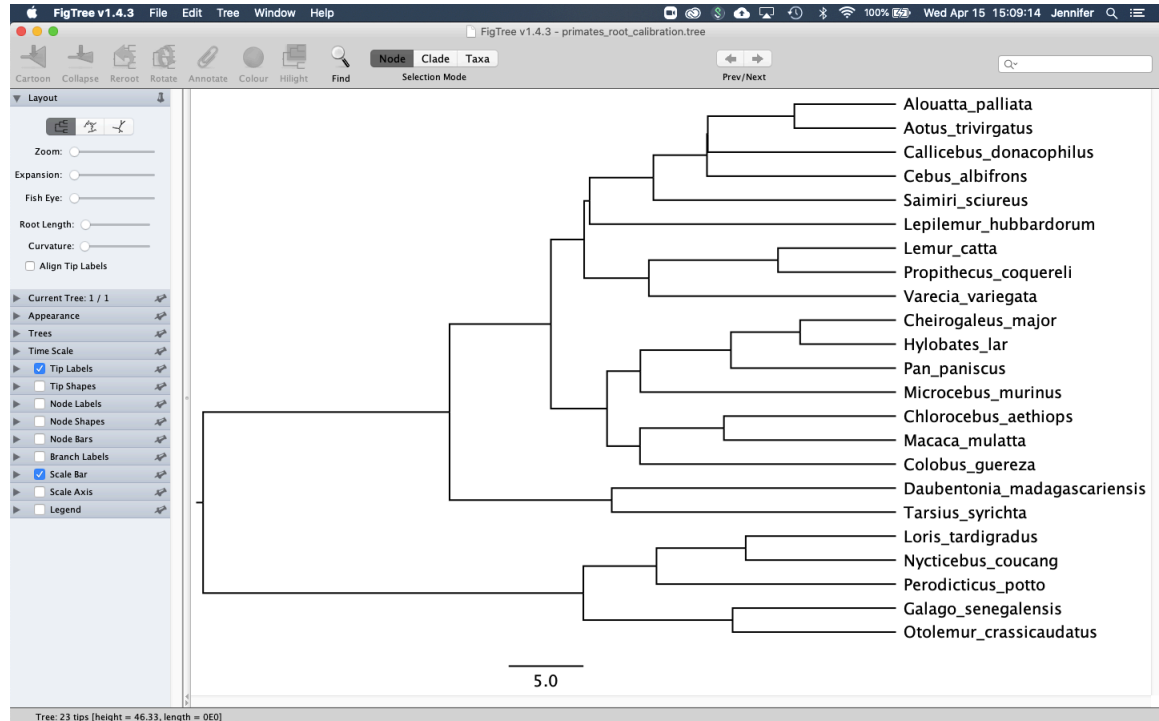


## Lab 12

Jennifer Lin

### Question 1

Send me a PDF or screenshot of your MAP tree (you can open it in FigTree).



### Question 2

Take a look at your log file in tracer. Does it look like your analysis has converged?

No, it looks like my analysis has not converged. First, the ESS of all the parameters are all quite low (less than 200). Second, when I open the Trace tab, most of the traces of the parameters do not look like a typical caterpillar-shape. Third, when I open the Marginal Density tab, the density curves of most of the parameters seems to have multiple peaks. This indicates that the possible values of the parameters have not converged.

I think there may be two reasons that make our result not converged. First, we used several gene trees in order to build the species tree. It is very hard to let all the parameters from different gene trees converge. Second, the other reason may be that we only ran 1000 generations mcmc. If we increase the generations we run, probably it can reach a better status where parameters all have converged.

### Question 3

Name a couple assumptions we made that seem unrealistic to you? According to your experience with RevBayes, do they seem to be easy to solve?

In this lab, we use constant-rate sampled-birth-death model. However, I think the birth and death rate may not be constant overtime. Consequently, using a model that allows different rates overtime would be an alternative choice. For example, we could use [Environmental-dependent Speciation & Extinction Rates Model], which estimating correlation between diversification rates and environmental characters.

We may don't have a perfect taxon sampling. Hence, we can use [Diversification Rate Estimation with Missing Taxa], which helps us to estimate diversification rates with incomplete taxon sampling.

### Question 4

Concatenation is still used much more frequently than the full multispecies coalescent. Summarize in a sentence or two the pros and cons of using concatenation versus the multispecies coalescent.

In concatenation, gene sequences are concatenated into a super-gene alignment, which is then analyzed to generate the species tree. In multispecies coalescent, phylogenies are inferred separately from each gene, and a consensus of these gene phylogenies is used to represent the species tree.

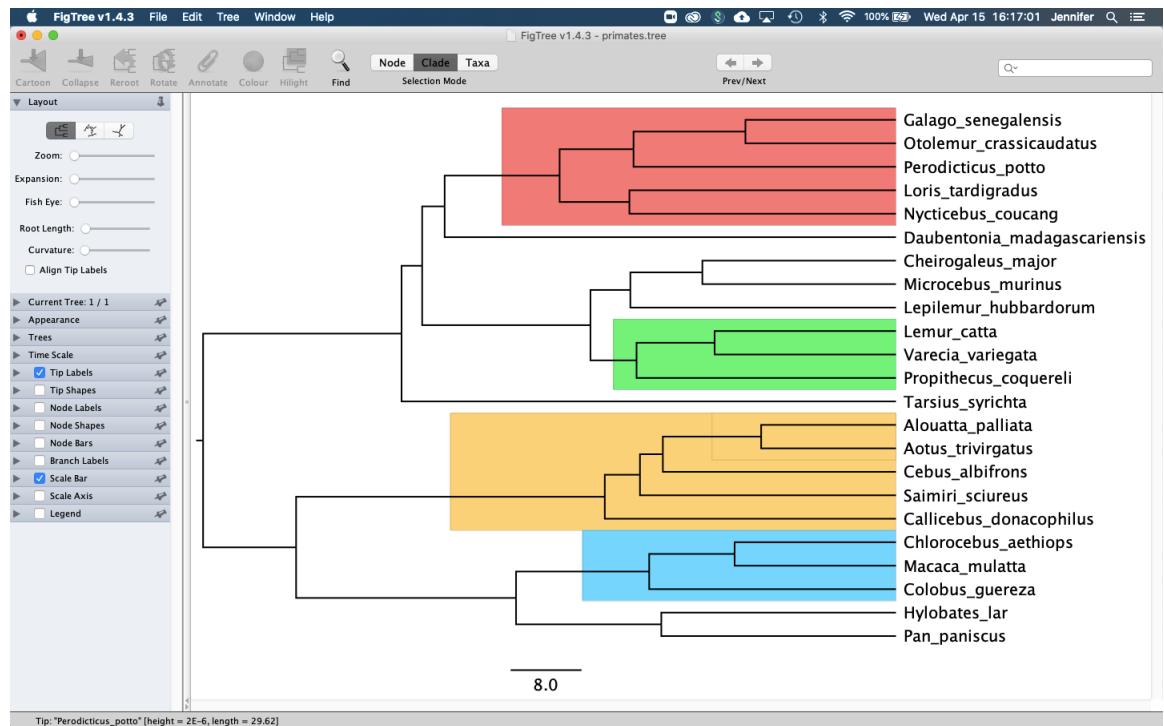
When using concatenation, we can simplified the parameters of the models because we treat them as a super-gene alignment. Thus, the parameters are more prone to be converged and we can use less computational power to get the results. On the other hand, multispecies coalescent requires more computational power and is harder to get converged (the excercise in this lab is an example). Consequently, concatenation is still used much more frequently than the full multispecies coalescent.

### Question 5

Contrast your output tree with the original tree we used as prior "data/primates.tree". Open them in Figtree. Do you observe differences in topology and time estimation? What big conclusion can you make from your observations.

For the topology, it seems like local phylogenetic relationships are similar in the two trees, which are highlighted in the same color in the following plots. However, when we try to trace back to deep branches, the topology are different between these two trees.

For the time estimation, the time unit in primates.tree is larger, which means it generally estimates the evolutionary time to be longer.



primary.tree



primates\_root\_calibration.tree