AWC Projects 2014/2015

Theme: 1. Genomics & Biomathematics
Activity: A. Developing cutting-edge genomic, bioinformatic and biomathematical technologies for New Zealand

Investigator	Project	Stakeholders
Bryant	**SNP based phylogeographic and demographic inference** **Team members**: David Bryant, Remco Bouckaert Single Nucleotide Polymorphisms (SNPs) are locations in a genome where individuals within a population or species differ by a single base-pair. They are widely used as `signposts' of the genome when detecting associations between genes and diseases or traits. We are developing tools to use these data for evolutionary and ecological genetics, identifying how species evolved and even estimating population sizes.	• Molecular ecologists using genetic data to infer population demographics • Agricultural and horticultural researchers
Cox	**Simulating how social behaviors influence genetic diversity in small human populations** **Team members**: Murray Cox, Elsa Guillot Perhaps uniquely among the world's species, humans are the product of both their natural and cultural environment. This is especially apparent in small human communities, where social behaviors, such as marriage practices, can strongly affect patterns of genetic diversity. This study develops new computational and theoretical tools to explore how cultural factors can affect genetic patterns, using the diverse populations of the Indonesian archipelago as a case study system.	• Eijkman Institute for Molecular Biology, Jakarta, Indonesia
Gemmell	**Simple sequences do not evolve simply: What is the role of recombination in microsatellite evolution?** **Team memebers:** Neil Gemmell Monika Zavodna, Andrew Bagshaw. Microsatellites are abundant, highly variable, repeated DNA sequences that are regarded as the most versatile genetic markers yet discovered. They have been a cornerstone of the recent revolution in genetics and are used in gene mapping, in DNA forensic work, and as population markers. Conclusions drawn from such studies in many cases depend critically on assumptions about how microsatellites evolve. Despite its importance, our understanding of microsatellite evolution remains surprisingly sketchy. Current models of microsatellite evolution are overly simplistic and almost certainly incorrect, potentially leading to widespread data misinterpretation. In particular, genetic recombination, whilst known to be the major generator of genomic variability, is widely regarded as a minor contributor to microsatellite evolution, if indeed it contributes at all. Here we seek to quantify the extent and nature of microsatellite mutation, and, for the first time, explicitly examine the role of sexual recombination in generating microsatellite variability, using near identical yeast strains that variously do and do not engage in sex, and thus recombination. If, as we suspect, recombination is an important force in microsatellite evolution, evolutionary models will need to be modified, potentially with far-reaching consequences for how we analyse and interpret microsatellite data and for our understanding of trinucleotide repeat disorders, such as Fragile X and Huntington's disease.	
Gemmell	**To what extent do tandem repeated sequences modulate gene expression?** **Team members:** Neil Gemmell, Sterling Sawaya (PhD Student), Andrew Bagshaw, Mik Black, Greg Jones, Martin Kennedy A small but significant portion of the genome is composed of short tandem repeats known as microsatellites. These sequences generally expand and contract at a rate that is much higher than the average rate of point mutation. Microsatellite expansion can result in disease, but not all microsatellite variation is harmful. For example, morphological or behavioral variation induced by microsatellite length variability may be beneficial in a changing environment. Recently, our group has found that microsatellites near the start of genes, the region we refer to as the gene's promoter, are often highly conserved in mammals. We have also discovered that microsatellites occur in human promoters more frequently than would be expected by chance and are strongly associated with regulatory elements. Together these observation suggest these sequences are potentially important sources of variation in gene expression and we are now actively working to understand the importance of these sequences on various human phenotypes.	

Investigator	Project	Stakeholders
Matisoo-Smith	**Development and application of new bioinformatic tools for understanding genetic change in a population through time** **Team members**: Post Docs - Ann Horsburgh, Darnell Kennedy; PhD students: Karen Greig, Anna Gosling We are working on several projects involving data from ancient and modern animal and human populations. In addition to applying standard Bayesian and population genetic modelling tools, we are working with colleagues in the US and Europe to develop new methods of analysis for population data through time.	• Archaeologists & other colleagues dealing with historical research • Pacific Island communities who are the descendants of the samples we are working with
Poole	**An RNA encyclopedia of Bacteria & Archaea** **Team members:** Anthony Poole, Dr Paul Gardner (UC), Dr Stinus Lindgreen (EU Marie Curie Fellow), Dr Patrick Biggs (AWC Affiliate), PhD students (UC): Sinan Umu, Wenting Liu Genome projects routinely annotate protein-coding genes, but rarely screen for non-coding RNA genes. These are often non-trivial to identify, but can be extremely important, as they are heavily involved in gene regulatory and other key processes. We will use phylogeny-informed comparative genomics to choose species that are at an optimal distance for the identification of non-coding RNAs from transcriptomes.	
Steel & Bryant	**Development of tools for genomic phylogenetics** **Team members:** Mike Steel and David Bryant One of the challenges arising from new sequencing technologies is how to use the large amount of patchy genomic data that is generated to learn more about evolutionary relationships. In our research to date, we have provided the first mathematical analysis of how much species coverage is required in order to have a high probability of being able to reconstruct a tree with confidence. Our results show that many large-scale data sets prove to be incapable of resolving certain evolutionary relationships. We have also described new approaches to estimating ancestral sequences in trees that provide sufficient taxon coverage, and developed new approaches for tree comparison. We plan to develop these and other related themes in 2012.	
Steel & Semple	**Development of network-based evolutionary analysis** **Team members***:* Charles Semple and Mike Steel Phylogenetic networks are an essential tool for representing evolutionary relationships when evolution involves reticulate (non-tree-like) processes, such as (i) the formation of hybrid species, (ii) lateral gene transfer within microorganisms, (iii) the ancestry of a diploid population subject to recombination. In research to date, we have provided the first mathematical treatment of a model which biologists have used to try to quantify the extent of lateral gene transfer. We have also explored different ways to define a 'phylogenetic tree' when reticulate processes are at play, and investigated fundamental algorithmic questions that arise in phylogenetic networks. In 2012, we will consider some mathematical and algorithmic approaches to pedigree graphs, which arise in considering population history (process (iii) above).	

Theme: 1. Genomics & Biomathematics
Activity: B. Using new genetic knowledge for molecular ecology & evolution

Investigator	Project	Stakeholders
Bryant	**Improved Network Based Inference Tools** **Team members:** David Bryant, Daniel Huson (International Advisory Panel), Jessica Leigh Phylogenetic trees are like family trees for different species. Split networks are like phylogenetic trees on steroids: they represent more complex relationships between species, and so can be used to make inferences when nature is too complicated for simple trees. AWC members past and present have made key contributions to the development of these methodologies, however there is much more to be done. In this project we are developing new network methods and software for both metagenomic and population level data, working with collaborators Daniel Huson and Naruya Saitou.	• Evolutionary biologists • Medical researchers

Investigator	Project	Stakeholders
Buckley	**Transcriptome evolution in New Zealand invertebrates** **Team members**: Thomas Buckley, Alice Dennis, Luke Dunning, Victoria Twort, Richard Newcomb We are using transcriptome data from New Zealand stick insects and weta to identify candidate genes involved in cold tolerance and reproductive isolation. The study of cold tolerance is directed at stick species from a range of habitats from the alpine zone to relatively warm, northern forests. Genetic variation at these candidate loci is being screened across these populations and species. Neutral markers obtained are also being used to obtain a robust phylogeny for the New Zealand species and their overseas relatives. The transcriptome sequencing in weta is targeted at reproductive tissues, and focused on highly threatened species, especially giant weta. These reproductive genes will be used to examine gene flow between species and populations.	• Department of Conservation (Primary)
Buckley	**Speciation and adaptation in New Zealand Invertebrates** **Team members:** Thomas Buckley, Richard Newcomb and Howard Ross We are using genome scale sequence data coupled with RAD tag data to test hypotheses on stick insect and giant weta population genetics and speciation. Genome scale assemblies will be used as a template for mapping RAD tag data for detecting loci under selection. Population genetic processes will also be identified and used to inform conservation genetic strategies. The target species include island endemics, relictual taxa and more widespread species. Each of these comparisons will give novel insights into processes of adaptation.	• Ngatiwai, Department of Conservation (both (Primary)
Cox	**Reconstructing the effects of admixture between Asian and Melanesian groups on the genomic structure of modern Indonesian peoples** **Team members**: Murray Cox, Armando Amaris, Aydar Aliev From a simplistic perspective, the Pacific region can be viewed as being settled in two waves: the first, ancestors of today's Melanesians, arrived ~50 kya, while groups with Asian ancestry appeared more recently, perhaps only ~5 kya. Most Indonesian individuals carry genetic markers from both sources, but how this ancestry is distributed along their chromosomes is poorly understood. This study uses high throughput SNP screening to infer how the admixture process has produced mosaic genome structures in modern Indonesians, and from this, aims to reconstruct aspects of the social and demographic practices that occurred during the admixture process.	• Eijkman Institute for Molecular Biology, Jakarta, Indonesia • Pacific communities
French	**Cows, starlings and *Campylobacter* in New Zealand: unifying phylogeny, genealogy and epidemiology to gain insight into pathogen evolution** **Team members:** Nigel French, Barbara Holland, Murray Cox In this study we use multilocus sequence typing and full genome analysis to discover how often, and how much, genetic material is exchanged between natural populations of the important zoonotic pathogens *Campylobacter jejuni* and *C. coli*. We examine how important recombination is relative to mutation for the emergence of new strains; and in which host species these events are most likely to occur. Ultimately we can learn how and why *C. jejuni* emerged to become such a prominent human pathogen; anticipate further evolution and restrict the emergence and spread of new strains.	• MAF-food safety • International infectious disease / pathogen evolution research community • Postgraduate students – training graduate students from NZ and internationally
Gemmell	**Parallel tagged next-generation sequencing in ecology and conservation** **Team members**: Neil Gemmell, Monika Zavodna, Catherine Gruber Next-generation sequencing (NGS) on pooled samples has already been broadly applied in human medical diagnostics and plant and animal breeding. However, thus far it has been only sparingly employed in ecology and conservation, where it may serve as a useful diagnostic tool for rapid assessment of species genetic diversity and structure at the population level. Here we undertake a comprehensive evaluation of the accuracy, practicality and limitations of parallel tagged amplicon NGS on pooled population samples for estimating species population diversity and structure.	**Conservation biologists and ecologists** interested in rapidly determining the genetic composition of species and populations
Gemmell	**Genotype-by-sequencing: a powerful new tool for molecular ecologists.** **Team members:** Neil Gemmell, Shannon Clarke (AgResearch) We are exploring the use of new Genotype-By-Sequencing GBS approaches that couple restriction site-associated DNA sequencing (RAD-seq that reduces the complexity of the genome) with tagging approaches to enable >10,000 informative SNP markers to be screened in a population set, even when only limited or no prior genetic knowledge was available for the species. Such an approach has particular appeal to species for which there is little prospect for high-density genomic tools to be developed, where only modest numbers of individuals will ever be analysed i.e. <1000. Currently our work is focused on applying these tools in an agricultural and aquaculture context, but we are interested in utilising these tools in a	**Ecologists, conservation and evolutionary biologists** interested in determining the genetic composition of species and populations in significant detail

Investigator	Project	Stakeholders
	conservation and ecological context on species such as: fur seals, native frogs, birds and other species.	
Gray	.**Network modelling of language evolution in the Pacific: testing hypotheses from genetic data** **Team members**: David Bryant, Lisa Matisoo-Smith In this project we will analyse lexical data with network methods to test hypotheses about hybrid histories in the Pacific	• Pacific Island communities • Pacific archaeologists, linguists and prehistorians • General public
Jamieson	How to conserve allelic diversity in small populations.	• Dept. of Conservation, conservation biologists and community groups conducting translocations and reintroductions.
Matisoo-Smith	**Development and Application of Next Generation Sequencing Technology to Highly Degraded Samples** **Team members**: Post Docs –Ann Horsburgh and Darnell Kennedy; PhD students: Karen Greig and Anna Gosling; Technician Olga Kardailsky Much of the current work in our ancient DNA laboratory involves the development and refinement of next generation sequencing technology for use with degraded ancient samples, such as those recovered directly from archaeological sites in the Pacific (Greig, Gosling) and Africa (Horsburgh) or from museum collections. While next generation sequencing technology is well developed for generation of data using fresh tissue samples, there are significant problems inherent in the analysis of highly degraded samples, particularly when those samples derive from modern human remains. The combination of high levels of contamination by museum staff and archaeologists and extremely poor preservation require novel approaches for identifying authentic DNA sequences from archaeological remains.	• Colleagues working with ancient DNA – within the AWC, nationally and internationally • students – training graduate students from NZ and internationally • communities who are connected to the samples we are analysing (NZ & worldwide) • Archaeologists and other historic researchers • Museum staff (NZ & worldwide)
Nelson	**Genomics of iconic species** **Team members:** Nicky Nelson, Kristina Ramstad We aim to undertake preliminary sequencing of the tuatara genome, focusing on two areas: (1) characterization of a partial transcriptome for tuatara, and (2) characterization of the Major Histocompatibility Complex (MHC) region. This project will enable us to assess the feasibility of assembling and annotating next-generation sequencing data from an evolutionarily distinct organism. For (1) we have used Illumina sequencing to obtain mRNA sequences from an early stage tuatara embryo, and are currently working on assembling and annotating these sequences to produce an Expressed Sequence Tag database for tuatara. In (2) we are collaborating with Scott Edwards at Harvard University to map the structure of the MHC, a genomic region which plays a central role in the immune system and provides a classic example of adaptive evolution. We have identified clones from a tuatara genomic library that span the core MHC region and are currently assembling sequence from these clones produced by GS-FLX 454 sequencing. Comparison of gene content in the tuatara MHC with that of other vertebrates will provide insight into the evolutionary history of this important genomic region.	
Poole	**Evolution of RNA editing.** **Team members:** Anthony Poole, PhD students (UC): Alicia Lai RNA editing is the process by which information in transcripts is altered prior to translation to generate protein products. It seems particularly prevalent in organelle lineages and endosymbionts. It has been suggested that genetic drift may lead to the fixation of editing sites in a three-stage process: 1. An enzymatic activity that is prerequisite for editing exists. 2. A site that can be edited emerges through mutation. 3. That site is fixed in a population through drift. We previously reported slippage-type editing in *Buchnera*, a bacterial endosymbiont of aphids. *Buchnera* is a close relative of *E. coli*, so we are testing this hypothesis using long-term evolution and genome resequencing in *E. coli*. Our experiment simulates the action of Muller's Ratchet, and we are screening for the emergence of slippage sites, of which thousands may emerge through a single point mutation in the *E. coli* genome.	
Rainey	Evolutionary implications of REPINs and their associated RAYTs **Team members**: Paul Rainey, Frederic Bertels (AWC student) and Xue-Xian Zhange (NZIAS) DNA sequences that copy themselves throughout genomes and make no specific contribution to reproductive success are by definition 'selfish'. Such DNA is a feature of the genomes of all organisms and evident by virtue of its repetitive nature. In bacteria the predominant repetitive sequences are short (~20 bp), extragenic and palindromic. These so-called REP sequences may occur many hundreds of times per genome, but their origins and means of dissemination have been a longstanding mystery. Recent work (with Frederic Bertels) has shown that REPs are components of higher order replicative entities termed REPINs that are themselves derived from	

Investigator	Project	Stakeholders
	the REP sequences that flanked an ancestral autonomous selfish element. In this ancestral state the REP sequences were critical for the movement of the selfish element but were devoid of any capacity to replicate independently. REPINs, on the other hand, have evolved to have a life of their own, albeit one that exploits – even enslaves – a genetic element upon which their existence depends. Current work seeks to understand the evolutionary origins and consequences of REPIN movement through experimental studies.	
Ritchie	**Coastal conservation management: The genetic stock structure of New Zealand snapper and grey mullet** **Team members:** David Ashton, Rachel Wilcox, Balam Jiménez, Mark Morrison (NIWA), and Peter Ritchie. Snapper is New Zealand's largest recreational fishery, and one of the country's largest coastal commercial fisheries. However, after a long period of over fishing stocks have failed to recover. We are using genetic methods to determine the breeding patterns of snapper and what contribution each nursery areas (e.g. Kaipara Harbour) make to the adult snapper populations. When the relatively contributions of these areas has been properly determined, restoration efforts can be focused on the issues that have the largest economic impact on the fishery. Well-informed management will ensure that recreational, customary, and commercial fishing can continue in the future	Ministry of Primary Industries, NZ Fishing Industry, Seafood Industry Council, and Regional Councils
Ritchie	**The genetic stock structure of Hapuku in New Zealand and Australia, and genome sequencing for new aquaculture developments** **Team members:** Henry Lane, Natalie, Jane Symonds (NIWA), Corey and Peter Ritchie Hapuku is an important New Zealand and Australian fishery species, but the specific number of stocks and their boundaries can only be loosely defined. This causes uncertainty when applying management plans and stock assessment models. We will use genetic marker to solving this problem. Our results will offers a real opportunity to maximise the benefits of this fishery resources but better matching management boundaries to biological productivity. The second part of this project is to develop genomics techniques to uncover natural adaptive genetic variation in species such as hapuku (which is the top priority finfish species for new aquaculture) and attempt to use the genetic variation that evolution has been moulding, to fast forward aquaculture developments and selective breeding. The whole genome will be sequenced and high-resolution genetic markers developed to 'capture' wild adaptive variation.	Ministry of Primary Industries, Western Australian Fisheries Department, NZ Fishing Industry, NZ Seafood Industry Council, Aquaculture New Zealand, Ngai Tahu Seafoods
Ross	**Estimation of the levels of confidence at which species of birds found in New Zealand may be identified by DNA Barcoding** **Team members:** Dr Howard Ross, Dr Craig Millar, Selena Patel, Thomas Garden (summer student) Previously (Ross et al 2008 Systematic Biology 57:216-230) the relationship between the amount of genetic variation in a species and the reliability with which it could be identified by genetic methods was determined by simulation. We are applying this relationship to the DNA sequence data being generated as part of the Barcoding of the Birds of New Zealand to gain estimates of the ease with which New Zealand birds can be identified in this manner.	**Primary:** **Conservationists** and others documenting or monitoring biodiversity, because it will enable them to quantify the reliability of species identifications made using DNA-based methods.
Ross	**Re-examination of the frequency of species-level paraphyly in animals** In a highly cited review article, Funk & Omland (2003 Annual Review of Ecology, Evolution, and Systematics 34:397-423) reported a high incidence (23%) of non-monophyly in animal species, on the basis of a review of published studies. There is the potential for this estimate to be biased upwards as a result of the publishing process. I am using the Barcode of Life Database (BOLD) of cytochrome c oxidase subunit I (COI) to reassess this result. These sequences have been collected as a part of the international initiative to document diversity and so develop a species identification tool. Consequently the species representation will be influenced by different biases. It has been argued that high levels of non-monophyly seriously diminish the power of the DNA Barcode Initiative. I hope to also assess the impact of non-monophyly on the efficacy of DNA Barcoding in species identification.	**Primary:** **Evolutionary Biologists**, because it will increase our understanding of how biodiversity is patterned. **Conservationists** and others documenting or monitoring biodiversity, because it will increase our understanding of the reliability of DNA-based methods of species identification.

Theme: 1. Genomics & Biomathematics
Activity: C. Analytic and predictive modeling in ecology and evolution

Investigator	Project	Stakeholders
Cox	**Determining universal rules underpinning the evolution of gene expression in allopolyploid species** **Team members**: Murray Cox, Austen Ganley Allopolyploid species are formed, near instantaneously, via the merger of two very different parent species. The resulting hybrid must maintain its competitiveness despite carrying two non-adapted sets of transcriptomes. This study employs high throughput RNA sequencing to identify, on a genome-wide scale, how expression of the two gene copies from each parent is regulated, with the aim of determining universal rules for this process across the eukaryotes.	• Agricultural sector
Rainey	**Evolution of multicellularity** **Team members**: Paul Rainey and Eric Libby (NZIAS post doc) The evolutionary transition from single cells to multicellularity posses significant theoretical and experimental challenges. Experimental analyses currently underway (with Caroline Rose and Katrin Hammerschimdt) have generated new insights that are currently being generalised via the construction of mathematical models and simulations. Of particular interest are frequency dependent coupling and negative feedback cycles between different developmental stages of organisms and their environmental effects.	
Spencer	I am working on several projects in this activity. An overarching theme is **the incorporation of epigenetic modifications into the standard models of population genetics**. I will be co-leading, with Jason Wolf (University of Bath) a NEScent-sponsored Working Group looking at what predictions different hypotheses make about imprinted genes and what sort of data is needed to test these theories. I continue to work on mathematical models that explain why natural populations of almost every species harbour significant levels of genetic variation. Finally, with collaborators Prof. Graeme Wake and Tony Pleasants, I am investigating mathematical models for optimal waiting and development times for plastic responses.	**Evolutionary Biologists**, because it will improve our theories of how gene expression evolves; **Empirical researchers in biomedical areas**, because it will direct them to experiments that will test evolutionary theories
Steel & Semple	**Modelling in biodiversity conservation and in speciation/extinction processes** **Team members**: Charles Semple and Mike Steel Quantitative methods are widely used in conservation biology for deciding which collection of species should be conserved. AWC research has provided the first mathematically-rigorous method of selection under a model of biodiversity in which species features arise and disappear. In previous research we have shown that selections can be very different under this more realistic model when compared to using the standard model for which features never disappear. We have also applied ecological constraints to the questions of minimizing and predicting the expected loss of phylogenetic diversity in the near future given the current high-rate of species loss. Future work will include analysing the loss of phylogenetic diversity in more realistic speciation-extinction processes.	