genomics aotearoa

ROYAL SOCIETY TE APĀRANGI

GSA
GENETICS SOCIETY OF AUSTRALASIA INC.

CENTRE FOR COMPUTATIONAL EVOLUTION

# Summer Institute in Statistical Genomics

23 November – 4 December 2026

Waipapa Taumata Rau
University of Auckland

Image credit: Sulthan Auliya on UnSplash: https://unsplash.com/@swafie

# About the Summer Institute in Statistical Genomics

Registration link:
https://events.humanitix.com/southern-ssig-auckland-2026

The Summer Institute in Statistical Genomics is open to anyone from all over the world who wishes to learn about any aspect of statistical and quantitative genetics from leading researchers. This includes students, post-docs, industry and government employees, teachers, policy makers, and others.

Professor Bruce Weir founded the Summer Institute in Statistical Genetics (SISG) in 1995 while he was Director of the Bioinformatics Research Center at North Carolina State University in Raleigh. SISG has since run annually in the United States, where it is now hosted at the Georgia Institute of Technology in Atlanta.

We are delighted that for the 3rd time in its history, a version of SISG will be run in New Zealand, hosted at the University of Auckland – Waipapa Taumata Rau. Questions? Please contact Prof Bruce Weir and Assoc Prof Anna Santure: bsweir@uw.edu; a.santure@auckland.ac.nz

All modules for the Institute will be held in seminar rooms in the central University of Auckland Campus, and instruction is in-person only.

You may choose just one, or up to four, modules; most participants choose two or three. You can pick and choose any combination when you apply, and in most cases will be assigned to your first preferences (depending on availability: most classes are capped at 40).

Coffee, tea and snacks will be provided at morning and afternoon tea each day. For lunch, participants can self-cater or purchase from food vendors on and near the University campus.

We gratefully acknowledge support from our Institute Partners:

genomics aotearoa

ROYAL SOCIETY TE APĀRANGI

GSA
GENETICS SOCIETY OF AUSTRALASIA INC.

CENTRE FOR COMPUTATIONAL EVOLUTION

# Summer Institute in Statistical Genomics Modules

Modules last 2.5 days, starting on Monday morning and ending Wednesday lunchtime, or from Wednesday afternoon through Friday afternoon. Each module runs as 10x 90 minute sessions, generally alternating lectures and practical exercises.

**Week One: Monday 23 November – Friday 27 November**

| Mon 23/11/26 - Wed 25/11/26 | Wed 25/11/26 - Fri 27/11/26 |
|---|---|
| 1. Statistical Genetics<br><br>Prof Bruce Weir | 2. Conservation and Population Genomics<br><br>Assoc Prof Anna Santure & Assoc Prof Emma Carroll |
| | 3. Introduction to Bayesian Statistics<br><br>Dr Brendon Brewer |

**Week Two: Monday 30 November – Friday 4 December**

| Mon 30/11/26 - Wed 2/12/26 | Wed 2/12/26 - Fri 4/12/26 |
|---|---|
| 4. Comparing Patterns of Quantitative Genetic Variation<br><br>Assoc Prof David Aguirre | 6. Applied Quantitative Genetics<br><br>Dr Ken Dodds & Dr Timothy Bilton |
| 5. Polygenic Risk Scores<br><br>Assoc Prof Yalu Wen | 7. Metagenomics and Gene Function Analysis<br><br>Assoc Prof Kim Handley & Dr Chloé van der Burg |

# 1. Statistical Genetics

This module concerns statistical methods for population and quantitative genetic analysis. It provides a unified treatment for the analysis of discrete genetic data, starting with estimates and sample variances of allele frequencies to illustrate genetic vs statistical sampling.

Concepts are demonstrated with exercises in R. We will take a detailed look at Hardy-Weinberg and linkage disequilibrium and how to characterize population structure using F-statistics. We base analyses on allelic matching within and between populations with individual inbreeding and relationship estimation as a special case, with reference to Weir and Goudet, Genetics iyag088 (2026).

**Learning Objectives**: After attending this module, participants will be able to:

1. Estimate allele frequencies from genotype counts and to estimate within-population inbreeding coefficients.
2. Determine sample sizes required to detect inbreeding and perform goodness-of-fit tests for Hardy-Weinberg equilibrium. Use open source software to perform exact tests for Hardy-Weinberg equilibrium.
3. Predict kinship levels for pairs of individuals in pedigrees or populations and know which software to use for particular applications.
4. Calculate allele-matching proportions within and between populations in order to estimate population structure parameters.
5. Evaluate population structure using open source software.

**Bruce Weir** is a statistical geneticist with interests in drawing inferences from population and quantitative genetic data, and with applications to forensic science. He is currently looking at SNP data for the North Island brown kiwi (*Apteryx mantelli*). He has academic appointments in Biostatistics at the University of Washington and in Statistics at the Universities of Auckland and Otago.

# 2. Conservation and Population Genomics

This module provides an introduction to foundational theory and methods to utilise genomic data to address key questions in ecological and conservation genomics. We will introduce and contrast common sequencing approaches in the field, including reduced representation sequencing and low coverage whole genome sequencing. The course will also cover quality control of genomic data, including what considerations are important for different sequencing methods and analyses. Students will then be lead through hands-on tutorials of key population genetic approaches including population structure analyses such as clustering and differentiation indices, current and historical effective population size, estimation of inbreeding and relatedness, and parentage assignment.

**Learning objectives**: After attending this module, participants will be able to:

1. Compare and contrast the costs and benefits of different sequencing approaches
2. Conduct quality control appropriate to different sequencing datasets and analyses
3. Apply common methods in population genomics, including clustering, the calculation of differentiation indices, estimation of inbreeding and relatedness, effective population size and historic demography
4. Use genomic data from within a population to conduct parentage analysis

The course assumes familiarity with a unix environment, familiarity with coding, and good confidence in R.

**Emma Carroll** is an Associate Professor in Molecular Ecology at the University of Auckland. After completing a PhD and then postdoctoral fellowship based at the University of Auckland, she took up a research fellowship at the University of St Andrews, Scotland before returning to New Zealand on a Rutherford Discovery Fellowship. Emma's research integrates her training is in evolutionary biology and population genetics with diverse cross-disciplinary methods and approaches including stable isotope ecology, behaviour and culture, quantitative statistical analyses and historic samples to inform current management decisions for marine mammals.

**Anna Santure** is an Associate Professor in Genetics at the University of Auckland. Following a PhD in quantitative genetics at the University of Otago, she held postdoctoral positions at the Zoological Society of London and University of Sheffield, UK, before returning to take up a lecturing position in Auckland in 2013. Anna is interested in predicting the ability of species to adapt in a rapidly changing world, with recent research projects focusing on both threatened taonga (culturally significant) and invasive species in Aotearoa New Zealand.

# 3. Introduction to Bayesian Statistics

This module will cover the basics of the Bayesian approach to statistics. The core idea is the use of probability distributions to describe uncertainty, rather than randomness or variability. This change in interpretation enables us to carry out data analysis in a principled and consistent way, which feels different from the common "bag of tricks" approach to statistics.

We will start with elementary problems that illuminate the conceptual foundations of the subject. To apply Bayesian inference to real problems, computational techniques including Markov Chain Monte Carlo (MCMC) will be introduced in the R language. We will revisit common statistical situations such as regression from a Bayesian point of view, and we will introduce hierarchical models which can express complex but (often) more realistic prior information to improve analyses. These ideas underpin many modern high-dimensional methods, including those used in contemporary genetic analyses. Finally, we will investigate Bayesian model comparison techniques.

**Learning Objectives**: After attending this module, participants will be able to:

1. Understand the distinction between Bayesian and classical/frequentist probabilities.
2. Understand what prior distributions, likelihood functions, and posterior distributions represent.
3. Compute posterior distributions in R for a single unknown quantity.
4. Implement hierarchical Bayesian models in JAGS.
5. Run Nested Sampling on simple inference problems and interpret the results.

**Brendon Brewer** obtained his PhD in Astrophysics from the University of Sydney in 2008. In 2012 he joined the University of Auckland's Department of Statistics, but this was not really a career change, as he has always had one foot in astrophysics and the other in statistics. His research interests include Bayesian inference, information theory, astrostatistics, and sports statistics.

# 4. Comparing Patterns of Quantitative Genetic Variation

This module will introduce the statistical techniques used to estimate additive genetic variances and covariances in captive-bred and wild populations, with or without strict controlled breeding. We will explore the concept of multivariate genetic constraints and demonstrate how to compare patterns of quantitative genetic covariation among populations or sampling units.

Concepts are demonstrated with exercises in R. Although a working understanding of R, matrix algebra and linear models will be beneficial, the course will provide an introduction to the relevant components of each topic. Naturally, this also means there will be some overlap in fundamental genetics and statistics content with other modules in the Summer Institute of Statistical Genetics.

**Learning Objectives**: After attending this module, participants will be able to:

1. Use linear models, including animal models, to partition phenotypic variation within a population to estimate additive genetic effects.
2. Use linear models to estimate additive genetic variances and covariances from data derived from controlled breeding and uncontrolled breeding.
3. Understand the concept of multivariate genetic constraints.
4. Compare patterns of quantitative genetic covariation among populations or sampling units.

**Dave Aguirre** is an Associate Professor of Ecology and Evolutionary Biology at the University of Auckland, Auckland, New Zealand. His research focuses on estimating additive genetic variances and covariances in captive and wild populations. In particular, he is interested in exploring patterns of multivariate quantitative genetic variation among populations or sampling units to better understand genetic constraints on phenotypic evolution.

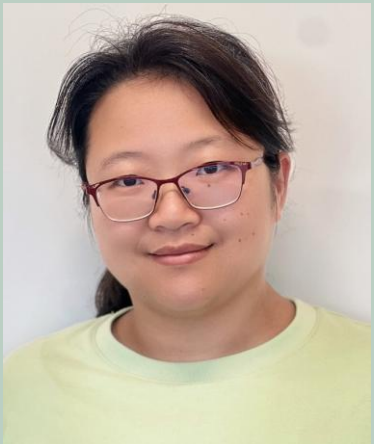
# 5. Polygenic Risk Scores

This module provides students with the fundamental tools to construct and evaluate polygenic risk scores (PRS) using summary statistics and individual-level data, with a strong emphasis on both single-ancestry and cross-ancestry prediction, and hands-on learning. Familiarity with R is assumed, and knowledge of PLINK and python is recommended.

Topics covered include construction of PRS using summary statistics and individual-level data; quality control procedures; population structure, admixture, and principal component analysis (PCA); clumping and thresholding (C+T) methods; Bayesian approaches (e.g., LDpred2, PRS-CS); development and use of linkage disequilibrium (LD) reference panels; evaluation of predictive performance; PRS development within single-ancestry settings; challenges in cross-ancestry PRS transferability; multi-ancestry PRS methods (e.g., PRS-CSx); integration of PRS with clinical and environmental risk factors; validation and tuning strategies; and considerations for study design, bias, and equity across diverse populations. An important component of this module is in-class software exercises, which provide students with hands-on experience analysing real data using state-of-the-art tools for PRS construction and evaluation.

**Learning Objectives:** After attending this module, participants will be able to:

1. Construct PRS using summary statistics and/or individual-level data.
2. Apply quality control procedures for PRS construction.
3. Apply common PRS methods, including C+T, LDpred2, and Bayesian approaches (e.g., PRS-CS).
4. Develop and evaluate PRS within single-ancestry populations.
5. Implement multi-ancestry PRS methods (e.g., PRS-CSx).
6. Select, develop and evaluate appropriate LD reference panels and assess their impact on PRS performance.
7. Integrate PRS with clinical and environmental risk factors in regression models.
8. Conduct validation and model tuning using training, validation, and test datasets where appropriate.
9. Apply PRS methods to real datasets and critically evaluate their limitations and appropriate use.



**Yalu Wen** is an Associate Professor in Statistics at the University of Auckland. She received her PhD in Epidemiology and Biostatistics from Michigan State University, USA. Her research focuses on the analysis of large-scale human genomics data using a range of statistical approaches, including genetic variant calling, genotype annotation, exome-wide and genome-wide association studies, and the development and evaluation of polygenic risk scores. Her current work centres on developing equitable polygenic risk scores across diverse populations, particularly those that are under-represented in global genomic resources.

# 6. Applied Quantitative Genetics

This module concerns the analysis of quantitative genetic data. These data typically consist of phenotypic measures of traits on a continuous scale, where the phenotypes are influenced by many genes and environmental effects. Quantitative genetics is used in plant and animal breeding, human health and behaviour, ecology and evolutionary biology. Some of the underpinning theory of quantitative genetics and statistics covered in this course overlaps with other courses in the Summer Institute of Statistical Genetics, but the focus here is the application of this theory in primary industries.

The course will include the fundamentals of: the genetic basis of quantitative traits, relatedness and inbreeding, contributors to phenotypic variation, estimation of genetic parameters (heritability, repeatability, genetic correlations, maternal heritability), estimation of genetic value, response to selection and the design of breeding programmes. The course will focus on the use of pedigree-based relatedness but will also include an introduction to the use of genomics in this context. There will also be teaching in the statistical methods (regression, linear models, mixed effects models) underpinning the genetic analyses. The methods will be demonstrated with exercises in R.

**Learning Objectives**: After attending this module, participants will be able to:

1. Understand the genetic models underlying quantitative genetics analyses
2. Calculate relatedness values from a pedigree
3. Estimate genetic parameters such as heritability and genetic correlations from phenotypic data on relatives
4. Undertake genetic evaluations incorporating information from the individual and/or their relatives
5. Plan efficient designs for experiments to estimate genetic parameters and for breeding programmes to make genetic progress
6. Understand how genomic data can be used in genetic parameter estimation and genetic evaluation

**Ken Dodds** is a Principal Scientist at the Bioeconomy Science Institute. He develops and implements statistical tools for genetics and breeding, especially in livestock and aquaculture. He has developed and used methods to use DNA information including methods for genetic evaluation with fractionally assigned pedigrees and for using partial genotype information from low-depth sequencing data.

**Timothy Bilton** is a Senior Statistician at the Bioeconomy Science Institute. His research is focused on the analysis of genetic data and implementation of genetic evaluations for applications in animal and plant breeding. He also has a strong interest in developing new statistical methods and tools for improving analysis of genetic data.

# 7. Metagenomics and Gene Function Analysis

This module provides an introduction to a metagenomics workflow from genome assembly to analysis of genomic features, such as gene function. Lectures and practical sessions will provide an introduction to metagenome assembly of short-read data, and multi-tool based contig binning for generating metagenome-assembled genomes, including how these tools leverage genomic compositional features and contig coverage. This part of the module will also cover the evaluation of assembly and genome quality.

The second part of the module will focus on genome analysis. We will cover use of metagenome-assembled genomes to determine genome coverage and population relative abundances, and utility of genome dereplication. Additional topics will include taxonomic classification of genomes, genome annotation, and the analysis of annotations to assess gene function with interpretation taking into account limitations presented by genome quality.

**Learning Objectives**: After attending this module, participants will be able to:

1. Assemble genomes from short-read metagenomic sequence data, and evaluate assembly quality.
2. Bin metagenome-assembled genomes, and evaluate genome quality.
3. Understand how to taxonomically classify genomes.
4. Use read mapping to assess population relative abundances based on metagenome-assembled genomes.
5. Annotate genomes, and use annotations to understand cellular features and processes.

**Kim Handley** is an Associate Professor in the School of Biological Sciences at the University of Auckland and a Science Advisory Forum member for Genomics Aotearoa. She is a microbial ecologist who uses metagenomics alongside other omics techniques to understand the role of microorganisms in the environment, such as microbial contributions to biogeochemical cycles (e.g. nitrogen), the potential for inter-species interactions, and how genetic traits govern the spatial and resource niches occupied by microorganisms.

**Chloé van der Burg** is the Bioinformatics Training Coordinator at Genomics Aotearoa. She completed her PhD at the Queensland University of Technology in Brisbane, Australia, before moving to Dunedin and working as a Postdoctoral researcher at the University of Otago. She has a research background in molecular biology and transcriptomics and has worked on a diverse range of organisms, from fish to fruit flies. In her current role she coordinates and delivers training workshops in bioinformatics and genomics to researchers across New Zealand.