

Botany 563: Phylogenetic Analysis of Molecular Data

Lab12: Testing correlated evolution of discrete characters

Learning objective: Familiarize yourself with different approaches to test two hypotheses regarding correlated evolution of discrete traits.

Goal: We will make use of two different approaches to test two hypotheses regarding correlated evolution of discrete traits, one in Asterids, and one in Primates. For the first study case we will use Maddison's Concentrated-Changes Test implemented in MacClade (Maddison and Maddison 2001); for the second one, we will use a program called BayesTraits (Pagel and Meade 2004).

TASK 1. Do single seeds evolve more often in fleshy-fruited lineages than in dry-fruited ones?

Data set: Asteridae_fruits.nex (kindly provided by Diane Ferguson).

The data set we will explore is relevant to understanding fruit evolution in the angiosperm subclass Asteridae. The tree is an rbcL/parsimony tree for 137 taxa of Asteridae and two outgroups (*Dillenia* and *Vitis*). There is information on two characters: 1) Fruit (dry vs. fleshy), and 2) fruit seed number (one vs. many). We will first look at character 1 (dry vs. fleshy fruit) to see what sort of information on its evolution can be extracted by looking at its distribution on the tree. Then we will look for evidence of correlated evolution between these two characters.

1.1) Open the file Asteridae_fruits.nex in MacClade.

In order to see the character and character-state definitions of a character, you can select the **data editor** (windows menu) and then hit "state names & symbols" (characters menu); or you can use the following display option: Display>Data Matrix Styles>wide w/state names.

1.2) What is the number of steps for character 1? _____ And for character 2? _____

This information can be obtained under the **tree window** (windows menu) by selecting "character list" (character menu). Also, if you choose to "trace" a character (that is, to paint it onto the tree), you can see the number of steps of character 1 by scrolling through the characters in the character legend window until character 1 (dry vs. fleshy) is selected.

1.3) How many most-parsimonious reconstructions are there for character 1? _____

For this you need to trace the character of interest, in this case, character 1. Go to "resolving options" (trace menu), and select "show all most parsimonious states at each node". This will turn on the "number of MPRs" option under the trace menu, which is what you want.

Repeat the procedure for character 2: _____

1.4) On the tree, what is the minimum and maximum number of gains of (fleshy → dry) and losses (dry → fleshy) for character one? Select "state change & stasis" (chart menu). Then look at minimum and maximum values for the traced character only. The area of these circles is proportional to the maximum and minimum number of changes. You need to click on the area to get the values.

Alternatively, you can go select "view table" (display menu).

GainsMin= GainsMax=
LossesMin= LossesMax=

1.5) To perform a test to see if the number of steps for character one on this tree is lower than would be expected by chance, you need to do the following:

- a) Select character one in the data editor. Use the “replicate” command (edit menu) to generate 100 copies of the character.
- b) Highlight characters 2-101 (the duplicates but not the original) and shuffle the states between taxa (i.e., permute) using the “shuffle” command (utilities menu).
- c) To chart out number of steps for each character (x-axis) against the number of characters, go to **tree window**, and select "character steps/etc" (chart menu).
- d) How many of your permuted characters have as few or fewer steps than the original character?

What do you make of this result? What do you think of the test?

1.6) Now we wish to test the hypothesis that single seeds evolve more often in fleshy-fruited lineages than in dry-fruited lineages. That is to say we are looking to see if there is a pattern of correlation in the evolution of fruit fleshiness (character one) and fruit seed number (character two). This will be done using Maddison's Concentrated-Changes Test using character one as an independent variable. You might want to delete the 100 replicates you created before to keep things tidy.

a) Set to ACCTRAN optimization (trace menu > resolving options). What is the total number of gains (one→many) and losses (many→one) of character TWO (seed number)?

Gains:

Losses:

b) To save you time we have counted-up the number of gains and losses of character two that take place on branches whose end-states of character one are fleshy (i.e., on yellow branches).

Gains (G_F): 3

Losses (L_F): 6

c) The Concentrated-Changes Test determines the probability of various numbers of gains and losses in the dependent character (here, seed fleshiness) given the number of gains and losses observed in a certain clade (here, G_F and L_F). The independent character is the traced character (in this case character one), and the null hypothesis is that gains and losses are randomly distributed. To perform this test in MacClade, you need to trace the character that is your independent variable (i.e., character one, fruit fleshiness), and select the “test correlation” tool from the tool palette (the name of a tool will appear as you click on it). Then select the base of the Asteridae (excluding the two outgroups: *Dillenia* and *Vitis*). Ask MacClade to do 1000 simulations of the Gains and Losses found in **1.6a** to see how many typically occur on fleshy branches. The “distinguished branches” are those having fleshy state, which is coded as zero. Set the ancestral state to "either" and use "actual changes." To find the probability of various gains/losses scenarios, simply select or de-select the appropriate boxes.

What is the probability of G_F or fewer changes on fleshy branches?

What is the probability of exactly G_F changes on fleshy branches?

What is the probability of L_F or more changes on fleshy branches?

What is the probability of G_F or fewer **and** L_F or more changes on fleshy branches?

d) Is there evidence that single seeds evolved more often in fleshy-fruited lineages than in dry-fruited ones?

TASK 2. Do female primates tend to prominently advertise their estrus when living in a multimale mating system? (Based on the BayesTraits manual by Mark Pagel and Andrew Meade)

Data sets: primates.trees, primates.txt

About *BayesDiscrete*

BayesDiscrete is part of *BayesTraits*, a computer package for performing analyses of trait evolution, which combines several previous programs, such as MultiState, Continuous, and Discrete. *BayesTraits* is available at <http://www.evolution.rdg.ac.uk/BayesTraits.html>.

BayesDiscrete is used to analyze correlated evolution between pairs of discrete binary traits. The test of correlated evolution compares the fit of two models of evolution, one in which the two traits evolve independently on the tree, and one in which they evolve in a correlated fashion (such that the rate of change in one trait depends upon the background state of the other). The independent model requires two parameters per trait (a total of four) and the dependent model has a total of eight rate parameters.

BayesDiscrete can do either ML or MCMC analyses. When doing MCMC, the prior for rates need to be set, and the test is done using Bayes Factors. Under ML the program will calculate the ML values of all parameters separately on each tree in the treefile, and the models can be compared using a likelihood ratio test statistic, as follows:

$$\text{LRT} = 2 * (\log\text{-likelihood (Dependent model)} - \log\text{-likelihood (Independent model)})$$

BayesTraits is run from the Terminal (should be in the doc), and it is the easiest to have all your files in the same folder as the application. Since we are testing for correlated evolution between two traits, we need two files: 1) a treefile: in nexus format with one or more rooted trees with branch lengths. The trees in the must be rooted such that the root forms a binary node at the base of the tree, with a positive branch length leading to the outgroup(s), and 2) an input data file: a plain text file with one line for each taxon in the tree with the binary data scored (the taxa names must be spelled exactly as in the tree, with no spaces within them, but need not be in the same order). In the case of *BayesDiscrete*, this file must have exactly two columns of data and they must take the values "0" and "1". If a data point is missing a '-' symbol can be used. The form of the command to run the program is: `./BayesTraits treefile.trees inputdata.txt`

In this lab, we will use *BayesDiscrete* to see whether or not there is a pattern of female primates to prominently advertise their estrus when living in a multimale mating system using a ML approach. Estrus advertisement in females consists of a prominent swelling of the perineum at or around the time of ovulation. A multimale mating system is one in which females mate with more than one male during their estrus (in contrast with monogamous or unimale mating systems, the latter being ones in which harem-holding males mate with each of the females in their groups). Theory predicts that females will advertise in multimale mating systems. The two traits in the file primates.txt were scored as follows based on published data: Estrus advertisement (yes=1, no=2), Mating system (multimate=1, otherwise=0). The file primates.trees contains 10 rooted trees, with branchlengths.

We will: 1) perform likelihood estimations using the dependent model first and ML analysis; 2) repeat the analysis with the independent model (also ML analysis); and, 3) compare both likelihood estimates with a likelihood ratio test.

Running the program

- a) Open the Terminal (Applications>Utilities)
- b) To change your working directory to BayesTraits, **TYPE:** cd Desktop/BayesTraits
- c) To launch *BayesTraits*, we need to “feed” it a treefile and an input data file. To do so, **TYPE:** ./BayesTraits primates.trees primates.txt
- d) A screen will appear asking you to select the model of evolution to use. You should type in the number of the model required. We will use a dependent model first (number 3).
- e) Once you’ve done this you will be asked to select either ML or MCMC as the method. Here we will do ML.
- f) Now, **TYPE:** mltries=25. This tells the program to use more than the default number (10) of optimization attempts in finding the likelihood.
- g) To start the analysis, simply **TYPE:** run.
- h) The output shows the tree number and its likelihood given the model, the values of the rate coefficients and the reconstructed probabilities of the two states at the root of the tree (2 binary traits give four possible pairs at the root). We will focus on the likelihood scores only.
- i) Rename the log file as primates.dependent.txt
- j) Repeat the analysis with the independent model.

You will get one likelihood score for each tree under both, the dependent and independent models. Unfortunately, there is no natural way to combine these results, which show that there is tree-to-tree variability. I suggest opening the two log files in excel, and computing the likelihood ratio test so that you can get an idea of the range of this statistic.

- k) What is the likelihood ratio test for tree 1?
- l) How many degrees of freedom?
- m) Is this significant?
- n) What does this test tell you about the evolution of mating system and estrus advertisement in female primates?