynamics correlated with growth and reproduction in Rosaceae. Horticulture Research. In prep.

Genomics Aotearoa provides genomic support for the FutureBees project

FutureBees, funded by MBIE, and led by Prof Peter Dearden aims to use genomic technologies to help beekeepers select the best bees for their business. FutureBees links remote telemetry data about beehive performance to genomic data to understand the genetic basis of key production traits for bee colonies.

Genomics Aotearoa has increased the efficiency of the FutureBees

programme by providing cost-effective ways to sequence genomes of large numbers of individual bees and deal with the thousands of genomes generated. Technology developed in the High Quality Genomes project, as well as Better Breeding Values, have been deployed to support the interpretation of honeybee population genomic data.

FutureBees is also involved in the new Genomic Graph technologies project, as part of its efforts to develop a 'New Zealand bee specific genome.' This *variome* captures the variation in New Zealand bees, and will reduce the cost of sequencing, ensuring that beekeepers can afford the use of technologies developed by FutureBees.

Our new tools and technologies, our training initiatives, and strong links to NeSI's computing power, highlights the value of Genomics Aotearoa's contribution to accelerate the development of smart new tools that will support New Zealand's \$335 million per annum honey exports.

New pipelines and knowledge sharing

A major goal for Genomics Aotearoa is to establish new tools and technologies from high quality genome sequences that can then be used in future projects. As well as the pipeline developed for stick insects using Chromium 10X linked-reads, visualising Hi-C data posed a problem for the team's researchers. The project developed a new R Shiny app³ for visualising and analysing Hi-C data. The app was developed and tested using data generated in this project and is being optimised for publication and sharing globally. Three genomes have been assembled with this tool so far, including *Gillenia* and bilberry.

In addition to these highlights, the project has been prolific in its first year for knowledge-sharing outputs, generating ten peer-reviewed publications, contributing nine invited talks, seminars, and conference papers, two popular articles and a national workshop on the Hi-C techniques and hands-on bioinformatics session. These workshops play a key role in lifting the genomics capability of the science community in Aotearoa.



Domestic and international collaborations inform and showcase our research

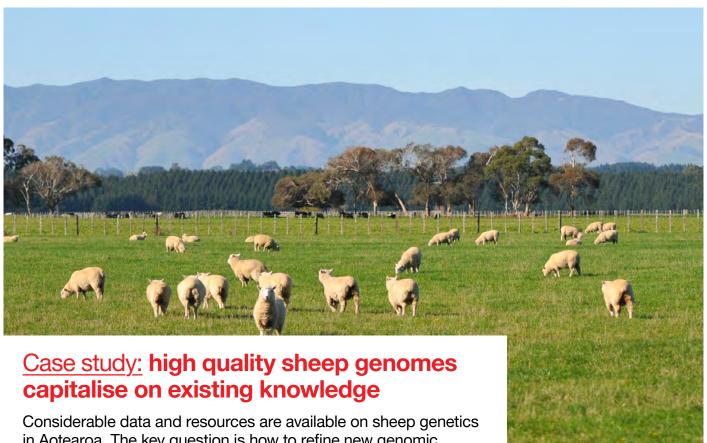
In the primary production sector, the project team have been working closely with Te Rarawa Anga Mua and the Komiti Kaitiaki for Warawara Ngahere for rewarewa sequencing for benefit of Māori honey industry partners. A new dialogue has been initiated with the fisheries committee of Taranaki Whānui ki te Upoko o te Ika and of Te Āti Awa for genome sequencing and supporting fisheries research.

Other domestic and international collaborations have been important to increase the profile of the project's successes in the primary production sector and share international knowledge. Key collaborations include:

- Australian National University for long read Oxford Nanopore sequencing at a low cost for various projects, including blue cod and kiwifruit.
- Neogen GeneSeek for access to new PacBio long read high fidelity (hifi) technology. Early tests of this technology in swamp maire have dramatically reduced time to assemble quality genomes. Neogen also offers highly competitive prices for barcoding DNA for re-sequencing large cohorts of individuals that we need for population genomics.
- Discussions with BioCommons Australia to identify areas of common interests where collaborations could be developed.
- Deep Water Group (DWG) contracted partner Plant & Food Research for genomic analysis of hoki stocks.
 The solutions developed by Genomics Aotearoa enabled the project to be planned and delivered within a year.

Genomics Aotearoa's largest project has demonstrated considerable achievements in its first year, building on the successes of previous learnings. The momentum will continue into the future, with plans for developing tools to characterise metagenomes, new statistical methods to analyse structural variation and SNPs, and generation of at least five new high-quality genomes.

^{3 &}lt;a href="https://github.com/GenomicsAotearoa/Hi-C-workshop/tree/main/Practical/HiCSuite">https://github.com/GenomicsAotearoa/Hi-C-workshop/tree/main/Practical/HiCSuite. R (https://www.r-project.org/) is widely used statistical computing software platform. Shiny is a tool within R to develop interactive web apps. Typically, tools developed within R are freely available for use and further development.



Considerable data and resources are available on sheep genetics in Aotearoa. The key question is how to refine new genomic technologies and incorporate them into the existing frameworks to create seamless processes for the sheep industry.

But optimising new genomic technologies to predict economic, environmental or welfare related genetic traits in sheep breeding is no trivial task – genetic tests need to be cost effective and results understandable and meaningful for breeders.

A collaborative approach to seek answers

Working with industry leaders, Genomics Aotearoa has developed new statistical approaches and new data sources that can be interpreted meaningfully.

This project brought together researchers from the horticulture, sheep, beef, and dairy industries via two Crown Research Institutes, two universities, and industry partners Livestock Improvement Corporation (LIC), and AbacusBio Limited, combining crucial skills and capabilities to complement other Genomics Aotearoa projects. Two postdoctoral researchers, managed across institutions, were employed to build capability.

New tools and approaches to predict breeding values

The project has produced tools for breeding value prediction, adapting techniques such as imputation (a process that uses statistical inference) to predict the worth of elite individual animals earlier and more accurately.

Novel approaches augment the wealth of genotypic information already available on animals with Whole Genome Sequence data. This combination adds considerable detail to understanding the role of genetic variations in the animal's DNA, and therefore its potential.

The approaches developed are being made

accessible across all primary industries, so that they can cost-effectively adopt genomic selection to suit their own circumstances. The techniques are also relevant to conservation, where management of threatened species requires similar genomic support.

The project is the culmination of a strong working relationship between Genomics Aotearoa partners, science, industry, and farmers in Aotearoa. Research organisations will now be investigating how to implement findings to improve farming of New Zealand cattle and sheep.

Moving to the future while making the most of the past

Incorporating the rapid growth in next generation genomic sequencing technologies into existing knowledge and systems provides the opportunity to better identify and use information from genetic variants in livestock improvement.

The new generation tools will further drive genetic improvements across meat production traits, while maintaining sheep health and disease resistance. Genomic breeding values will increase the rate of genetic gain, more accurately predicting the worth of an individual animal, helping breeders get the most from their investment.

The result will be future proofing, ensuring our biological industries benefit from the increased accuracy and predictability, and are competitive internationally.

Led by Dr Shannon Clarke (AgResearch)



which presents opportunities for developing New Zealand's

A BioCultural Notice recognises the rights of Indigenous peoples

blueberry production.

What is also special about this project is that for the first time a BioCultural Notice has been added to a gene sequence in Aotearoa.

The BioCultural Notice is a digital identifier that recognises the rights of Indigenous peoples to define the use of information, collections, data, and digital sequence information generated from the biodiversity and genetic resources associated with their traditional lands, waters, and territories. More information about BioCultural Notices and their importance can be found at https://localcontexts.org/.

Genomics Aotearoa Researcher David Chagné is part of the Te Tuakiri o te Taonga Catalyst Project funded by the Royal Society of New Zealand which developed the BioCultural Notice and Labels for Aotearoa.

Close relationships between bilberry and cultivated blueberries and cranberries enable understanding of key traits

Bilberry (Vaccinium myrtillus) belongs to the same genus that includes blueberries and cranberries. Prized for its taste and health properties, it has provided essential nutrition for Northern European Indigenous populations for centuries. High concentrations of phytonutrients, including the important purple anthocyanins, are found in both skin and flesh.

Unlike its cultivated relatives, bilberry remains largely undomesticated, with berry harvesting almost entirely

from the wild. With a lack of human-assisted genetic enhancement it is ideal for genomic analysis, providing comparisons with the domesticated Vaccinium species.

The bilberry genome assembly, sequenced by an international team of researchers, builds on the existing genomic resources and knowledge for other Vaccinium to help understand the genetics underpinning some of the quality attributes that breeding programs aspire to improve.

Bilberry genetic diversity has not been well understood and its placement within the Vaccinium genus is not well established.

The bilberry genome will be used as a reference to study genomic diversity in bilberries and provide information on the loci linked to adaptive traits and the phytochemical composition of the berries.

One key difference between blueberries and bilberries is the localisation of anthocyanins, the compounds responsible for the berry's colour that have antioxidant properties. In blueberries these are restricted to the skin but in bilberries concentrations also accumulate in the flesh. This study analysed the locus for the anthocyanin-regulating transcription factors (MYBA) and identified a complex locus controlling berry anthocyanin composition.

The close relationship between bilberry and blueberry genomes enables comparison between them to discover the genes responsible for key traits like anthocyanins, which evolved from a common ancestral gene.

Led by Dr David Chagné (Plant & Food Research)



The completion of the native rewarewa (Knightia excelsa) tree high quality genome demonstrates how Aotearoa is leading genomics for its native species.

Native species like rewarewa are important to our environment and have taonga significance. They also provide a rich resource for honey production

Rewarewa, only found in Aotearoa, occurs mostly in the North Island and the top of the South Island, and is common in coastal. lowland, and lower montane habitats.

Bees are also attracted to rewarewa and produce a delicious honey - described as "dark, malty, and complex with a sweet finish."

The Genomics Aotearoa collaboration between Manaaki Whenua - Landcare Research, Plant & Food Research and University of Otago has developed an impressive set of techniques benchmarking genomic solutions which are useful for the Aotearoa genomics research community and internationally.

A genomic resource for Māori honey producers

The rewarewa genome is now being used to better understand the tree's biology. The goal is to enable the burgeoning rewarewa honey industry to commercialise their product based on its whakapapa (provenance), as has been done for mānuka.

Establishing a high-quality genome for this taonga species has specific cultural importance to Māori. This tree has several traditional uses including medicinal, spiritual, and in building construction.

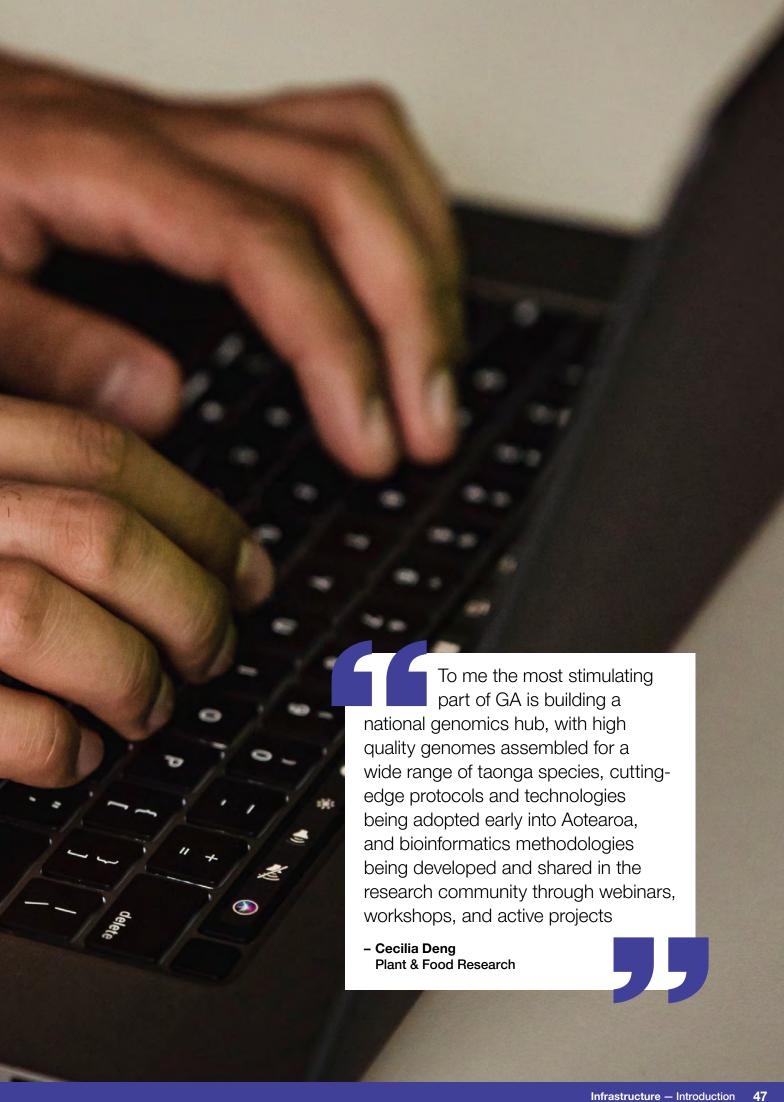
The bioinformatics processes developed through this research are also adaptable to conservation projects. The research may contribute a deeper understanding of which genes contribute to adaptation of this species to a range of climatic conditions, a question that is of critical importance in times of rapid climate change.

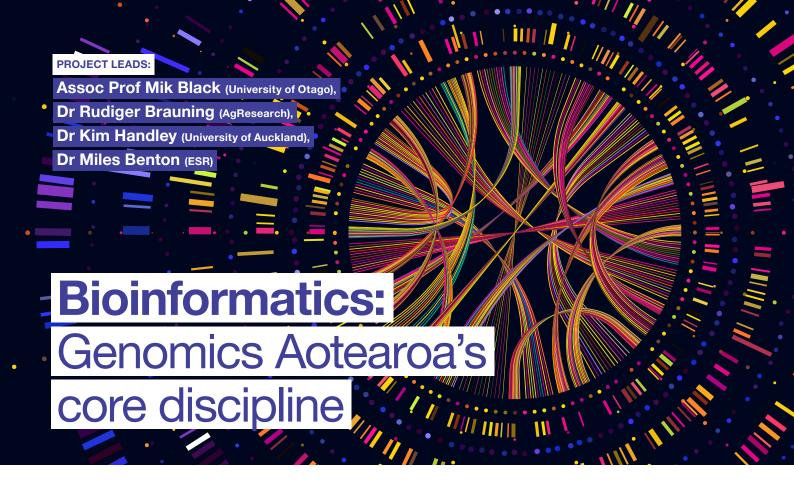
Led by Dr David Chagné (Plant & Food Research)

Gold-standard computing infrastructure and support is essential to enable the growth of genomics research in Aotearoa, for Aotearoa.

Bioinformatics capability for researchers, secure state-of-the-art computing facilities and data curation, mentoring future bioinformatics leaders and culturally grounded approaches to data curation and access are all part of Genomics Aotearoa's infrastructure repertoire.

All of Genomics Aotearoa's projects are underpinned by the on-going development of the national infrastructure, which together enable and grow the skills and capability of our genomics researchers. Our Infrastructure links projects together through common tools and skill sets, helps to define a structure for capability building (e.g., training) and community development/engagement, and enables a level of coordination and resourcing at a national level.





Bioinformatics combines biology, computer science, mathematics and statistics and is a discipline that lies at the heart of Genomics Aotearoa's work. While researchers arrive at bioinformatics through a variety of fields, they all share the same problem: how to analyse large, complex sets of data to extract biological meaning.

Bioinformatics provides the tools and strategies needed to analyse genomic data: the greater the capability in bioinformatics, the more we can do with the information we gather. Understanding the bioinformatics requirements of the research community drives our other infrastructure needs: computing resources, data management, and training.

National leadership in bioinformatics is needed to support and develop on-going capacity and capability in Aotearoa. Linkages among life science researchers, applied and academic bioinformaticians, and industry practitioners increase capability. Bioinformatics postdoctoral opportunities are critical to nurture novel ideas and technologies and provide international-level training and career pathways.

Genomic Aotearoa's approach is one of learning by doing. The aim is to increase genomics capability by enabling researchers to analyse their own data through training opportunities, infrastructure provision, and building new tools and technologies into projects. Leading bioinformaticians are linked to emerging postdoctoral bioinformaticians working on Genomics Aotearoa projects, providing them with the training and tools to increase capacity. Training in bioinformatics techniques is offered to all researchers in these

projects, using an instructor training approach to further spread capability. We believe that the best way to improve the uptake of genomic and bioinformatics technology is to lead by example – using excellent research to define and deliver the tools and training that are most relevant for our researchers.

Building on strong foundations

In Genomics Aotearoa's initial years, the foundations were laid for on-going support and capability building for bioinformatics. These foundations include a strong leadership team comprised of experts from several institutions, the establishment of working groups to define and progress roadmaps for bioinformatics, computing and data management, recruitment of a training co-ordinator, and a digital content strategy.

Although the pandemic has prevented international travel, the past year has seen further engagement with national research communities through the eResearch NZ conference in 2020 and 2021 and many training and workshop activities. The infrastructure to support bioinformatics has progressively been refined according to the roadmaps.

Genomics Aotearoa is on track to achieve its goals of providing national leadership in bioinformatics and mentoring, networking, and support for Genomics Aotearoa bioinformaticians.



New Zealand eScience Infrastructure (NeSI), with its specialised computing infrastructure, support-staff, and computational science expertise, provides critical underpinning support for bioinformatics research platforms in

Aotearoa. NeSI has a strategic commitment to support research that aligns with Genomics Aotearoa's commitment to increase the capability of our genomics researchers. This strategic alignment makes us ideal partners.

To support the partnership, NeSI has recruited bioinformaticians to work alongside genomics researchers to better understand their current and future needs. They assist with bioinformaticians support bespoke software optimisation and deployment, third-party application installation, onboarding researchers new to NeSI's platform, among other user support activities.

As well as this highly specialised bioinformatics support, NeSI provides cluster computing, virtual environments for training, a dedicated 6TB RAM node for large-scale genome assembly, and dedicated data allocations for the Aotearoa Genomic Data Repository. Just one example of the success of the partnership has been the sequencing and assembly of the complex stick insect genome.

NeSI resources are now routinely used for the computational components of Genomics Aotearoa research projects, as well as in the delivery of workshop-based bioinformatics training. We continue to assist NeSI with their development of a virtual computing environment tailored to genomics work, and Genomics Aotearoa has recently assisted with benchmarking genomics pipelines for new NeSI GPU systems.

Together, NeSI and Genomics Aotearoa have the domain knowledge, technical skills, and infrastructure to tackle the skills gap facing Aotearoa's genomics and bioinformatics sector. "Supporting genomics research today and tomorrow requires a specialised and powerful technology ecosystem, and an inclusive partnership with Aotearoa researchers, communities, and Te Ao Māori," says Nick Jones, Director of NeSI.

⁴ https://www.nesi.org.nz/case-studies/how-genomics-aotearoa-and-nesi-tackled-big-problem-and-won https://www.genomics-aotearoa.org.nz/projects/high-quality-genomes/how-stick-insect-helping-bioinformatics-practice



Genomics Aotearoa is continuing to develop a national genomics data archive, with appropriate guardianship, to support our research. Genomics Aotearoa manages data storage and access within a Māori values context.

Genomics Aotearoa has contracted NeSI to provide technical expertise and compute/storage resources for the establishment of a national genomic data repository. Initiated in 2019, the Aotearoa Genomics Data Repository (AGDR) has been progressively enhanced over the last two years.

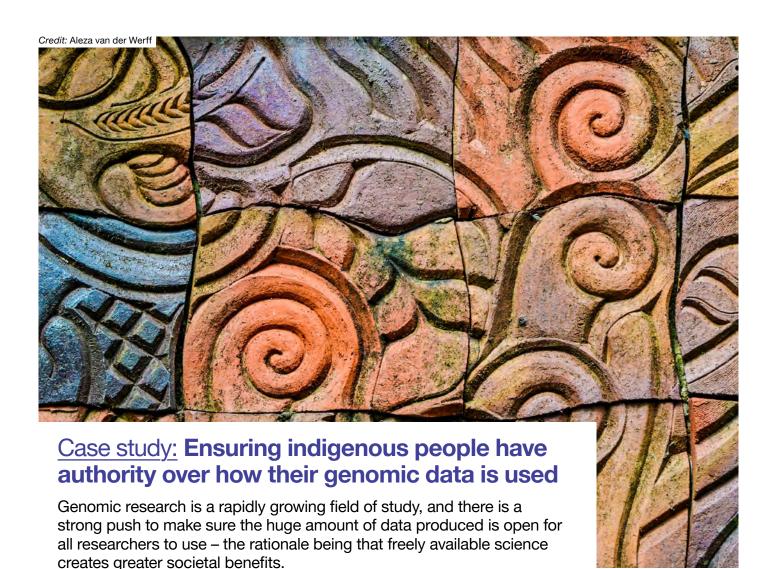
The focus of the data repository has been on providing storage for genomic data from taonga plants and animals. Local hosting is a critical requirement because if data are moved outside Aotearoa's legal jurisdiction (e.g., on "cloud" servers) any Tiriti o Waitangi rights are extremely difficult to enforce.

Tikanga for both data submission and retrieval were developed by former Genomics Aotearoa Vision Mātauranga Coordinator, Ben Te Aika. Data set submitters must provide information not only about the species and samples, but also on the Māori consultation process that was followed and the kaitiaki of the data being submitted. This information is also

used for data access: requests must be approved by the kaitiaki of that data set.

Technical support for the AGDR is provided by NeSI's local implementation of the Gen3 platform housed at NIWA's Greta Point facility. Over the past year the NeSI team has improved the user experience for both data submission and repository usage. This work has included consultation with relevant research communities on the minimum metadata requirements for submitted data sets, as well as the development of extensive documentation.

Although still in its infancy, the repository is fully functional and accepting submissions. The AGDR already contains genomic data from nine species, with a large amount of additional data scheduled for storage over the coming year. Both the interface and the data submission processes are being further refined. Once these are complete, a wide range of researchers will be encouraged to submit their genomic data to the repository.



But a group led by Genomics Aotearoa researcher Maui Hudson from the University of Waikato has warned of injustice if there is unrestricted access to genomics data. The concern is that the Indigenous voice will be lost if their perspectives are ignored, and the benefits of the research will simply not reach these communities.

Their paper "Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data" was published in the prestigious international journal Nature Reviews Genetics in April 2020. The review was a unique collaboration of prominent research organisations in the United States, Canada, Australia, South Africa, and Spain, as well as the Universities of Waikato, Otago, and Canterbury, Plant & Food Research, and Manawa Ora, Tauranga.

The publication of this work in a prestigious journal reaffirms the importance of ensuring equity is integral to research.

The article identifies enduring negative effects on Indigenous peoples, minority populations, and socially disadvantaged groups across numerous research projects. An open approach is not equitable as these communities carry substantial risk but see few of the benefits of genomic research.

Researchers must cease the practice of disempowering Indigenous communities by appropriating their genomic data and genetic heritage. Developing access and benefit-sharing protocols is a key pathway to move

forward.

Kaupapa for indigenous research

Greater diversity and inclusion of Indigenous peoples in genomic research can be supported by adopting three broad kaupapa:

- 1. Building trust Indigenous communities decide whether their data are publicly available or only accessible on request.
- 2. Enhancing accountability the provenance of Indigenous samples and data must be transparent, disclosed in publications, and kept with the data.
- 3. Improving equity credit is attributed to Indigenous communities to support future use and benefit-sharing agreements as appropriate.

The authors suggest these principles will be given effect only through efforts to develop closer relationships between genomic researchers, Indigenous communities, and Indigenous genomic scientists.

Involving Indigenous communities more deeply in genomic research will ensure the benefits of research are shared more equitably and improve the quality and utility of the research products.

Led by Maui Hudson (University of Waikato)



Genomics Aotearoa has a strategic commitment to develop genomics and bioinformatics capability in Aotearoa researchers. Pressure on the genomics sector to produce and interpret high quality data drives a need for researchers with a wide range of bioinformatics skills.

High school and undergraduate courses are not yet able to cover the entire range of topics required, creating a significant capability gap. To fill this gap, researchers are increasingly seeking out opportunities to attend training courses and workshops to develop the skills they need to succeed in their field. Genomics Aotearoa provides these courses in a flexible environment, either online or face-to-face, in close partnership with NeSI.

Genomics Aotearoa's training offerings are spearheaded by Training Coordinator Ngoni Faya. Since 2019 Ngoni has worked closely with Dinindu Senanayake at NeSI to plan and deliver training sessions in a range of in-person and online formats. Building a strong partnership with NeSI has been crucial to the success and high quality of the training courses. NeSI offers free access to high performance computers for training sessions, allowing researchers to experience the power of these systems, often for the first time. Having the chance to discover what NeSI has to offer means that researchers often choose to use NeSI services for their own research in the future.

Since April 2019, more than 750 researchers from 21 Universities and Research Institutes have attended Genomics Aotearoa courses. Of the 31 courses delivered so far, seven have been run as online webinars. In the last year alone 445 learners participated in 18 events. Regular topics include Genomics Data Carpentry, RNA-seq, eDNA, Imputation and Genotyping-by-Sequencing.



From February 2021, Ngoni and Dinindu have offered a follow-up service designed to help researchers navigate through customised genomic data analysis pipelines as they start applying new knowledge to their own projects. Straightforward follow up queries are handled via email while one-on-one Zoom sessions are scheduled for more complex problems.

An important training component involves Genomics Aotearoa postdoctoral researchers participating in the instructor training activities that are offered as part of our membership of the international Carpentries training programme. This two-day module teaches the pedagogical skills that help attendees pass on their learnings to others when they return to their home institution, thereby spreading the capability-building net even wider.

The first pandemic lockdown provided significant opportunities for expansion into online education. A test webinar was planned for March 2020 to assess the viability of online training for the future. When the country-wide lockdown was announced, this webinar moved from being a test, to being the first of four online courses, providing training opportunities to 120 students in two-day blocks. The success of these online courses became a fantastic proof of concept that encouraged a shift to a permanent mix of online and in-person courses once the country left lockdown.

Courses are free to attend and fully subsidised by Genomics Aotearoa, with support and resources provided through partnership with NeSI. Offering free courses removes a significant barrier for researchers, particularly postgraduate students. A three-day inperson workshop in Australia would cost around \$500 NZD, while a four-day workshop in the UK would be closer to \$2,000 NZD. Online courses are less expensive but there is still the cost of the facilitator. A two-day online course in the UK costs around \$600 NZD. Genomics Aotearoa believes that providing training that is free helps reduce inequitable access to education while offering high quality courses for all researchers.

Offering a centralised focus for genomics training for junior researchers has several benefits. The pressure of a mentoring workload on more advanced researchers is reduced. Centralised training encourages networking

among participants and facilitators, who may usually be located far from each other. New collaborations can eventuate, leading to exciting new research possibilities.

Our website outlines our catalogue of regular courses⁵, which include RNA-seq analysis, Imputation, Genomics Data Carpentry, Spring and Summer Schools. One-off events have included Genome Assembly, Genomics Workflows – Containers and a Hi-C data analysis workshop.

Just a few of the positive responses from participants have included:

The broad overview of data processing pipelines gave me an idea of the capabilities of these packages and how I could use them for the specific use cases within my project

- METAGENOMICS SUMMER SCHOOL (AUCKLAND)

For so long, I have been afraid to start Genomic data analysis and this workshop has really given me the confidence to grow my data analysis skills. The learning environment was friendly and the instructors were accommodating

- GENOMICS DATA CARPENTRY (CHRISTCHURCH)

Beautiful dive into Hi-C data analysis. Very clear introduction to Hi-C method. And of course, face-to-face interactions:)

- HI-C WORKSHOP (AUCKLAND)

Great workshop as always (3rd time attending your workshops). Apart from the knowledgeable instructors and helpers, the workshop material is logical and can be easily used for learning even outside the workshop.

- OTAGO BIOINFORMATICS SPRING SCHOOL (DUNEDIN)

It was a great opportunity to network and meet new faces. I also really appreciated hearing about all the different applications of Hi-C, I'm leaving inspired

- HI-C WORKSHOP (AUCKLAND)

The organiser of the workshop being available for questions and personalised assistance, even after the workshop - RNA-SEQ WORKSHOP (PALMERSTON NORTH)

Courses can also be developed for specialised audiences. One-off courses have included *Postdoc Upskilling* which extended Bioinformatics topics to include science communication and networking, a *Clinical Interpretation of Disease-Causing Variants* workshop on curating and interpreting genomic data from disease causing variants, a *Data Repository Metadata Workshop* for Genomics Aotearoa affiliated researchers to participate in database design for taonga genomic datasets, and *High Throughput Sequencing* for Ministry of Primary Industries staff.

For information on customised courses, future training offerings, attending a regular training event or subscribing to the Genomics Aotearoa training newsletter, contact Dr Ngoni Faya (ngoni.faya@otago.ac.nz).

⁵ https://www.genomics-aotearoa.org.nz/education-events#on-this-page--1



Genomics research is a developing frontier with an increasing focus on the needs of Māori populations, taonga, and other indigenous species. Addressing Māori ethical issues within the context of genomic research is important, so the solutions developed are appropriate for the cultural context of Aotearoa.

Māori are still significantly under-represented in genomic sciences. One way of building capability is to develop understanding of genomics among Māori. New Zealand's annual Summer Internship for Indigenous Genomics (SING) Aotearoa programme, co-funded by Genomics Aotearoa, is designed to develop that indigenous understanding of genomics.

SING Aotearoa brings tauira Māori interns together with Genomics experts in an intense five-day maraebased residential programme. Topics covered include genomics, bioinformatics, and ethical and cultural issues in relation to genomics research. The nature of

the programme provides opportunities for non-Māori researchers to facilitate sessions and engage with participants. The initiative has created internships, research collaborations, and scholarship pathways. The SING Actearoa programme is modelled on the SING programme developed in the USA. The partnership with SING-USA has provided opportunities for faculty and student exchanges. SING is now one of the key capability-building initiatives supported by Genomics Actearoa.

This year the sixth SING programme was held at Rehua Marae in Otautahi Christchurch in January

2021. Fifteen tauira, seven convenors, and a further 12 specialists contributed to korero on the themes of Health, Conservation, and Primary Production, as well as exploring wet lab and bioinformatics techniques. The interns were overwhelmingly positive about their SING experiences:

I learnt a lot more about how genetic research can really advance Māori, but also the huge ethical issues that come with research & data. I also learnt about the legal element to research & ethics, contracts, benefit sharing agreements etc.

The basis of the entire spectrum of issues & content that is involved in this mahi, ranging from whakaaro Māori right to the micro level of scientific ideology and hegemony. The concepts of empathy and

sympathy at every level of a researchers mahi and interactions in this space.

The key issues around Māori & Science, that the meaning of those two communities together is going slowly but making progress. Key issues around consent/governance/Māori involvement etc & how it needs to be taken into consideration for research design.

SING draws on expertise within Aotearoa, providing additional exposure for Genomics Aotearoa projects and reciprocal capacity building for both interns and the presenters. The perspectives of both interns and presenters shared through SING have contributed to Te Nohonga Kaitiaki (see below). The next SING event will be held in early 2022.

A SING Aotearoa experience - Jordon Lima (Ngāti Porou)

Attending her first SING Aotearoa conference was transformational for Jordon Lima.

Jordon was studying biomedical sciences at the University of Otago when lecturer and SING Faculty member Dr Phil Wilcox (Ngāti Rakaipaaka, Rongomaiwahine, Ngāti Kahungunu ki te Wairoa) encouraged her to attend the 2020 SING conference; an inaugural forum that brought together SING consortia from the US, Canada, and Australia.

Jordon also attended the SING Aotearoa workshop in 2021.

"All throughout my studies, I have contemplated how my cultural identity fits into my science. At SING, I was surrounded by fellow Indigenous researchers and students who shared their cultural practises and experiences and had all contemplated the same thing throughout their respective studies," Jordon recalled.

"Indigenous scientists often feel isolated within our institutions around the world, but SING provides a safe environment for us to connect over these shared experiences and to support one another."

"Both SING workshops were amazing. The guest speakers, who are experts in their respective fields, shared their knowledge and experience with attendees on how best to incorporate Te Ao Māori (the Māori worldview) into our research practices."

"Most importantly, close interactions with these experts and our peers validates and builds a sense of confidence that is key for embracing our indigeneity in science. It's been great."

"Thanks to SING Aotearoa, I am now encouraged to further develop my understandings of how Te Ao Māori can be incorporated into my research. I am also more involved in providing mentorship for other young Māori looking to pursue careers in genetics, as I was provided mentorship through SING.

And I have made long-lasting connections through this programme; I hope together we can help to shape the space for cultural identity in science in New Zealand for the future."

Jordon is working on her PhD in the Centre for Translational Cancer Research at the University of Otago, under the direction of Professor Parry Guilford and Associate Professor Karyn Paringatai. The Centre takes a genetic approach to improving the survival and quality of life of cancer patients with new and more effective options for diagnosis and treatment.



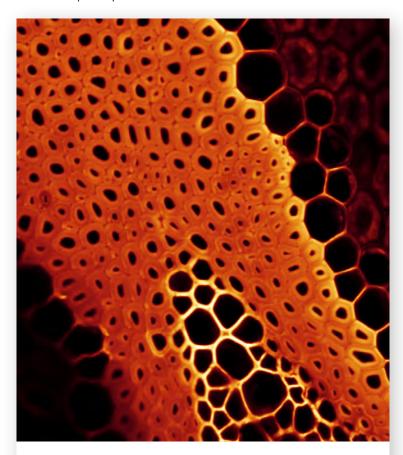


Te Nohonga Kaitiaki (Place of the Guardians) are guidelines to assist the planning and execution of genomic research in a way that honours taonga, kaitiaki, and mātauranga Māori. The guidelines can also help hapū and iwi formulate protocols for research relevant to their own tikanga and mātauranga. Both the role of traditional guardians of taonga species and the responsibilities of institutional stewards are recognised in the guidelines.

Genetic and genomic research on taonga species supports practical conservation efforts, biotechnology and agriculture as well as contributing to theoretical knowledge. Kaitiaki need to be involved in decisions about uses of the information generated from these projects to ensure mana whenua and Māori overall benefit from the value created.

"Despite Te Tiriti o Waitangi affirming Māori kaitiaki interests and rights over taonga, the application of these rights to species and their data has historically been overlooked by Pākehā science," says researcher Dr Maui Hudson. In recent years ease of access to genomic technologies has resulted in a proliferation of research into taonga species, highlighting the need for researchers to appropriately recognise the tikanga and mātauranga of mana whenua.

Māori led, the Te Nohonga Kaitiaki project team engaged with both Māori and researcher communities over the past three years. Genomic research on taonga species was viewed from both a cultural and commercial perspective. The project team engaged and consulted with diverse Māori networks, agencies, and communities. The project responded to feedback by extending the engagement and consultation phase to ensure full participation.



Te Nohonga KaitiakiGuidelines for Genomic
Research on Taonga Species



Collaboration and engagement enhance Te Nohonga Kaitiaki Guidelines

Te Nohonga Kaitiaki was completed on target this year. The feedback on the draft guidelines was positive and the document is undergoing final layout prior to publication. The next step is to launch and disseminate the guidelines through the Māori engagement channels developed through the project and the science research community.

The findings of the project have been shared widely. The draft guidelines were presented at the Māori Genetics, Genomics, and Gene Editing Wānanga and Ngā Pae o te Māramatanga Indigenous Research Conference in 2020, and the SING Aotearoa Workshop in early 2021.

Lead researcher Maui Hudson is a member of the Ethical, Legal and Social Issues Subcommittee of the Earth Biogenome Project and an invited speaker for an Access and Benefit Sharing Initiative. Five peer-reviewed publications, four conference and seminar presentations and several reports and online media commentaries have emerged from the project this year.

While the project may be complete, plans are underway for dissemination and application of the guidelines more broadly. The team is aiming for a Vision Mātauranga workshop on the guidelines at the New Zealand

Molecular Ecology conference in 2022. Opportunities are being explored to leverage the guidelines through the Turbocharging Plant Production project, led by Plant & Food Research. This project investigates gene editing to enhance plant breeding.

A key area identified by the guidelines is data access, aligning with emerging initiatives supporting Indigenous Data Sovereignty, including the use of Traditional Knowledge and Biocultural Labels research space. Funding has been received from MBIE, the Mellon Foundation and Te Puni Kokiri for activities to advance the use of Labels. The team is also contributing to work around a Māori Intellectual Property Strategy for Genomics Aotearoa through a partnership with the National Science Challenge Science for Technological Innovation.

Printed and electronic versions of the guidelines will be made available from 30 September 2021.

Genomics Aotearoa strongly encourages researchers to read the guidelines to gain a better understanding of the importance of the appropriate engagement of mana whenua in genomics research. Please contact Genomics Aotearoa to be advised of the guidelines' release or to reserve a printed copy.



Genomics Aotearoa's public engagement activities include a fortnightly Friday seminar series, interactions with the media, a YouTube video channel and public events held throughout the year.

Friday seminar series

The Friday afternoon zoom seminar series is now a recognised feature of Genomics Aotearoa's engagement with the New Zealand genomics community and beyond.

Established as a way of sharing research with fellow genomics researchers, it has grown substantially from its origins in the first pandemic lockdown 2020, attracting a range of researchers and students in the genomics space in a collegial way that hasn't been done consistently in Aotearoa before. Seminars are now open to all and promoted widely. This has resulted in a significant increase in requests to join in each fortnight, including from some senior science classes in high schools. Between 60 and 80 people tune in each fortnight.

We have hosted 18 seminars over the 2021 period, all from university or science organisation researchers sharing research milestones, but just as importantly, engaging in questions and discussions on the technical aspects of their genomic and bioinformatic decision-making processes. The seminars are

also proving a good opportunity for early career researchers to present.

This year we were thrilled to have two international speakers speaking live via zoom – British geneticist, author, and broadcaster Adam Rutherford, presenting from the UK on *Genetics, Race, and Scientific Bias: A History.* And Sergey Karen from the National Human Genome Research Institute, NIH, USA talked about the latest technologies applied to producing highly contiguous, complete, telomere to telomere genome assemblies. Speaking live from the USA, the presentation featured one of Genomics Aotearoa's inaugural post-doctoral fellows, Ann McCartney, who is now working at the NIH. Both seminars attracted great attendance and generated follow-up interactions between Genomics Aotearoa researchers and the international experts.

Media highlights

Genomics Aotearoa was highlighted in a New Zealand Herald feature article that marked the 30th anniversary of the human genome project, with Prof Peter Dearden putting genomics into a New Zealand context. Peter Dearden also featured in a TV3 news item about the discovery of new micro-organisms in glaciers (this also featured in the NZ Herald), and in a TVNZ documentary – Fight for the Wild – on genomic technologies for conservation. We also publicised our world-first stories – on the wasp genome project and the rewarewa genome sequencing. Genomics Aotearoa continues to provide valuable monthly commentary and easy to understand technical explanations via the Science Media Centre's Sciblogs. And it also contributes regularly to Twitter.

YouTube Videos

We have set up a YouTube channel and are in the process of uploading the 25 seminars, which have been recorded since April 2020. To date, our most popular video has been on Māori participation in genomics research. This is heartening as Vision Mātauranga is at the heart of what we do.

Events

Genomics Aotearoa again co-hosted the eResearch NZ conference in partnership with NeSI and REANNZ. Held at Te Herenga Waka - Victoria University of Wellington and as a hybrid event for the first time, the event attracted 153 people - 92 in person and 62 virtually. Forty of the virtual attendees were overseas. Attendees were very appreciative of the opportunity to network and discuss issues around building data science capability. Dr Ayesha Verrall, Associate Minister of Research Science and Innovation also attended.

Despite the restrictions of COVID, we continue to contribute to genomics conferences and meetings here and internationally. Just some of those contributions include:

Otago Museum science fair: As part of the Women in STEM research project, in which two of our post-doctoral fellows featured, we also helped organise a genetics section at the Otago Museum science fair to profile genomics as a career option to young people.

Resbaz: Genomics Aotearoa continues to have a high profile in the University of Otago's Research Bazaar programme promoting the digital literacy through a series of workshops, demonstrations and talks to researchers at all levels from all disciplines.

Genomics Aotearoa researchers continue to play a prominent role in conferences here and overseas via zoom, including:

- Bioprotection Aotearoa conference
- NZ Apiculture Conference
- Genomics Aotearoa and Bioplatforms Australia
- Genomics Aotearoa and Biocommons Australia
- Genomics Aotearoa HQG2 project
- The Genetics Society of Japan
- The Frozen Ark project governance group
- Genome biodiversity conference
- NZ Institute of Agricultural and Horticultural Science Inc conference
- Federated Farmers conference

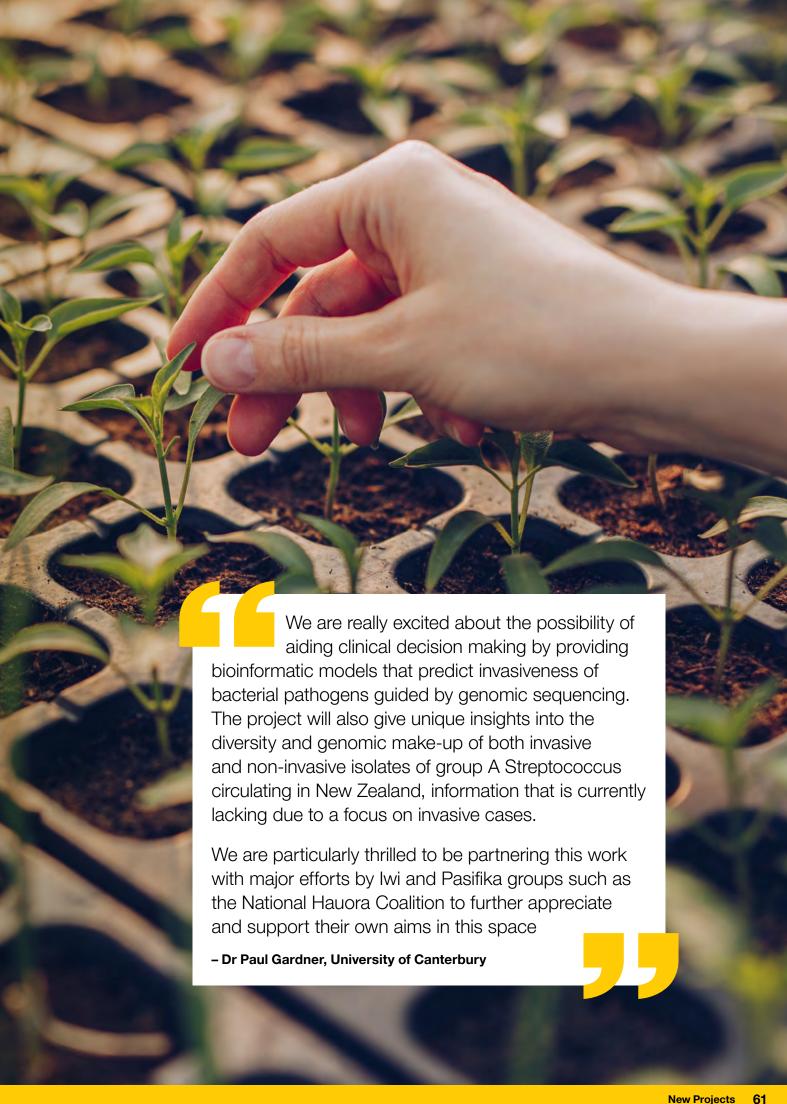
As conference organisers are increasingly making presentations available online, the opportunity to extend our influence further in the online space is expected to increase.



This year has seen the initiation of five new projects: Genome Graphs; Identifying the drivers of Streptococcus pyogenes; Invasomics; Ruatau; Environmental Microbiomes. Establishing five new projects in a single year is a major milestone for Genomics Aotearoa and is indicative of the growing desire for genomics-based research capability nationwide.

Genomics Aotearoa welcomes collaborative project proposals from researchers based in Aotearoa that are consistent with our vision. For more information on how to complete a proposal, contact Genomics Aotearoa.

Genomics Aotearoa uses a thorough vetting process to ensure the highest science quality and strategic value of accepted projects. www.genomics-aotearoa.org.nz/about/how-we-develop-new-projects



Genome Graphs to unravel pangenomes

The move towards pangenome analysis (a collection of many individual genomes) is a step-change in bioinformatics practice. Graph-based pangenome analyses offer an improved approach to detect genomic variation. While graph-based methods are a relatively recent bioinformatic development, they are experiencing strong uptake internationally, though not yet in Aotearoa.

Graph-based pangenome-based methods identify a core genome (the part of a genome or gene set that is shared across many individual genomes). The methods are particularly relevant for the analysis of complex structural variants in populations with mixed ancestry.

To establish pangenome analysis capability within the Aotearoa genomics research community the project will assess and apply Genome Graph methods to improve variant detection in a collection of case studies. Case studies selected are particularly relevant to Aotearoa and include both large (human-scale) and small (microbial-scale) genomes. Graph-based methodology is relevant to a range of on-going

Genomics Aotearoa projects across all themes, including the High Quality Genomes and Population Genomics, Environmental Metagenomics, Variome and Rakeiora projects. These projects provide ideal models for case studies.

The Genome Graphs project is closely aligned with Genomics Aotearoa's goal of increasing bioinformatics capability and will contribute best-practice analysis workflows and training for the country's genomics research community.

Led by Assoc Prof Mik Black (University of Otago), Dr Joep de Ligt (ESR)

Identifying the drivers of *Streptococcus pyogenes* to better understand disease causation

Group A Streptococcus (GAS, or *Streptococcus pyogenes*) is a global health problem, responsible for more than 600 million infections per year worldwide. In Aotearoa GAS disproportionately impacts Māori and Pasifika peoples, with significant numbers of children affected. Like many other bacterial infections, GAS infection can result in a wide variety of clinical outcomes, from minor to serious disease, but we still do not fully understand the genetic drivers associated with these differing outcomes.

The relationships between genotype (genetic) and phenotype (invasiveness or resistance to antimicrobials) are difficult to predict from genome sequences. The high m disease incidence rates in Aotearoa and the high morbidity rates among Māori and Pasifika communities highlight an urgent need to better understand the genetic drivers of the range of infections cause by GAS.

To do this, the project aims to: 1) gain insights into the genetic diversity and population structure of invasive and non-invasive GAS isolates circulating amongst children in Aotearoa and 2) identify genetic drivers responsible for invasive phenotypes (genotype-phenotype prediction).

We are using a machine learning algorithm to identify genetic variants that are associated with more invasive bacteria (i.e., are more likely to

cause severe disease), which requires a large and diverse dataset to *train* the machine. The dataset will be comprised of a collection of GAS genomes from samples collected in those communities that are most affected by this disease, as well as genomes from GAS isolates where the bacteria invaded the body.

The project will further apply the machine learning model on a data set of *Campylobacter jejuni*, with the goal of making the method applicable to different bacterial species.

The information gained will bring us a step closer to the ability to predict potential outcomes of bacterial infections from genomic data.

Led by Assoc Prof Paul Gardner (University of Otago), Dr Joep de Ligt (ESR)

Invasomics for better biosecurity and invasive species management

Invasive species threaten native biodiversity, primary industries, and health. Growing international trade, increasing mobility, and climate change make Aotearoa New Zealand's border highly vulnerable to new pests and diseases. Our unique ecosystems add a layer of complexity to our responses to these threats. Innovative and customised solutions that address the threat of new pests arriving are therefore urgently needed by policymakers and practitioners.

Predicting the invasive potential of species would enable us to better prioritise species for prevention, and management. The project aims to address key knowledge gaps in prediction of inherent invasiveness by building a generalisable tool that characterises and compares invasive and non-invasive populations and species to identify the genome, microbiome, and phenome traits that characterise successful invaders.

The project has chosen three high risk species that could significantly impact primary production if they arrive in Aotearoa (Brown marmorated stink bug, Queensland fruit fly and Spotted wing *Drosophila*) as models to: (1) Identify 'omic signatures that underlie successful biological invasion; and (2) Use 'omic signatures to predict the invasive potential of future invader

Members of the project team will work closely with biosecurity agencies (e.g., Ministry for Primary Industries, Māori Biosecurity Network, Better Border Biosecurity, Bio-Protection CoRE), as well as major primary production sectors. The project aligns closely with the on-going High-Quality Genomes and Population Genomics project.

The Invasomics project furthers Genomics Aotearoa's goals to connect bioinformatics researchers and end- users, nationally and internationally, by building on and extending the existing skill set.

Led by Dr Manpreet Dhami (Manaaki Whenua – Landcare Research), Ang McGaughran (University of Waikato)

Ruatau: connecting Māori genomic scientists and communities

Empowering Māori communities to make informed decisions about important genomics issues such as hauora (health) in people, or kaitiakitanga (guardianship) of taonga species is critically important. Māori scientists (kaipūtaiao Māori) can help this through understanding community perspectives, priorities, and concerns, but Māori are underrepresented in academia and in genomics research.

Furthermore, many Māori genomics researchers grew up Te Ao Pākehā and feel that they lack a sufficiently strong grounding in Te Ao Māori.

Developing cultural competency in these Māori genomic scientists in collaboration with Māori communities will therefore increase the understanding of genomics and associated risks and benefits in Māori communities and will enable the kaipūtaiao Māori to increase their understanding of mātauranga Māori/tikanga and reconnect with their own whānau/hapū.

The Ruatau project will achieve this by bringing Māori communities and kaipūtaiao together in

wānanga (workshops) focused on genomics. The knowledge of both groups will be extended in a two-way tuakana-teina relationship through the examination of problems and questions about genomics that are relevant to Māori communities.

Ultimately, cultural upskilling and genomics wananga materials will be made available for use by other researchers/communities.

Led by Dr Alana Alexander (University of Otago), Dr Catherine Collins (University of Otago), Dr Simon Hills (Massey University)

Environmental Microbiomes, from virus to eukaryote

The metagenomics approach is rapidly transitioning into a mainstream tool in microbial ecology. However, there is a need for approaches to reconstruct and analyse challenging mixtures of prokaryotic (bacterial and archaeal) community sequence data along with the genomes of other understudied organisms (viruses and microeukaryotes). The Environmental Microbiomes project looks to meet this need by providing clear methods and guidelines.

The Environmental Microbiomes project encompasses three work streams: (1) virome, (2) microbial strain heterogeneity, and (3) microeukaryote (tiny multicellular organisms) analyses.

The project will develop best practice guidelines for DNA and RNA virome analysis, and an approach for viral genome fragment dereplication in multisample datasets. This work capitalises on the virus genomes generated from the Environmental Metagenomics project. Additionally, two model systems will be used to explore microbial strain heterogeneity, improving assemblies using data

subsampling to reduce sequence complexity and using chromatin-based sequencing. Finally, a new protocol will be developed to generate microeukaryote genomes from metagenomic data.

The project aligns with key objectives for Genomics Aotearoa capability building and advanced genomics research, through the contribution of step-by-step workflows and hands-on-training, and research at the leading edge of microbial ecology

Led by Dr Kim Handley, (University of Auckland)

Scion

Genomics Aotearoa welcomes Scion as a new partner, joining our other nine partners — AgResearch, ESR, Plant & Food Research, Manaaki Whenua - Landcare Research, and the Universities of Auckland, Massey, Otago, Waikato, and Te Herenga Waka - Victoria University of Wellington.

Scion is a Crown Research Institute that specialises in research, science and technology development for forestry, wood product, wood-derived materials, and other biomaterial sectors. Scion is using genomics technologies to assemble the *Pinus radiata* genome and is developing molecular tools to assist the

genetic improvement of new *Pinus radiata* varieties. In addition, Scion uses genomics to characterise the function and genetic diversity of genes responsible for key forestry and wood product traits for the economic, social, environmental and cultural benefit of Aotearoa - New Zealand.



Contributing researchers

Aotearoa Variome

| Stephen Robertson | University of Otago |
|--------------------|------------------------|
| Phil Wilcox | University of Otago |
| Murray Cox | Massey University |
| Allamanda Faatoese | University of Otago |
| Lisa Matisoo-Smith | University of Otago |
| Tony Merriman | University of Otago |
| Andrew Sporle | University of Auckland |
| Huti Watson | Ngati Porou Hauora |

Clinical Genomics

| Stephen Robertson | University of Otago |
|-----------------------|---------------------------|
| Cheng-Yee Chan | Canterbury DHB |
| Kate Gibson | Genetic Health Service NZ |
| lan Hayes | Genetic Health Service NZ |
| Natasha Henden | Auckland DHB |
| Colina McKeown | Genetic Health Service NZ |
| Jo Martindale | Capital and Coast DHB |
| Kate Neas | Genetic Health Service NZ |
| Padmini Parthasarathy | University of Otago |
| Rachel Stapleton | Genetic Health Service NZ |
| Patrick Yap | Genetic Health Service NZ |

Genomic Translational Oncology

| Cris Print | University of Auckland | |
|--------------------|------------------------|--|
| Henry Chan | Waitemata DHB | |
| Jaclyn Ting Fowler | Auckland DHB | |
| Kim Gamet | Auckland DHB | |
| Lindsey Harbour | Auckland DHB | |
| Richard King | Canterbury DHB | |

Rakeiora

| Oris Print | University of Auckland |
|-------------------------|------------------------|
| Stephen Robertson | University of Otago |
| Phil Wilcox | University of Otago |
| Donia Macartney-Coxson | ESR |
| Miles Benton | ESR |
| Ben Curran | University of Auckland |
| Kimiora Henare | University of Auckland |
| Jennie Harre Hindmarsh | Ngati Porou Hauora |
| Martin Kennedy | University of Otago |
| Frances King | Ngati Porou Hauora |
| Joep de Ligt | ESR |
| Polona Le Quesne Stabej | University of Auckland |
| Hugh Stuart | Ngati Porou Hauora |
| Ben Te Aika | University of Otago |
| | |

Environmental Metagenomics

| Kim Handley | University of Auckland |
|-------------------------|--------------------------|
| Carmen Astudillo-Garcia | University of Auckland |
| Jian Shen Boey | University of Auckland |
| Michael Hoggard | University of Auckland |
| Gavin Lear | University of Auckland |
| Charles Lee | University of Waikato |
| Chanenath Sriaporn | University of Auckland |
| Matt Stott | University of Canterbury |
| Hwee Sze Tee | University of Auckland |
| David Waite | University of Auckland |
| Louise Weaver | ESR |
| Bevan Weir | Manaaki Whenua |
| Susie Wood | Cawthron Institute |

High Quality Genomes and Population Genomics

| Thomas Buckley | Manaaki Whenua |
|---------------------|--------------------------------------|
| Shannon Clarke | AgResearch |
| David Chagné | Plant & Food Research |
| Anna Santure | University of Auckland |
| Alana Alexander | University of Otago |
| Rachel Ashby | AgResearch |
| Rudiger Brauning | AgResearch |
| Chris Couldrey | LIC |
| Peter Dearden | University of Otago |
| Manpreet Dhami | Manaaki Whenua |
| Natalie Forsdick | Manaaki Whenua |
| Neil Gemmell | University of Otago |
| Joseph Guhlin | University of Otago |
| Charles Hefer | AgResearch |
| Andrew Hess | AgResearch |
| Elena Hilario | Plant & Food Research |
| Jeanne Jacobs | AgResearch |
| Michael Lee | University of Otago |
| Libby Liggins | Massey University |
| Tom Oosting | Victoria University of Wellington |
| Duckchul Park | Manaaki Whenua |
| Jessie Prebble | Manaaki Whenua |
| Peter Ritchie | Victoria University of Wellington |
| Tammy Steeves | University of Canterbury |
| Roy Storey | Plant & Food Research |
| Susan Thomson | Plant & Food Research |
| Amali Thrimawithana | Plant & Food Research |
| Maren Wellenreuther | Plant & Food Research |
| Annabel Whibley | University of Auckland |
| Chen Wu | Plant & Food Research |

Epigenome-Wide Association Study

| Greg Jones | University of Otago |
|----------------------|---------------------|
| Basharat Bhat | University of Otago |
| Aniruddha Chatterjee | University of Otago |
| John Pearson | University of Otago |

Bioinformatics

| Mik Black | University of Otago |
|-------------------------|------------------------|
| Rudiger Brauning | AgResearch |
| Kim Handley | University of Auckland |
| Miles Benton | ESR |
| Rachael Ashby | AgResearch |
| Carmen Astudillo-Garcia | University of Auckland |
| Peter Dearden | University of Otago |
| Ngoni Faya | University of Otago |
| Michael Hoggard | University of Auckland |
| Dan Jones | Plant & Food Research |
| Leah Kemp | ESR |
| Libby Liggins | Massey University |
| Roy Storey | Plant & Food Research |
| | |

SING Aotearoa

| Maui Hudson | University of Waikato |
|---------------------|------------------------|
| Kimiora Henare | University of Auckland |
| Tuti Nikora | University of Waikato |
| Katharina Ruckstuhl | University of Otago |
| Phil Wilcox | University of Otago |

Te Nohonga Kaitiaki

| Maui Hudson | University of Waikato |
|-------------------|-----------------------------|
| Chris Battershill | University of Waikato |
| Robert Brookes | University of Waikato |
| Jason Mika | Massey University |
| Tuti Nikora | University of Waikato |
| Jacinta Ruru | University of Otago |
| Matt Stott | University of Canterbury |
| Ariane Thompson | University of Waikato |
| Lisa Warbrick | Massey University |
| Phil Wilcox | University of Otago |
| | |

Glossary of genomic terms

| TERM | DEFINITION |
|------------------------|---|
| Assembly | DNA sequencing technology cannot read whole genomes in one go; instead, it reads short pieces of bases from a genomic sequence. Sequence assembly aligns and merges fragments from a longer DNA sequence to reconstruct the original sequence. |
| Bioinformatics | The science of analysing genomic data. |
| Candidate genes | Genes of interest related to phenotypes or disease states. |
| Clinical geneticist | This is a medical doctor with special training in genetics who meets with patients to evaluate, diagnose, and manage genetic disorders. Clinical geneticists also assist in the management of genetic diseases by identifying preventable complications through early and accurate diagnosis and surveillance. |
| CRISPR-Cas9 | CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) is a method of genetic manipulation consisting of two key molecules that introduce a change into the DNA. The molecules are: |
| | An enzyme called Cas9 which can cut strands of DNA at a specific location in the genome so that short sections of DNA can then be added or removed. |
| | A piece of RNA sequence called guide RNA (gRNA), located within a longer RNA which guides Cas9 to the correct part of the genome to cut. |
| DNA methylation | DNA methylation is one type of epigenetic mechanism that modifies the expression of genes and therefore their effects. |
| | DNA bases are modified by addition of a methyl group. Methylation can change the activity of a DNA segment without changing the genome sequence. |
| Epigenetics | Epigenetic mechanisms are those that modify inherited (genomic) gene expression i.e., how, and when those genes are switched on (or off). These epigenetic switches may be due to environmental exposure or lifestyle factors. |
| | Non-genetic factors include the environment like diet, gut microbiota, toxin, and drug exposure, psychological and physical stressors, and levels of activity throughout life. Measuring the epigenetic changes that occur in diseases, including cancer and heart disease, can provide understanding of the underlying mechanisms. |
| Epigenome | The set of all epigenetic modifications to an individual's genome. |
| Eukaryote | An organism whose cells contain a nucleus surrounded by a membrane and whose DNA is bound together by proteins (histones) into chromosomes. Animals, plants, and fungi are eukaryotes. |
| Exemplar research | A research model that provides leadership and examples for further research in similar fields. |
| Finishing the sequence | Finishing an assembly involves refining the genomic sequence to eliminate sequencing errors and to close gaps. |

| TERM | DEFINITION |
|--|---|
| Genetic admix/ admixture | The presence of DNA in an individual from a distantly related population or species, because of interbreeding between populations or species who have been reproductively isolated and genetically differentiated. Admixture results in the introduction of new genetic lineages into a population. |
| Genetic counsellors | These are healthcare professionals with training in human genetics and counselling who guide patients and their whānau with a genetic disorder through the process of understanding and making informed healthcare decisions. |
| Genetic-linkage mapping | Illustrates the order of genes on a chromosome and the relative distances between those genes. |
| Genome annotation | The process of attaching biological information to genetic sequences. |
| Genome map | Helps scientists to define which parts of the genome are physically linked to each other. The landmarks on a genome map that aid navigation might include short DNA sequences, regulatory sites that turn genes on and off, and genes themselves. |
| Genome variation types | Genome variations include mutations and polymorphisms. Mutation is often used to refer to a variation that is associated with a specific human disease, while the word polymorphism implies a variation that may or may not affect a physical characteristic. Genetic variations also include gene deletions, gene additions and structural variations. |
| Genome-wide association studies | A search for parts of the genome associated with characteristics of interest, one example being human diseases. |
| Genotype (noun) | An organism's set of genetic variations. |
| Genotype (verb) | To determine genetic variation in a genome. |
| Germline | The sequence of cells which develop into eggs and sperm. |
| Imputation | The mathematical process of replacing missing data with substituted values. |
| Introns | Extra sequences of DNA inside genes that are non-coding. |
| Long read sequencing | Involves new forms of sequencers that can read long distances down one strand of DNA. There are currently two effective technologies: |
| Also known as: Third generation sequencing | Pacific Biotechnology (PacBio) - an imaging approach that allows the detection of the incorporation of single labelled base pairs one after another into a strand of DNA being replicated (pacb.com/videos/video-overview-of-smrt-technology) |
| | Oxford Nanopore (Nanopore) - uses tiny, charged pores that a strand of DNA is drawn into, and as each base passes through the hole it changes the charge in a way that can be measured (https://nanoporetech.com/how-it-works). These changes in charge are then assigned to each base and the sequence is built up from there. |
| Linked read technology | Uses a unique barcode system to label short DNA sequences from individual molecules that are close to each other on the genome, so they can be linked to create longer sequence reads. |
| Metabolomics | This process detects chemicals or metabolites and provides a read-out of what chemicals are in a tissue at any one time. |
| Mutations | The changing of the structure of a gene, caused by the alteration of single base units in DNA, or the deletion, insertion, or rearrangement of larger sections of genes or chromosomes. The resulting variant form may be transmitted to subsequent generations. |

| TERM | DEFINITION |
|--|--|
| Non-coding DNA | Genes (coding DNA) account for a small percentage of the DNA in the genome - knowing the entire genome sequence will help scientists study the parts of the genome outside the genes. Non-coding DNA includes the regulatory regions that control how genes are turned on an off, as well as long stretches of DNA of unknown function. |
| Nonsense mutation | A DNA mutation that results in a non-functional protein. |
| Nucleotides and bases | A single base of DNA is made up of Adenine (A), Cytosine (C), Guanine (G) or Thymine(T). Base pairs are the two opposing nucleotides (a base, a sugar molecule, and a phosphate group) on a double-stranded DNA molecule. Adenine pairs with Thymine, and Cytosine pairs with Guanine. |
| Phenotype | Observable characteristics influenced both by an organism's genotype and by the environment. |
| Pathway | In genetics, a pathway is a set of genes that work together in a biological process. |
| Pharmacogenomics | This is the study of how genomic variation within the individual or their disease (including gene expression, epigenetics, germline, and somatic mutations) influences one person's response to drugs. The aim is to optimise drug therapy by maximising therapeutic effect and minimising adverse effects. |
| Pipeline | A process for the preparation, development, production, and analysis of genomic data. |
| Prokaryote | Unicellular microbial organism that lacks a nucleus. All bacteria are prokaryotes. A related group are Archaea – these are also unicellular but are different to bacteria. |
| Proteomics | RNA (made from turned-on genes) is translated into protein. Proteomics is a technique to look at the broad range of proteins in a cell or tissue (using mass spectrometry). We can usually identify which proteins are present in a cell and what they are doing. |
| RNA | RNA (ribonucleic acid) is the molecule that takes information from DNA to make protein and has many other activities. RNA is only made from genes in the DNA that are turned on. |
| Short read sequencing | Small segments of DNA strands put in order then assembled in the genome. This is assembled by looking at the sequence of each chunk and finding sequences that overlap (aligning). |
| Also known as: Next generation or second-generation sequencing | |
| SNPs | About 90 percent of human genome variation can be accounted for by single nucleotide polymorphisms, or SNPs (pronounced "snips"). These are variations that involve just one nucleotide, or base. These markers are used to genotype individuals to better understand population genetic variation and can be used to detect genes that are under selection. |
| Somatic | Refers to the cells of the body in contrast to the cells that make sperm or eggs. |
| Transcriptomics | Sequencing RNA from a tissue or cell to measure the set of active (expressed) genes. |
| Variome | The complete set of genetic variations found in populations of species. |
| Whole genome sequencing (WGS) | The process of determining the complete DNA sequence of an organism's genome. |
| | |



