

Phylogenetic Analysis of Molecular Data (Botany 563)

Computer Lab 06: Maximum Likelihood I

Learning objectives:

- Familiarize with Maximum Likelihood (ML) methods using PAUP*

Dataset: **Fantastiflora.nex**

Task 1 Get familiar with some of the models of molecular evolution available for maximum likelihood analysis.

- a) Open and execute the dataset **Fantastiflora.nex**
- b) Do a parsimony search (random addition, hold-1; tbr; is fine). If you get more than one tree, use TREES-FILTER TREES to temporarily delete all but tree #1. Otherwise, you can also do a quick neighbor-joining analysis with default settings. Report the kind of search you did.
- c) Determine the likelihood score of the tree under the default likelihood model (HKY) with a transition:transversion ratio of 2:1. Go to TREE – Tree Scores – Likelihood and ask to get a full-likelihood evaluation of tree 1.
What is the likelihood score?
How long did the estimation take?
- d) Repeat step c but, before calculating the likelihood score, change the likelihood settings to “estimate” the transition transversion ratio.
What is the Likelihood score?
How long did the estimation take?
What is the estimated Transition:Transversion ratio?

Is the Likelihood score higher or lower than that generated in step c? Consider why this is necessarily so.

- e) Calculate the likelihood score for the same tree under three other models: In each case (except JC) select “estimate” for the substitution parameters.
 - a. JC: “All rates equal” (or transition:transversion = 1); “Assume equal base frequencies”
 - b. K2P (like JC but $t_i \text{ rate} \neq t_v \text{ rate}$)

- c. HKY (like K2P but base frequencies not equal: set base frequencies to “empirical” or “estimate”)
- d. GTR (like HKY but all transition types can be different = “General Time Reversible”)

Model	-Ln L	Relative substitution rates
JC		
K2P		
HKY		
GTR		

What trend do you see in the Likelihood scores? Consider why this is necessarily so.

Task 2. Try a likelihood search.

- a) Change the optimality criterion to likelihood (ANALYSIS – LIKELIHOOD)
- b) Change the likelihood settings back to HKY with a ti-tv ratio as estimated in 1d (type it in or recalculate it on a tree in memory and then hit “previous”).
- c) Begin a heuristic search using “asis” taxon addition and NNI branch-swapping. Watch the search progress closely.

Most likely the number of rearrangements tried will be much less than is typical during parsimony. Why might that be?

- d) Note the likelihood score and the number of trees found. How does it compare to the score obtained on your first tree (1d)? How does the topology compare?

Task 3. Look at a phylogram of the ML tree and save trees with branch lengths

- a) Open the likelihood settings interface and change the ti:tv ratio to “previous” (this saves time when you calculate branch lengths because it avoids the program having to estimate the ratio again)
- b) With the tree still in memory and with the optimality criterion still set to likelihood, go to TREE-PRINT TREE. Select “phylogram” and show branch lengths. Preview the tree (save it to a PDF file if you like). Be patient it may take a few minutes.

What is the meaning of the branch lengths? Why should we be happy that they are much less than 1?

- c) Select TREES-SAVE TREES. Before saving the trees, select OPTIONS-STORE BRANCH LENGTHS and “6” decimal places. Save tree to a file.
- d) To see if you succeeded, read the treefile back into memory. Select TREES-GET TREES. . Before opening check “store branch lengths if present.” Select the appropriate file and get the tree. You can preview the tree again as in 3b except that you should check “use user provided branch lengths.” This should quickly show you the same tree as 3b. **Why is this much quicker than 3b?**

Task 4. Model selection using likelihood ratio tests.

Refer to the table you generated in Task 1.

- a) Perform a series of likelihood ratio tests and report the results in the table below. The P-value can be obtained from any statistical table (a link to an online χ^2 calculator is in the course website).

Comparison	2x L difference	Degrees of freedom	P-value
JC vs. 2KP			
K2P vs. HKY			
HKY vs. GTR			

- b) Based on these results, which is the optimal model to analyze the data?

Task 5. For the optimal model (the one you selected on task 4d), see if accommodating rate heterogeneity with a discrete approximation to a Gamma (Γ) distribution significantly increases your ability to explain the data. Use 4 discrete categories for now.

- a) Likelihood score of the model + Γ =
- b) Value of the shape parameter (α) =
- c) Do a likelihood ratio test:

Comparison	2x L difference	Degrees of freedom	P-value

- d) Do your results suggest that the data show site-to-site heterogeneity?
- e) Explore alternative category numbers. How do the likelihood and computation time compare to the 4 categories rate approximation?

Task 6. Do a ML search using the preferred model and the parameters calculated above. If necessary, you could speed things up by deleting some of the outgroups. Also, you can start with a MP tree and do NNI swapping.

- a) Copy and paste the “heuristic search settings” block of text from the PAUP console and your resulting tree to a word file. Also report the likelihood score of the tree and the computation time. Name it as: lab06.yourlastname.pdf and put it in the folder Classroom_Scratch>Botany563>2009.

Optional:

Task 7. Familiarize yourself with ML searches using Garli and RaxML through the CIPRES portal.

- a) Go to <http://8ball.sdsc.edu:8889/cipres-web/Home.do>
- b) Conduct a search with 100 bootstrap replicates.