

Lab 05
Jennifer Lin

Question 1

DNA and amino acids are discrete characters. Phylogenetic CTMCs can be used to model all types of discrete character evolution.

What would it mean to use the GTR model for a morphological character with 3 character states?

The GTR model describes the rate of the transition of each morphological character state. We can use a GTR model to describe how are the rates of the 3 morphological character states change.

Is this an ordered or unordered character state evolution model?

It is an **ordered** character state because we can indicate the probability changing from one character state to another character state. The character states are not equally exchangeable to each other.

What would the parameters be, and how many would there be?

	state a	state b	state c
state a	-	$\pi_b r_{ba}$	$\pi_c r_{ca}$
state b	$\pi_a r_{ab}$	-	$\pi_c r_{cb}$
state c	$\pi_a r_{ac}$	$\pi_b r_{bc}$	-

The π_a , π_b , and π_c are the proportion of the a, b, and c state respectively.

The r_{ij} means the rate of changing from i to j state.

There would be 3 (π_a , π_b , and π_c) + 6 (r_{ij}) = 9 parameters

Question 2

Which model was selected using AICc? jModelTest, like most other phylogenetic software uses log-likelihood (lnL) values. What is the log-likelihood value for the selected model?

TPM2uf+G was selected using AICc.

The log-likelihood value for the selected model is **5710.0932**.

Perform the BIC calculation (remember to select Write PAUP* block). Was the same model selected as with AICc? What is the log-likelihood value for this model?

HKY+G was selected using BIC, which is different from the AICc result.

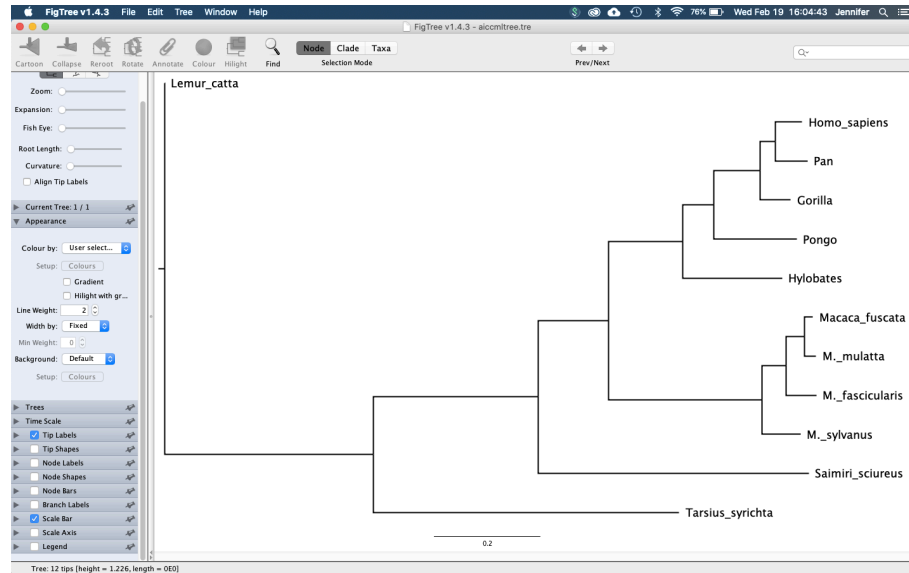
The log-likelihood value for the selected model is **5711.939**.

What if these two criteria differ in their model selection?

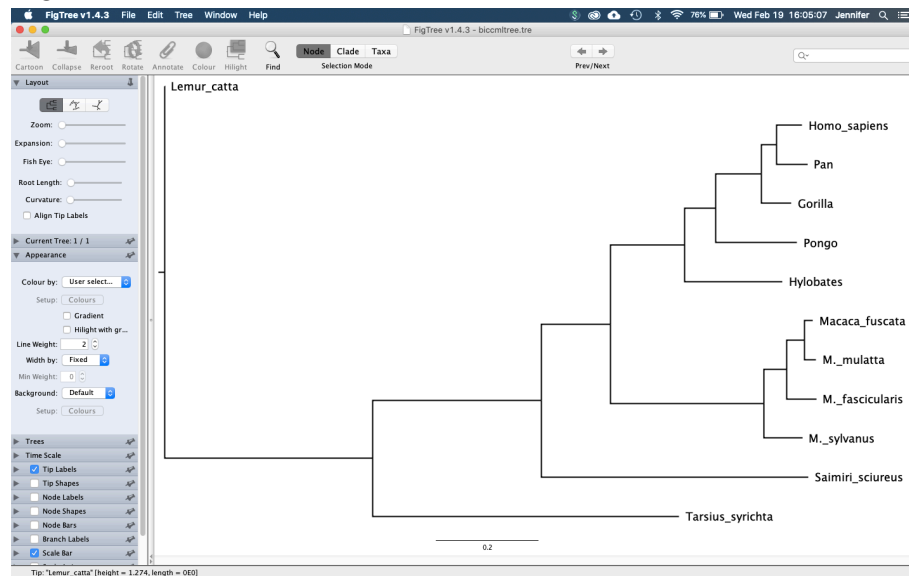
We can compare the two models first. If there are a lot of parameters are actually the same between these two model, that means the results of AICc and BIC actually give similar results given this data. Yet, if the two selected models are quite different, maybe we can keep both of the two selected model and run the following analysis using both models. After doing some further analyses using the two models, we can compare the results.

Question 3

AICc



BIC



Information from AICc

```
BEGIN PAUP;  
Lset base=(0.3585 0.3207 0.0844 ) nst=6 rmat=(2.0810 22.7853 2.0810 1.0000  
22.7853) rates=gamma shape=0.3720 ncat=4 pinvar=0;  
END;
```

Information from BIC

```
BEGIN PAUP;  
Lset base=(0.3637 0.3199 0.0815 ) nst=2 tratio=5.3965 rates=gamma  
shape=0.3630 ncat=4 pinvar=0;  
END;
```

Question 4

Describe the Phylip file format for me. What do the numbers at the top of the file indicate?

The first line for this data is "12 898". The 12 means there are 12 taxa in this file. The 898 means for each taxon, the length of the DNA sequence is 898 nucleotide long (gap included).

For the second line to the last line, each line describes one taxon. It starts with the name of the taxon, followed by 6 spaces, and end with the aligned DNA sequence of this taxon.

Question 5

What was the likelihood value?

The likelihood value is **-5719.449865**.

How does this compare with your previous analyses from today?

Today's analyses make me more confident about the results because we select the best model rather than using all default values before running the analyses.

Load your tree into FigTree and make sure the bootstrap values are visible, take a screen shot, and send it to me.

