

Phylogenetic Analysis of Molecular Data (Botany 563)

Computer Lab 2:Parsimony searches

Learning goals

- Become familiar with some of the basic operations of PAUP*
- Learn procedures for conducting effective parsimony searches
- Be able to obtain information from trees and integrate it with other information to make biological inferences.

Part one: The Basics. Familiarizing yourself with PAUP*

Task 1. Execute the dataset **fantastiflora.nex**. Go to Data>Character Status.

- 1) How many taxa and characters does this data matrix have?
Taxa: _____ Characters: _____
- 2) How many characters are: variable _____; Parsimony informative _____?
- 3) How many named partitions does this data matrix have? _____
Provide their names: _____

Task 2. Familiarizing with consensus methods.

- 1) Exclude the g3pa and ITS character sets.
- 2) How many characters are there left? _____
How many are variable _____; parsimony informative _____?
- 3) Conduct the following search (more detail on searches below, for now just follow this). Go to Analysis>Heuristic and select the following using the top menu in the Heuristic search window:
 - a. starting tree options → stepwise addition; swap on best only
 - b. stepwise addition options → simple; hold=10
 - c. branch-swapping options → NNI; MulTrees on; swap on best
- 4) How many most parsimonious trees did you find? _____
- 5) Compute a strict and majority-rule consensus tree based on the trees you have found. Go to Trees>Compute consensus and select the appropriate options. (If you have only one tree repeat the search using KEEP to retain extra suboptimal trees). Compare these two consensus trees so as to get a feel for their differences. You can do this in two ways: (a) directly in the PAUP buffer, or; (b) go to Tree>Print Consensus Tree(s) and play with number of trees per page and page setup to see them side by side.
- 6) In the majority-rule consensus tree, what do the numbers above branches mean?
- 7) Is the following statement True or False? (Circle one):
For a given set of trees, the strict consensus tree will never be more resolved than the majority rule consensus tree.
Explain:
- 8) Save ALL the trees you found to a file: Trees>Save trees to file. Name the file appropriately so you can use it later.

Task 3. Visualizing trees.

- 1) Load **ONLY** the first one of the trees you saved in the previous section. To do this: Trees>Get Trees; select '**options**', and tell the program you want only the first tree.
- 2) Comparing Plot types. Go to **Tree>Print trees**. Locate the **Plot type** menu and try out different types of plots.
What is the difference between a phylogram and a cladogram?

- 3) Select the **Rooting** option to root the tree (for now, it doesn't matter what taxon you use to root the tree). Explore the different rooting options.

Task 4. Comparing heuristic search strategies.

Use **ALL** the data in **fantastiflora.nex**. In PAUP*, under the **Analysis** menu:

- 1) Start each of the following searches and report what happens after 2 minutes (proportion of the search completed or time that took to complete; length/score of shortest tree(s) found; number of shortest trees):
 - a. Exhaustive search _____
 - b. Branch and bound search (default settings are OK for now)

 - c. Heuristic Search with the following settings:
Simple addition sequence, hold = 10; NNI branch swapping

 - d. Heuristic Search with the following settings:
Random addition sequence, 100 reps, hold =10; TBR branch swapping

- 2) Based on this comparison, which method is fastest? Which is more thorough?

Part two. Using tree-thinking, PAUP*, and MacClade, to make fantastic inferences.

The *Fantastiflora* group is an example of radiations in islands, endemic to the Fantasy Island. Based on previous morphological analysis, this group is thought to be monophyletic, and related to the following groups: *Wonderlandia* (distributed all along Wonderland), *Neverlandia* (endemic to the island of Neverland), and *Gulliveria* (two species in two locations: Blefuscu and Lilliput). It is known that all these four groups form a clade, and that while *Mythopoeia succulenta* is not part of it, it is a close relative.

1. What is an appropriate outgroup to use in your analysis?
2. Is the *Fantastiflora* group monophyletic? Conduct a **thorough** parsimony search using **fantastiflora.nex** in PAUP. Write down your search parameters:

Report the number of most parsimonious trees you found and their score.

3. Save your tree(s) with an appropriate name (you will use it for the next part!).
Trees>Save tree to file
4. Save your tree as a .pdf. Name is at follows: lab02.yourlastname.pdf and put it in the folder Classroom_Scratch>Botany563>2009.
5. How many species does the *Fantastiflora* clade have?
6. What is the sister group/taxon to the *Fantastiflora* clade?
7. Find a synapomorphy that defines a subclade of *Fantastiflora*. To do this you need to open the **fantastic_morphology.nex** dataset in MacClade. Then go to Windows>Tree Window. Load your tree from part (3). Use the Trace>Trace character option.
Subclade:
Synapomorphy:
8. Based on your preliminary analysis, does the new species discovered (Sp.nov.) belong to the *Fantastiflora* clade? Is there any other evidence that is consistent with your findings?
9. Based on this preliminary analysis, does the data support the hypothesis that the ancestor of *Fantastiflora* colonized Fantasy Island from Wonderland? Explain.

OPTIONAL (as time allows)

Task 1: Explore character weighting.

- 1) Change the weight of some subset of your data, for example, downweight/upweight third positions, exons, a more slowly evolving gene.
- 2) Conduct a new search using the same method you used in Task 3. Note any changes in tree topology. Was there a significant change in the search speed?

Task 2: Apply a stepmatrix to weight characters.

- 1) Write a stepmatrix into you data file (within a PAUP block). For example, here is a 2:1 transversion:transition matrix. It can be written on one line as follows:

```
Usertype titv2 stepmatrix = 4 a c g t . 2 1 2 2 . 2 1 1 2 . 2 2 1 2 . ;
```

Or you can add returns [and explanations in square brackets] to make it more interpretable:

```
Usertype titv2 [name of the usertype]
Stepmatrix = 4 [number of states]
      a c g t [names of states]
[a] . 2 1 2
[c] 2 . 2 1
[g] 1 2 . 2
[t] 2 1 2 . ;
```

- 2) Execute the file.
- 3) Select *set character type* and highlight the characters to which you want to apply this step-matrix (maybe all). Change them to usertype - "titv2" (or whatever you named your stepmatrix).
- 4) Conduct a new search using the same method as before. Note any changes in tree topology. Was there a significant change in the search speed?