

Phylogenetic Analysis of Molecular Data (Botany 563)
Computer Lab 3: Statistics in a Parsimony Framework

Learning goals

- Learn about diagnostics of phylogenetic signal
- Learn how to implement non-parametric bootstrap analyses
- Understand and calculate clade decay indices

Dataset: **fantastiflora.nex**

Part One: Assessing phylogenetic structure in data.

Task 1 Obtain a g1 skewness index for the data.

- 1) Analysis>Evaluate Random Trees. For a clearer visualization select a histogram with 30 classes.
- 2) Evaluate 10,000 random trees and note the g1 skewness statistics: _____
- 3) Refer to the table below (from Hillis and Huelsenbeck, 1992) to get a rough estimate of whether we can reject the null-hypothesis of a lack of phylogenetic (hierarchical) signal in the data. What do you conclude based on this statistics?

	Number of taxa							
	5		10		15		25	
# characters	P=0.05	P=0.01	P=0.05	P=0.01	P=0.05	P=0.01	P=0.05	P=0.01
100	-0.71	-1.66	-0.36	-0.31	-0.15	-0.19	-0.09	-0.20
250	-0.54	-1.26	-0.22	-0.35	-0.15	-0.20	-0.04	-0.09
500	-0.60	-0.84	-0.20	-0.27	-0.10	-0.16	-0.04	-0.08

Task 2 Conduct a PTP test of phylogenetic signal

- 1) There is a bug in PAUP* that can be overcome as follows. Type the following PAUP block into the data file and then execute:

```
Begin PAUP;
  Constraint name = ((1,2));
End;
```

- 2) Go to Analysis>Permutation Tests
- 3) Do a PTP test with 1000 replicates.
- 4) Configure a quick heuristic search strategy. What parameters did you use?

- 5) Does the resultant table suggest that there is significant signal in the fantastiflora dataset?

Part Two: Assessing clade support

Task 1: Do a bootstrap analysis

- a) Go to Analysis>Bootstrap/Jackknife. Do 100 Bootstrap replicates. Select a “full heuristic” search but speed it up. We suggest doing: SIMPLE, HOLD=1, TBR.
- b) Look at the bootstrap consensus tree (Trees>Print Bootstrap Consensus) and identify a weakly supported clade.

Clade _____

What is the BS support value of that clade? _____

What does this number mean?

- c) Repeat the bootstrap analysis but now with 1,000 reps. Did the bootstrap value of that weakly supported clade change? _____. If so, how much? Is this expected or not? Why? _____

- d) Save the bootstrap consensus as a pdf file. Name is as: lab03.yourlastname.pdf and put it in the folder Classroom_Scratch>Botany563>2009.
- e) Based on your parsimony bootstrap analysis, what can you say about the sister group to the Fantastiflora clade?

Task 2: Estimate the decay index of a clade using the KEEP option.

- a) Do a heuristic parsimony search (For example 100 random addition replicates, hold=10; TBR). Note the tree length: _____
- b) Look at the tree (or a strict consensus tree) and identify the following clades:
Clade A: *F. officinalis* + *F. illicifolia*
Clade B: *F. personata* + *F. tithymaloides*
- c) Do another parsimony search after setting KEEP to 1 higher than the shortest trees you found in 3a. Go to Analysis>Heuristic Search, and look for Keep under the General Search Options menu. (If it gets too slow, you can use SIMPLE taxon addition and set MAXTREES to automatically increase).
How many trees did you get? _____ score _____
- d) Generate a strict consensus tree of all trees in memory. Are clade A and/or B still there? _____. If so repeat step 3d while setting KEEP to one step higher until both clades collapse. What are the decay indices for clades A and B?
Clade A: _____ ; Clade B: _____

Task 3: Estimate the decay index of clades using a converse constraint search.

- a) Write clades A and B as constraints into the command line (or in the data file in a PAUP block –if you do this latter option, remember to save and execute). Constraints are defined in tree form with trees written in parenthetical notation, with one extra set of parentheses. You do not need to list all taxa, only those in clade A (or the set that of taxa that are outside clade 1, if that is a smaller list). You can refer to taxa by name or use their taxon number (=position in the matrix). For example:

```
Constraint cladex = ((5, 15, 16));
```

- b) Check whether the constraint is accurate. Go to Analysis>Show constraints
- c) Search for the shortest tree (most parsimonious tree) that does not have cladeA. To do this: Go to the heuristic search menu>general options. Select “Enforce topological constraint”, and “only consider trees that are NOT compatible with the constraint”. Do a search (same parameters as above) and note the length of the optimal tree(s): _____.
- d) Repeat for clade B. Length: _____
- e) What are the decay indeces for clades A and B?
Clade A: _____ ; Clade B: _____
- f) Did you get the same answer as Task 2 (using KEEP)? If not, why?