



Research development workshops

Thursday 21 November

**Auditoriums 1 and 2,
Administration Building (F23),
University of Sydney**

Workflows for big data

Time: 9.00 am – 10.30 am

Dr Angela McGaughan

Australian National University

This workshop will take participants through common workflows for big data. We will start with experimental design and move through genomic library preparation to common bioinformatic problems. We will finish with a worked example of population genomics in action, where we use code in R to understand the effects of translocation and founder events on the genetic diversity of Australian populations.

For those who want to follow along with the demonstration, the package [PopGenome](#) will need to be installed in R.



Models in genomics and biodiversity analysis

Time: 11.00 am – 12.30 pm

Dr Xia Hua

Australian National University

This workshop will introduce some basic mathematical models in population genomics, phylogenomics, and phylodynamics. Some of these models are relevant to the afternoon workshops, including species distribution models in biodiversity and spatial analysis, multi-species coalescence process and birth-death process in phylogenomics, and regression models in genome-wide association studies. The workshop will focus on the mathematical foundations of these models so that participants will have a rough idea on how to develop a mathematical model for their own studies. No software is required for this workshop. Participants are assumed to have a basic understanding of calculus, probability and statistics.



Biodiversity and spatial analysis

Time: 1.30 pm – 3.00 pm

Dr Dan Rosauer

Australian National University

This workshop will introduce concepts and methods for mapping spatial patterns of biological diversity. It will consider the meaning, measurement and use of diversity metrics such as phylogenetic diversity and endemism. Through prepared examples, participants will learn how to link spatial and phylogenetic data for a group of species to both visualise and measure the spatial component of evolution, for example to identify areas with distinct or unique biota. **The software [Biodiverse](#) will be used for this** familiar with interpreting phylogenetic trees but do not need prior experience with spatial information.

Phylogenomics

Time: 1.30 pm – 3.00 pm

Dr David Duchêne
Australian National University
Professor Simon Ho
University of Sydney

This workshop will present an introduction to the challenges associated with phylogenetic analyses of genome-scale data. Participants will learn about data filtering, gene tree incongruence and inferring species trees. A hands-on session will involve the analysis of a multigene dataset to resolve the phylogenetic relationships among marsupial families.

The software required for this workshop are [MEGA](#) and [ASTRAL](#).

Participants are assumed to be familiar with interpreting phylogenetic trees and some of the basic principles of molecular evolution.

Metagenomics and pathogen discovery

Time: 3.30 pm – 5.00 pm

Dr Jackie Mahar and Dr Mang Shi
University of Sydney
Dr Jemma Geoghegan
Macquarie University

This workshop will provide training on how to find potential pathogens in meta-transcriptomic data. This will include guidelines for (i) *de novo* assembly of meta-transcriptomic data, (ii) taxonomic assignment and annotation of assembled transcripts/genomes and (iii) comprehensive characterisation of potential pathogens, including identification of false positives.

Participants will be guided through a bioinformatics pipeline using a reduced dataset on a high-performance computer with the required software and databases already installed. **The software required for this workshop are [PuTTY](#) (if working from a Windows machine) and an FTP**

client such as [FileZilla](#). Participants are assumed to have a basic understanding of command line.

Genome-wide association studies

Time: 3.30 pm – 5.00 pm

Associate Professor Shyam Gopalakrishnan
Assistant Professor Ida Moltke
University of Copenhagen

This workshop will give a brief, basic introduction to genome-wide association studies (GWAS). This will include hands-on experience with running a GWAS and plotting the results, and will cover some of the potential caveats and related important quality control and data filtering steps.

We will be working on a server with all necessary programs installed. The only software required for this workshop is a program that allows you to log in into this server using SSH and open figure files remotely. **For Windows users, this requires that you install [MobaXterm](#). Mac users can use Terminal, but please make sure you also have [XQuartz](#) installed (default on some Macs, but not all, and it makes viewing of figure files via SSH a lot easier).** Linux users should not need to install anything. We assume very little prior knowledge except for a basic understanding of what it means to test for association between a specific genetic variant and a given phenotype. It will be helpful but not essential to have some knowledge about what statistical tests can be used to test for such an association.

