**Practical 1d: Assessing temporal structure in TempEst**

**Data set**

* Maximum likelihood tree inferred in PhyML:

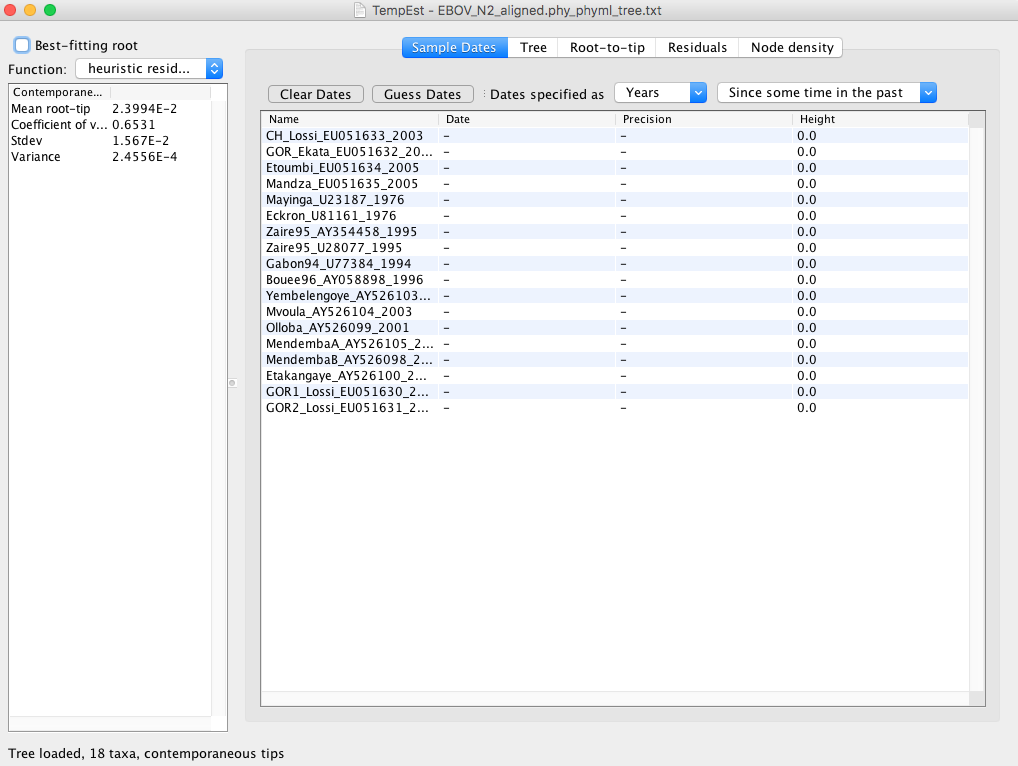
EBOV\_N2\_aligned\_early\_samples.phy\_phyml\_tree.txt

**Software**

* TempEst

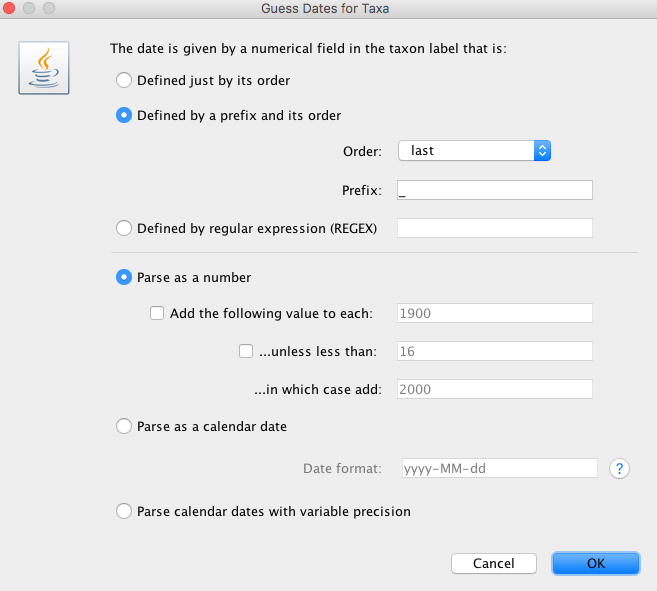
TempEst is a tool for visual inspection of temporal structure in tip-dated phylogenetic analyses. It is available for OSX, Windows, and Linux machines. However, some of the most recent version did not work under OSX, but you can download the Linux version and work with the .jar file.

Double click on TempEst.jar. This will open a window asking you for an input file. In our case, we will use the tree estimated in PhyML. Find EBOV\_N2\_aligned\_early\_samples.phy\_phyml\_tree.txt and click *Open*. TempEst will then open. The interface is a window with several tabs; *Sample Dates*, *Tree*, *Root-to-tip*, *Residuals*, and *Node density* (Fig 1).



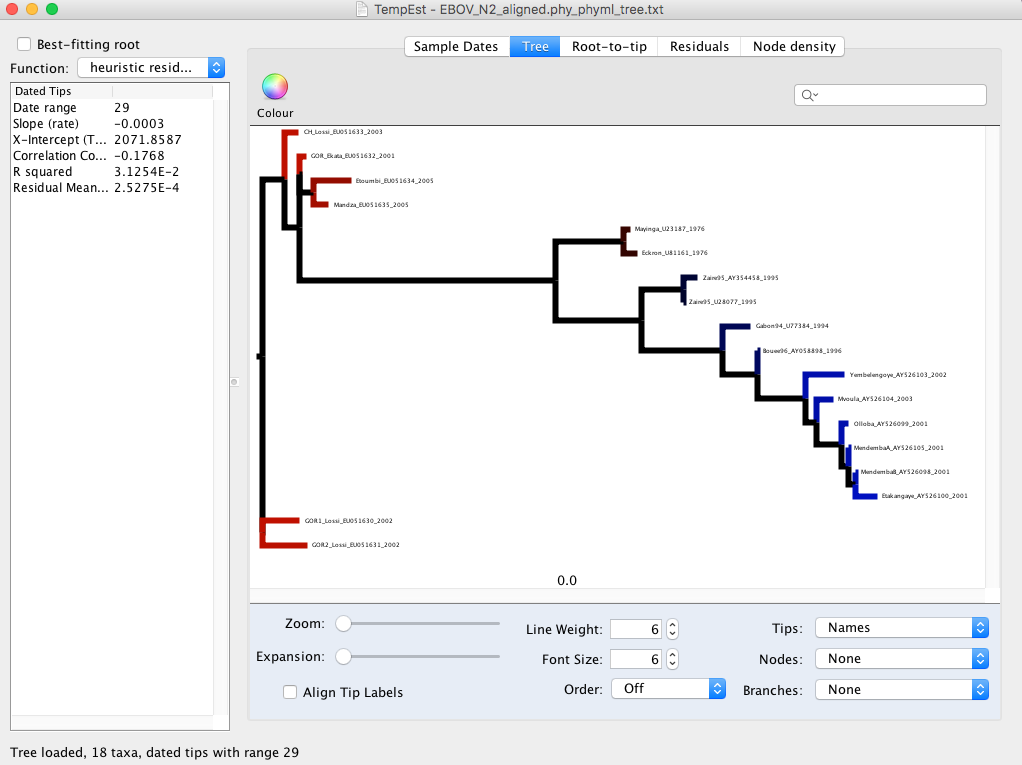
**Fig 1.** TempEst interface with tree loaded.

Note that at the bottom of the window it says that there is a tree loaded, with 18 taxa and with contemporaneous tips. We first need to tell TempEst that the tips of our tree have dates. They are specified as year of sampling after the underscore in each of the taxon names. To set this in TempEst click on *Guess Dates*. This will open a tool to extract the sampling times for different taxon names and formats. Set it up as shown in Fig 2 and click OK.



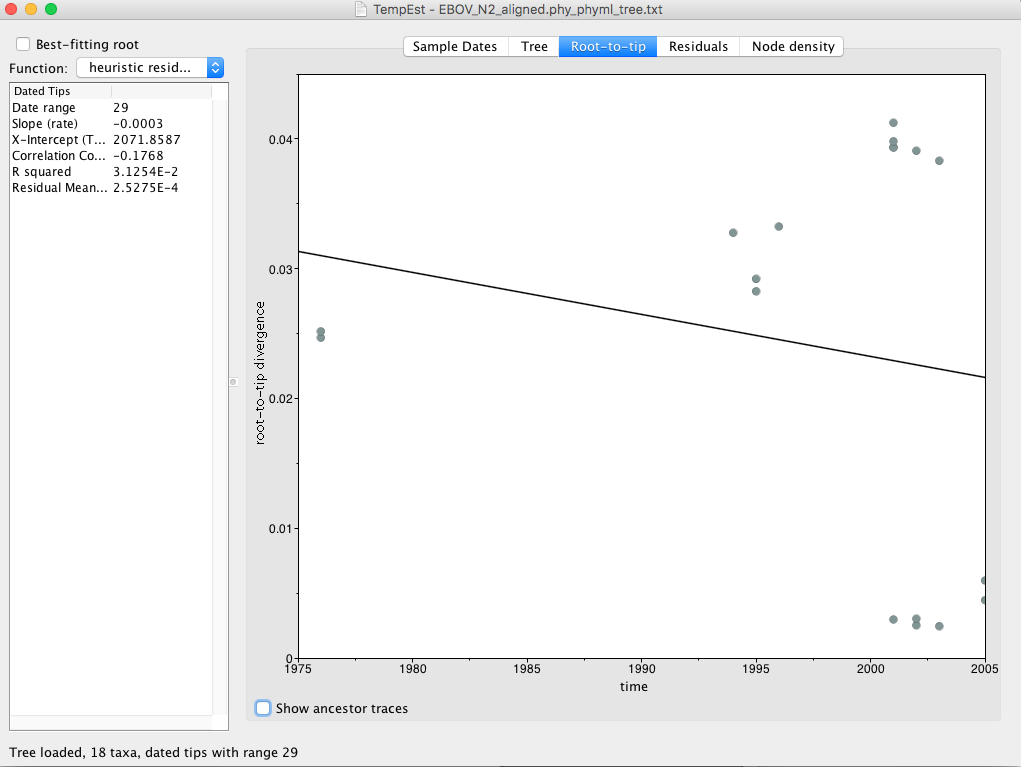
**Fig 2.** Set up to extract sampling times from taxon names.

The TempEst window will update with a list of the taxa and their corresponding sampling times. Check that everything looks correct. It is also possible to edit any dates that were not extracted correctly and to set the uncertainty (precision) around them. Click on the *Tree* tab to inspect the tree. Note that terminal branches have different colours. Increase the line weight if you cannot see them clearly (Fig 3).



**Fig 3.** Tree display in TempEst.

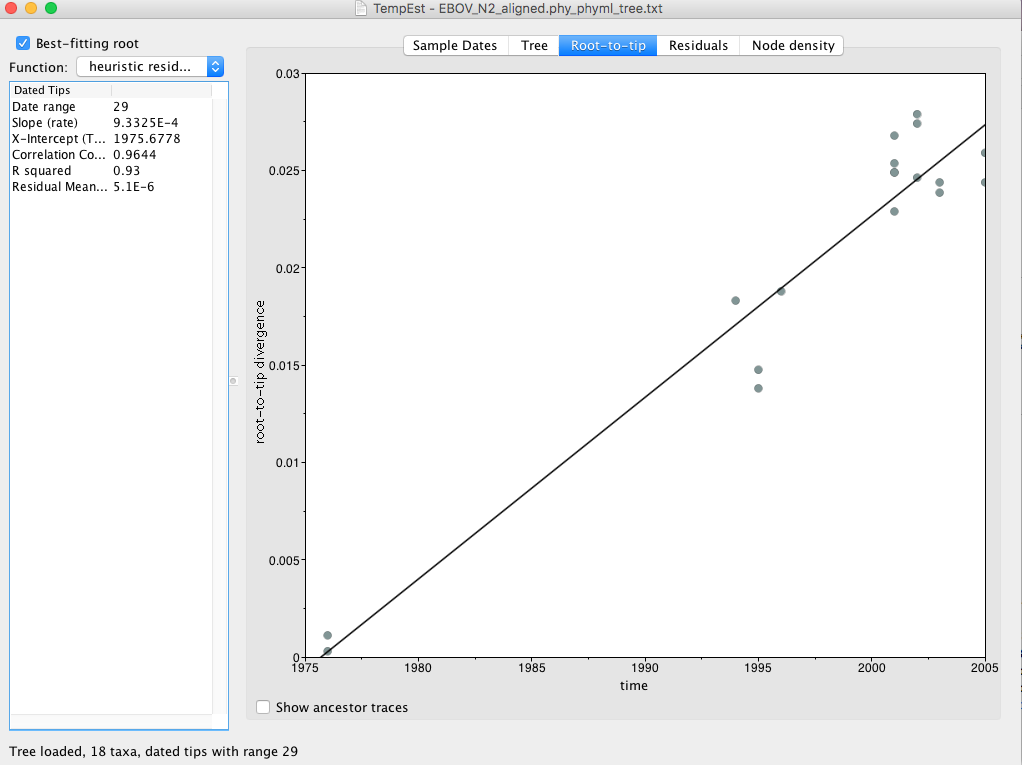
Branches in blue and red have larger and smaller root-to-tip distances than expected, respectively. Click on the *Root-to-tip* tab to see the regression (Fig 4). Also, notice that the regression statistics are printed on the left panel.



**Fig 4.** Root-to-tip regression in TempEst.

**Question 4.1:** Do the regression statistics indicate that there is temporal structure in these data?

Tick the box at the top left that says *Best-fitting root* (Fig 5). Try setting different options in the *Function* drop-down menu. These are used to optimise the position of the root of the tree. For example, the R-squared function finds the position of the root of the tree that maximises R-squared.

**Fig 5.** Root-to-tip regression in TempEst using the best-fitting root.

**Question 4.2:** What does the x-intercept represent? Do you think this is a reasonable estimate for EBOV?

**Question 4.3:** Rate estimates for EBOV have been reported to range from 1×10-3 to 2×10-3 subs/site/year. Is the estimate here reasonable? Can we consider these data set to have sufficient temporal structure?

**Question 4.4:** According to the R2 would you say that these data follow a strict molecular clock?

Go back to the *Tree* tab and try again different tree rooting functions. Notice how the position of the root appears to be robust to the different functions. Select all samples from non-humans (GOR1\_Lossi\_EU051630\_2002, GOR2\_Lossi\_EU051631\_2002, GOR\_Ekata\_EU051632\_2001, and CH\_Lossi\_EU051633\_2003) (you might need to click on CTRL or COMMAND to select these only). Go to the Root-to-tip tab. The non-human samples will be highlighted in blue. Do they appear to evolve slower, higher, or at a similar rate, compared to human samples?

Click on the *Residuals* tab. These can be interpreted as those from an ordinary regression. Finally, select the *Node density* tab. The points correspond to the tips and nodes in the tree.

**Question 4.5:** Clearly, the position of the root of the tree is important to determine whether there is temporal structure in the data. Can you think of a method that is not based on optimising the regression to select the root of the tree?

**Optional exercise**

For the practical session tomorrow, we will use an other data set of flu sequences from the 2009 H1N1 pandemic. The aligned sequences are in NorthAm.Nov.phy. Obtain a phylogenetic tree in PhyML as we did yesterday, and use tempest to determine whether there is temporal structure in these data.

**Question 4.6:** Do the H1N1 data appear to have temporal structure, what is their evolutionary rate and time of origin? Does this match the observation that this outbreak started between late 2008 to early 2009?