

Course Description

Introduction to Unix for Next Generation Sequencing

Objective

The new generation of high throughput DNA sequencers (often denoted NGS) are able to output genomic sequence data in ever increasing speeds and quantities. Speedily processing and analysing this workload requires a robust base system and powerful techniques. The command-line environment of Unix/Linux has proved capable at handling this challenge, as long as the user has the associated competence. The aim of this course is to deliver such skills to the participants.

Location and Date:

Swallowgate Computer Classroom, 9am to 1pm, 20th April 2017

Intended Audience

Postgraduates, postdocs and staff of the University of St Andrews who are currently, or will soon be, conducting research using Next generation Sequencing datasets.

Prerequisites:

This is a course for beginners, so no prior Unix/Linux command-line experience is assumed. Participants should however have some basic experience of DNA data handling, maybe obtained in a Windows or MacOSX environment. Participants must be embarking, or currently working, on research projects where genomic data will be handled.

Content

The chief focus of the course will command-line handling of NGS files. There will be one short introductory session mentioning some theoretical and orientative aspects and three hour-long practical sessions starting with some simple exercises and thereafter increasing in ambition.

Outcomes

At the end of the course, participants will feel at home with the Unix command line and be able to list out, navigate, view and organise large and multiple datasets efficiently. They will gain familiarity with how Unix solves some common bioinformatics problems and take away some useful skills which they can practice and perfect afterwards.

Format of course

This is a half-day course, made up of three main practical sessions. Laptops are not required as each participant will have a terminal at their disposal. There will be one coffee break.

Instructors

Joseph Ward and Ramon Fallon of the St Andrews Bioinformatics Unit. There may also be some potential assistance for the practicals from colleagues from the Scottish Oceans Institute.

Detailed Schema:

Time	Type	Target skills	Contents
08:50 – 09:00	Meet	Registration	
09:00 – 09:30	Introduction	Understand	* challenges of NGS datasets * usefulness of Unix/Linux for today's NGS data * connecting to remote machines

			<ul style="list-style-type: none"> * orientation to the command-line * Sharing, permissions.
09:30-10:30	Practical No. 1	Access/Navigate	<ul style="list-style-type: none"> * Getting help * moving around directories/files * naming, copying, linking * \$HOME, pwd; cd; ls, tab-completion * Executables, \$PATH, history.
10:30-11:00	Coffee Break		
11:00-12:00	Practical No. 2	View/Organize	<ul style="list-style-type: none"> * echo, cat, more, less, grep, vi/vim * wc, head, tail, sed for large files * wildcards for multiple files
12:00-13:00	Practical No. 3	Extract/Parse	<ul style="list-style-type: none"> * grep, for tr; cut; sed. * wild cards and simple regex * pipes, awk * processes and screen