





# Course Description Introduction to Unix for Bioinformatics and Genomic Analysis

# Objective

Next generation sequencers (often denoted NGS) produce ever-increasing amounts of genomic sequence data. Speedily processing and analysing these data requires a flexible and robust computing environment and powerful techniques. The computer operating system Unix, and its open source equivalent, Linux, together with their related tools and command-line environment have proved capable at handling this challenge. The aims of this course are to familiarized users with the Unix operative environment, and provide participants with basis skills to handle and manipulate NGS data sets.

## Organisational details:

The course shall be held at the Swallowgate Computer Classroom, 6th October 2016 and is organised by the St Andrews Bioinformatics Unit (StABU) in collaboration with the Scottish Oceans Institute with partial funding from CAPOD.

## **Intended Audience**

Research postgraduates intending to work with NGS data or large genomic datasets

#### **Prerequisites:**

Participants are expected to be embarking on research projects where genomic data will be handled and processed. No experience of command-line usage is needed.

#### Content

The chief focus of the course will command-line handling of genomic data via Unix terminals and the advantages this holds over Graphical User interface processing. The course is made up of four sub-modules whose themes are 1) "Navigate" 2) "View" 3) "Modify" 4) "Do", each of which consist of an initial explanatory/theoretical half of 45 minutes, followed by a practical half where participants complete exercises covering the theme's topics.

#### Format of course

This is a one-day course, made up of theoretical and practical sessions. Laptops are not required, as each participant will have a computing terminal at their disposal. There will be one lunch break and two coffee breaks.

## **Registration and instructors:**

Numbers are limited so we recommend early registration at the following site: <u>http://bioinformatics.st-andrews.ac.uk/course-registration</u>

The instructors shall be:

- Ramon Fallon from the St Andrews Bioinformatics Unit
- Daniel Garcia de la Serrana Castillo, Scottish Oceans Institute
- Luke Holman, Scottish Oceans Institute
- Christopher Hollenbeck, Scottish Oceans Institute

#### Schedule:

Please see next page.

Time	Туре	Theme	Contents
08:30 - 09:00	Meet	Registration	
09:00 – 09:45	Explanatory	"Access/Navigate"	<ul> <li>* challenges of genomic big data</li> <li>* introduction to Unix</li> <li>* orientation within Unix</li> <li>* throughput / performance issues</li> <li>* command-line vs. GUI</li> <li>* local vs. remote</li> </ul>
09:45-10:30	Practical		* connecting / finding one's way around * \$HOME, pwd; cd; ls * \$PATH/directories/variables/find
10:30-11:00	Coffee Break		
11:00-11:45	Explanatory	"View/See"	* directory / folder structures * file types; conversions * large file challenges * compression * extraction
11:45-12:30	Practical		* echo; cat; head; gzip; wc; * more; less; man * pipe; grep; for; tr; cut; sed.
12:30-13:30	Lunch Break		
13:30-14:15	Explanatory	"Edit/Modify"	* Editors * Regular expressions
14:15-15:00	Practical		Dos2unix; vi/sed/replace/regular expressions
15:00 - 15:30	Coffee Break		
15:30 – 16:15	Explanatory	"Do/process"	Tabular Data /File formats
16:15-17:00	Practical		* Awk; advanced options; * background processes; screen * Parallelism
17:00-17:10	Wrap-up		