# MSc phylogenetics practical

In this practical we will carry out some simple analyses that use phylogenies (evolutionary relationships inferred from genetic sequences) to answer some interesting biological questions. You’ll be using Linux because 90% of bioinformatics servers also do, and running it on the Raspberry Pi because we think they’re cool :-)

Throughout the practical, commands you should type into the terminal window will be rendered in this font, e.g.:

MacBook-Pro:pi\_biolinux\_cluster joe$ echo 'hello, world!'

hello, world!

Some of the datasets in this practical have been simulated so that no two workstations will give identical answers to every question. If you have questions during the practical, ask your instructors. They are:

Joe Parker

Jim Clarkson

Alex Papadopulos

## Part 0 – Getting started (10-15mins)

*# Yes, most things in computing are numbered from 0… n-1, not 1… n. Get used to it!*

You should already have read [“How to read a phylogenetic tree”](http://epidemic.bio.ed.ac.uk/how_to_read_a_phylogeny) by Andrew Rambaut, and “[Basic Shell Commands](http://www.hongkiat.com/blog/basic-shell-commands-for-bloggers/)” (which you can run interactively on [Tutorialspoint](http://www.tutorialspoint.com/execute_bash_online.php)). Let’s get started:

1. Your desktop should have a folder visible called ‘phylogenetics’. This is where your data will live, but we’ll be using a window that connects directly with the operating system (a ‘terminal’) which uses the bash language. Open a terminal now, by holding the Ctrl-Alt-T keys together.
2. You should now have a terminal window open. The black screen you can see looks like it dates from the 1970s because it does! In fact, pretty much all computers, deep down, date from the 1970s, because not much has changed at that level since then.
3. The line you can see with a blinking cursor at the end is called a *prompt.* It is split into three parts, telling us:

**User@user-pi:~$** (username @ computer name: directory$)

1. “~” is a funny name for a directory, which you’ve probably not seen before. We can use the ‘pwd’ command (‘**p**rint **w**orking **d**irectory’) to find out where we really are. Type that now, followed by a return (you always type return to enter your input).
2. We want to move to the directory where the data is, that ‘phylogenetics’ folder you can see on the desktop. To change directories in Linux, Unix, or Mac terminals, we use the ‘cd’ command, followed by the name of the directory we want to go to. These names must be specified relative to the current directory, so if we’re currently in a directory called /home/trouble/ahead and we want to move into /home/trouble/ahead/for/starters , we need to move using the command:

* cd for/starters not
* cd starters

as the computer can’t find the ‘starters’ directory, without first being told to look in the ‘for’ directory.

1. To go up a directory we can use the special file ‘..’ e.g. cd ‘..’ will take us up to the parent directory. Type these commands, looking at the output

* cd Desktop/phylogenetics
* pwd
* cd ..
* pwd
* cd Desktop/phylogenomics/part\_1

… we’re now ready for the practical.

## Part 1 – Aligning sequences, building trees, and looking for error (30mins)

*# Unfunny pithy remark here, spot the pattern?*

You should already have read

* Align [muscle] a matK dataset with a random human one in
* Make a tree [raxml] and spot the outlier
* Rebuild the tree and note it is betterer

## Part 2a – Phylogeography of island chains (30mins)

*# Commenting code liberally is good. Comments in Linux typically start with a ‘#’*

* There are two datasets from two genera of flowering plants, both distributed on the hawaian islands. One is wind-dispered, the other is dispersed by mammals. We will investigate their districutopns and look for a phylogeographic pattern.
* One dataset is drawn from the tree
* One is shuffled (4≤*n*<10 times)
* Make a tree of each
* Colour in [figTree]
* Compare patterns
* There are two datasets from two genera of flowering plants, both distributed
* How can we be sure?

## Part 2b – Rates of evolution of island chains (30mins)

*# Commenting code liberally is good. Comments in Linux typically start with a ‘#’*

* [BEAST] tells us about demography
* Pre-run BEAST chain in Tracer – show them the fuzzy worms
* BEAST tree in figtree, show root 95% hpd
* BaTS tells us about tip-trait association. Markup for one species
* Open for another and run
* Which is more associated? Significantly?

## Part 3 – Speciation in tropical rainforests (30mins)

*# Most code comments debase into wisecracks of dubious value soon enough…*

You should already have read

### Glossary of useful terms

Poo

Bum

Wee

Phylogeny

Foo barr

Clade

Tip

Taxon

Branch

Node

MCMC

Chain

Likelihood

Alignment

posterior