

Molecular Dating Practical

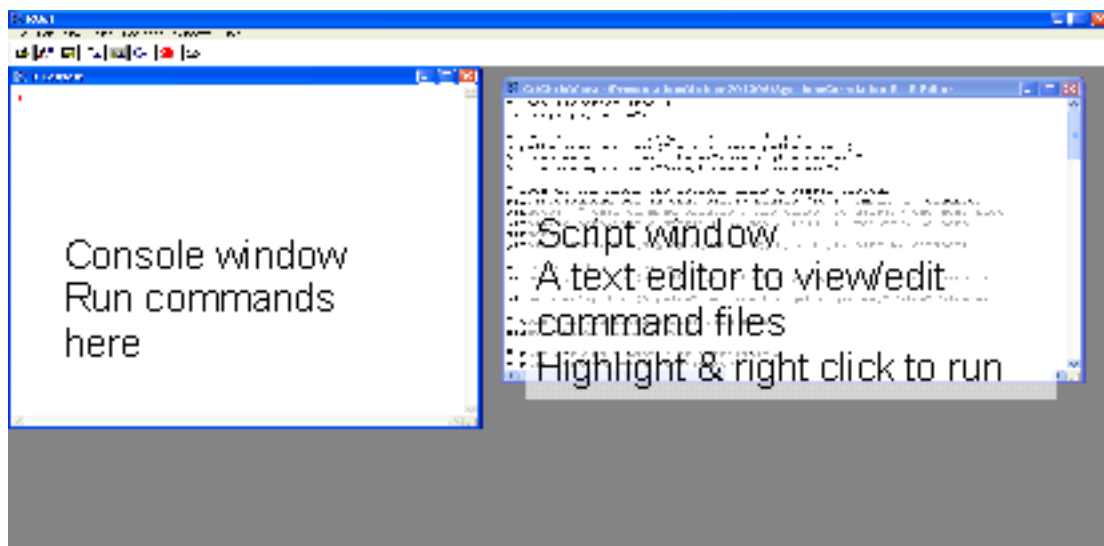


R is a free software environment for statistical computing and graphics (<http://www.r-project.org/>). R is primarily a command driven environment, and we will be using command scripts to run our analyses. R has multiple package libraries, developed by many different people for particular applications. We require the phylogenetics package “ape”, the niche modelling package “adehabitat” and the phyloclimatic modelling package “phyloclim”. R packages can be installed using the `install.packages` command. Type the following commands into an R console window to install these packages.

```
install.packages(c("phyloclim", "ape", "adehabitat",
"SDMTools"))
```

A quick note on using R

This is not a tutorial about R, but we are doing some examples with R. When you open R you will see a command window, you can type instructions here to run analyses. Generally it is a good idea to store your commands in a text file (called a script file and given a .R extension). Use File-Open to open script files. This creates a script window (essentially just a text editor). Run lines of code by right-clicking then and selecting run-line-section. You must run each line in order as later commands may depend on earlier ones. Any line starting with # is a comment, please read the comments as they describe what is going on. **To run the example script files you must set the current directory to the folder with our test files. Use File-Change dir and select the directory with the R scripts.**



The practicals using R will involve opening script files and running one line at a time.

Molecular Dating Practical

Data can be downloaded from via the link to the phyloclimatic modelling workshop at <http://www.zsl.org/chrisyesson/>. You will need MolecularDatingPracticalData.zip

Open R and change the R directory using File- Change dir to select the location of your unzipped data

```
## load in the library for manipulating phylogenetic trees
library(ape)
```

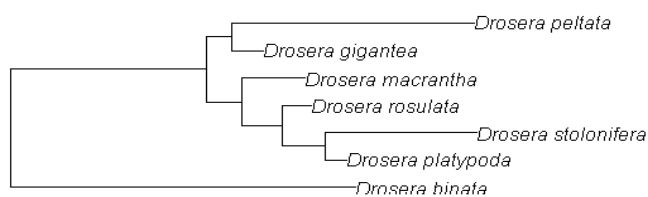
```
## load a tree
tr<-read.nexus("Data/YessonCulham2006Subtree.tre")
## view the tree using the plot command
plot(tr)
```

```
## simple Penalised Likelihood can be done with the chronopl command
## age.min signifies the age to assign to the root, in this case 20Ma
## lambda is a smoothing parameter that scales the rate variation
tr.pl<-chronopl(tr,age.min=20,lambda=1)
```

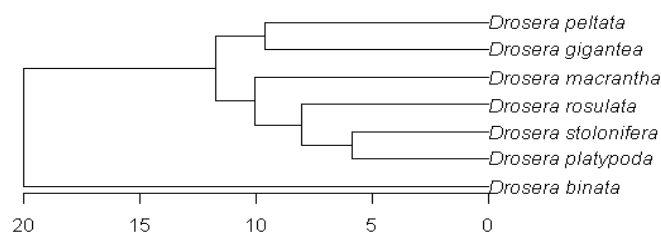
```
## show both trees on the same plot
layout(matrix(1:2, 2, 1, byrow = TRUE))
plot(tr)
title("The original tree")
plot(tr.pl)
axisPhylo()
title("Penalised Likelihood tree")
```

Ref Sanderson, MJ (1997) *Molecular Biology and Evolution*, 14, 1218-1231.

The original tree



Penalised Likelihood tree



Dating using mean path lengths

other rate smoothing can be done in R too. MPL rate smoothing has been applied successfully to large trees. MPL provides explicit tests of deviation from a clock at each node as well as standard errors on the ages

perform an MPL rate smoothing

```
tr.mpl<- chronompl(tr)
```

find the root length by taking the maximum node depth

```
tr.mpl.root.length<-max(branching.times(tr.mpl))
```

scale root to 20Ma

```
tr.mpl.root.calibration<-20/tr.mpl.root.length
```

calibrate the tree to set the root to age 20Ma

```
tr.mpl$edge.length <- tr.mpl$edge.length * tr.mpl.root.calibration
```

plot a tree with the results

```
layout(matrix(1:4, 2, 2, byrow = TRUE))
```

```
plot(tr)
```

```
title("The original tree")
```

```
plot(tr.mpl)
```

```
axisPhylo()
```

```
title("The dated MPL tree")
```

```
plot(tr)
```

```
node.labels(round(attr(tr.mpl, "stderr"), 3))
```

```
title("The standard-errors")
```

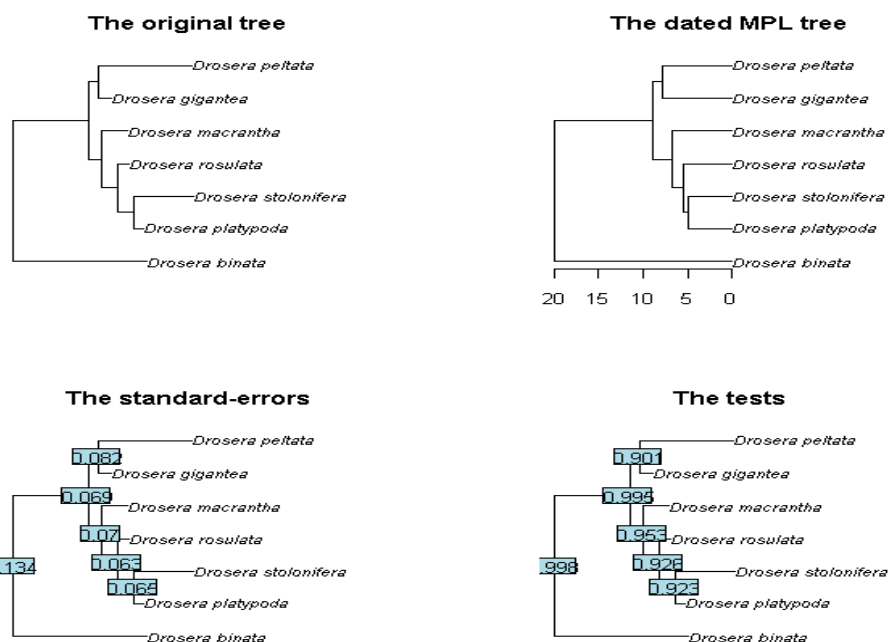
```
plot(tr)
```

```
node.labels(round(attr(tr.mpl, "Pval"), 3))
```

```
title("The tests")
```

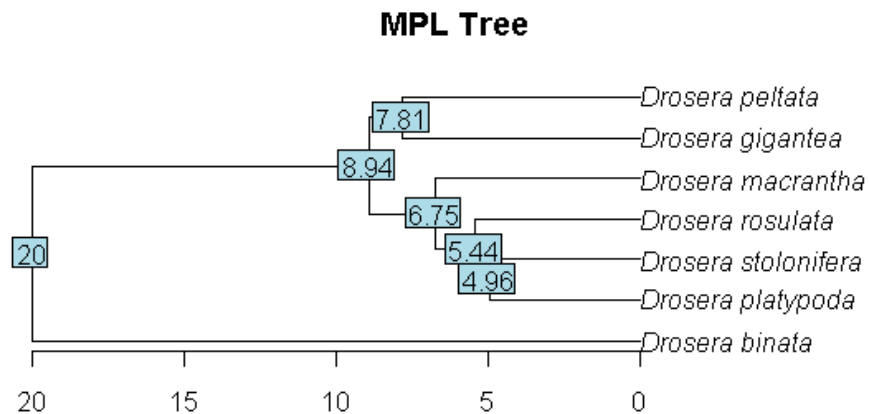
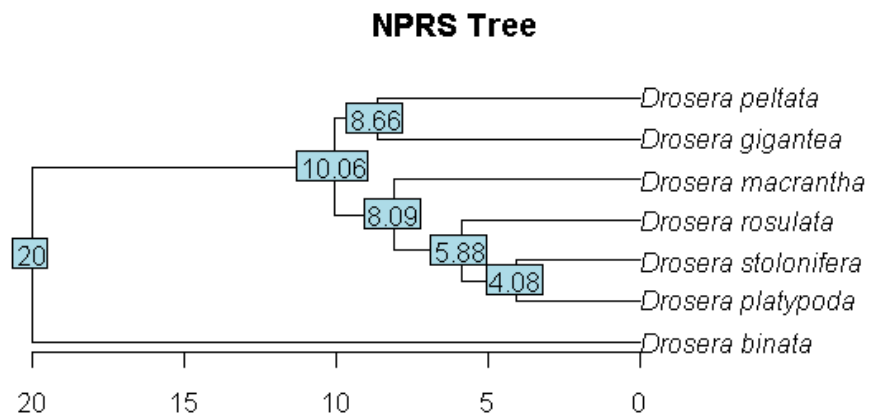
this is adapted from the example given in the chronompl help page

Ref: Britton, T, et al.(2002) Molecular Phylogenetics and Evolution, 24:58-65



different methods produce different dates

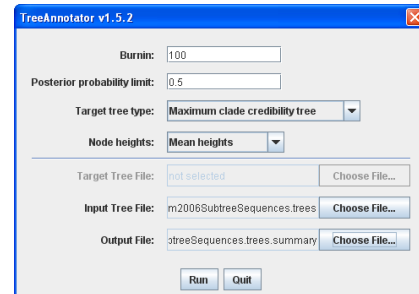
```
## display both trees side-by-side
layout(matrix(1:2, 2, 1, byrow = TRUE))
plot(tr.pl)
title("Penalised Likelihood Tree")
nodelabels(round(branching.times(tr.pl),2))
axisPhylo()
plot(tr.mpl)
title("MPL Tree")
nodelabels(round(branching.times(tr.mpl),2))
axisPhylo()
```



Combining multiple trees for confidence intervals

Once you have your tree file containing multiple ultrametric trees you can calculate confidence intervals using Treeannotator - part of the BEAST software package.

- Open Treeannotator
- Select the input as the tree file
YessonCulham2006SubtreeSequences.trees
- Select an appropriate output tree file



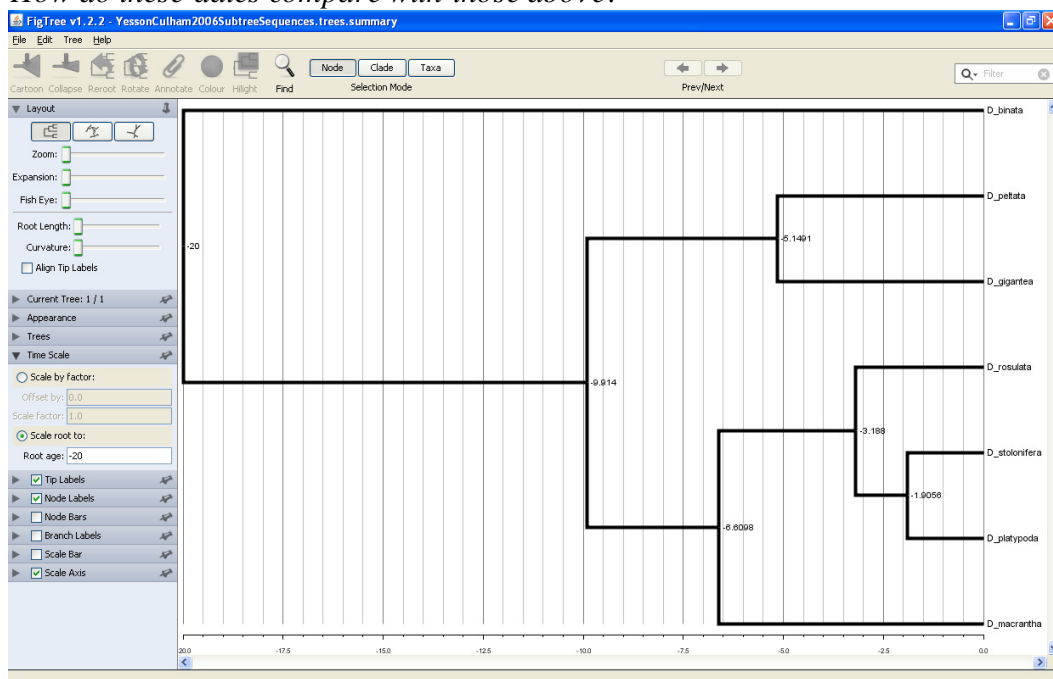
YessonCulham2006SubtreeSequences.trees.summary

- Set the burn-in (number of trees to discard) as 10 and click run
- This produces a tree file with confidence intervals encoded as node-labels

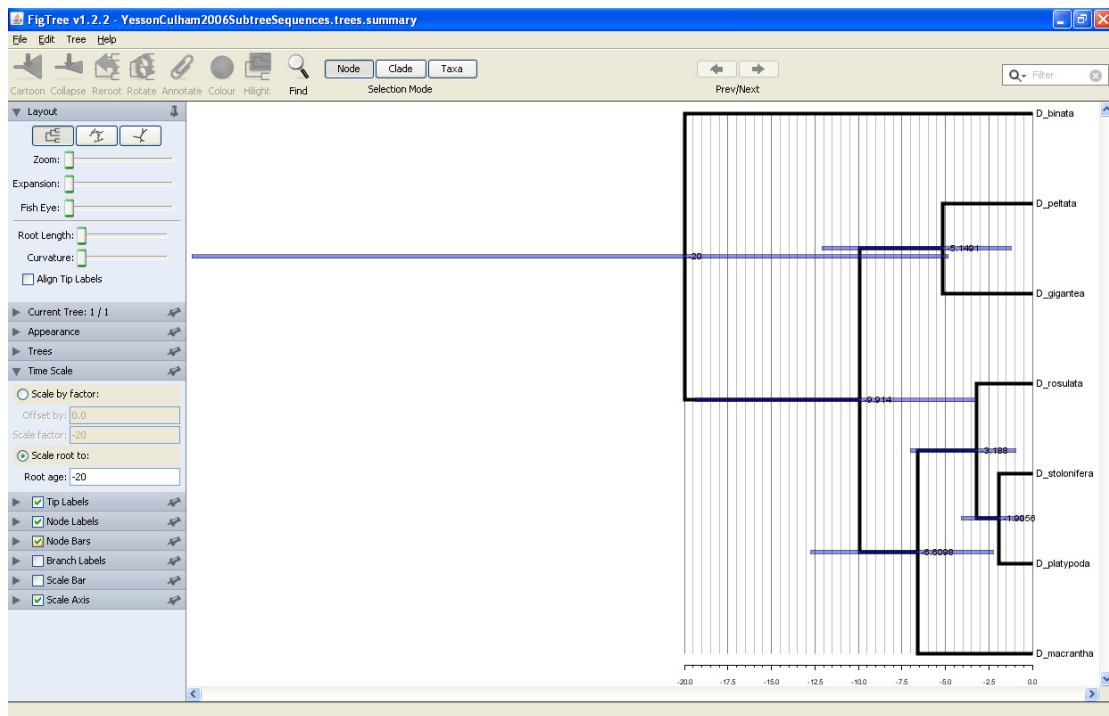
Using Figtree to view the summary tree. The developers have Beast also developed a treeviewing package called Figtree for displaying chronograms. (Download the zipped executable from <http://tree.bio.ed.ac.uk/software/figtree/>)

- Open figtree
- Open the summary tree output from treeannotator
- Click the scale axis box to view a time axis on your tree.
- Click node-labels to see the node ages
- As the tree is not calibrated we can scale the root to our desired calibration (20Ma) using the time scale option. Type 20 into the scale root to box.

How do these dates compare with those above?



- Display node bars using the node bars box.
- Note the larger confidence intervals as you move deeper into the tree



Note: This is a quick and very simple example and does not cover many of the important features of tree searching or rate smoothing functionality. It is meant as a simple introduction to the Beast package. Please look at the Beast website for further tutorials and details (http://beast.bio.ed.ac.uk/Main_Page)

This is the end of the practical – there is an example below on using beast but this is to take away and do in your own time.

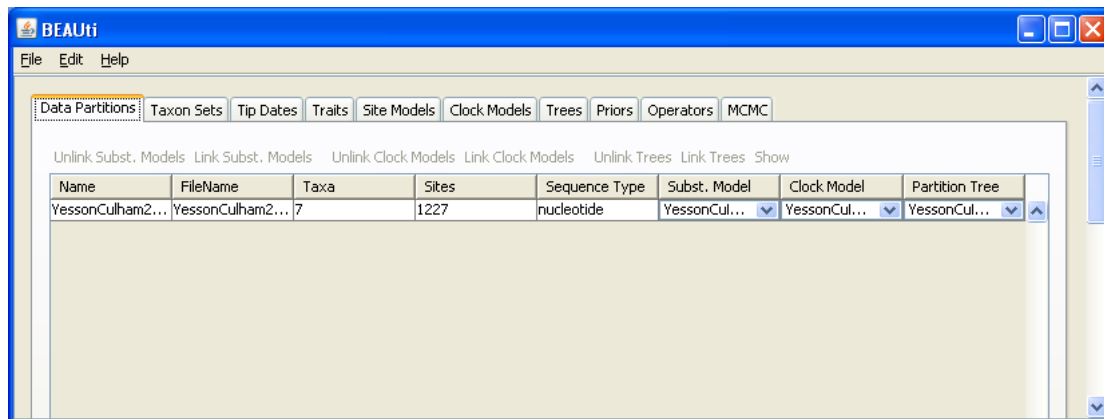
Simple analysis with BEAST

*Note: this practical assumes some knowledge of phylogenetic analysis using Bayesian techniques. You can skip this and move straight to the section **Combining multiple trees for confidence intervals**.*

BEAST performs tree searches incorporating models for rate smoothing using Bayesian analysis. If you don't have the BEAST software installed, download it from http://beast.bio.ed.ac.uk/Main_Page#Downloads or I can provide it on a data stick. All the associated programs are downloaded as zipped executable files, simple unzip and double-click the appropriate .exe file to run the program.

First we must prepare the analysis file using a nexus file of DNA sequences. We do this with the package Beauti (part of the beast package).

- Open Beauti
- File-Import alignment – select the file
“YessonCulham2006SubtreeSequences.nex”



Set the model of sequence evolution in the site model tab. For now we will keep the defaults.

Setting the clock model

- Select the clock model tab
- Select the relaxed clock lognormal model for rate smoothing

Setting the MCMC parameters

- Select the MCMC tab
- A typical analysis will run 10 million replicates, but we don't have time. Type 100000 in the length of chain box
- Finally click “Generate Beast file” in the bottom right corner. Save the file as “TestBeastAnalysis.xml”

This simple model will produce a set of uncalibrated ultrametric trees. Loading constraints and calibrations is the next step, but beyond the scope of this practical. See the Beast website for more details (http://beast.bio.ed.ac.uk/Main_Page)

BEAUti

File Edit Help

Data Partitions Taxon Sets Tip Dates Traits Site Models Clock Models Trees Priors Operators MCMC

Length of chain: 10000000

Echo state to screen every: 1000

Log parameters every: 100

File name stem: YessonCulham2006SubtreeSequences

☒ Add .txt suffix

Log file name: YessonCulham2006SubtreeSequences.log.txt

Trees file name: YessonCulham2006SubtreeSequences.trees.txt

☐ Create tree log file with branch length in substitutions:

Substitutions trees file name:

☐ Sample from prior only - create empty alignment

Data: 7 taxa, 1 partition; Fix clock rate to 1.0

Generate BEAST File...

Running the analysis

- Open BEAST
- Select the saved file “TestBeastAnalysis.xml” and run the analysis

Examine the output files with Tracer (<http://tree.bio.ed.ac.uk/software/tracer/>)

- Open Tracer
- File-Import trace file – select the file from your analysis YessonCulham2006SubtreeSequences.log.txt
- Examine each of the parameters by selecting them in the window on the left.
- View histograms or line charts by selecting the tabs on the right
- Use these views to select an appropriate burn-in period

