Advanced MacClade

Introduction

Today we will be using some of the various functions of MacClade that are useful for testing hypotheses about the evolution of characters. The exercises today are based on portions of three labs from IB 200B, so don't worry if you don't finish everything. You can always refer to the instructions post hoc as you're working on your project. With this in mind, feel free to look over the lab and focus on the features you feel might be most useful for your project. Also, be aware that this is not a complete treatment of all of the features of MacClade. If you're interested, IB 200B focuses on these kinds of comparative analyses in much more detail…

Step Matrices

The first thing we will consider is the step matrix. As you may recall, a step matrix is method that is used to weight changes from one character state to another. For example, we could make it twice as hard for character state A to change to State Y as it is for it to change to state B. Usually step matrices are used for molecular data sets, because we know from biochemistry that transitions are more difficult than transversions. Theoretically, we could also make a step matrix for morphological characters, but usually we don't know enough about their probabilities of change to make an accurate step matrix. A typical step matrix might look something like:

	A	B	C	D
A	0	1	2	4
B	1	0	3	2
C	2	3	0	1
D	4	2	1	0

Step matrices can be constructed and used in PAUP and MacClade.

--Open one of the molecular data sets (the perennial favorite Primate mtDNA will work fine for this operation) in MacClade.

--Go to the characters menu select the 'character list' option. A box with the characters and their status will appear. Highlight some of the characters. Hold down Shift to select multiple characters.

--Now go to the 'characters menu' and select 'edit type.' A dialogue box will appear. Scroll through the different 'built in' step matrices and compare how they are different.
--Now push the 'new' button, and give your step matrix a name. Then adjust it so that it has 4 states using the 'change size' button.
--Enter character state change values into the matrix. You can select multiple boxes by holding down Shift. Normally you would want to have a reason for assigning the different character state change weights, but we'll ignore that technicality for now. You now know how to make a step matrix in MacClade.

You can also have MacClade automatically make a step matrix in which the cost of a change between two amino acids is the minimum number of amino acid changes required to convert between the amino acids. However, this requires the use of another of MacClade's functions; converting a sequence of bases to a series of amino acids.
--Open the data set 'primate mtDNA.' If it's open already, deselect the charactes you selected in the character list.
--Go to the 'characters' menu and select 'translate to protein' which can be found under the 'data format' option. A dialogue box will appear. Hit ok. MacClade will translate the sequence into a protein.
--Go to the character list and select some of the characters.
--Now go to the 'characters' menu and select 'edit type.' Select 'new.'
--Next check the 'protpars type based on genetic code' box and then press 'create'. MacClade will create a step matrix based on the currently selected code (which you can modify) that shows the minimum number of changes between amino acids.
When you close the file, make sure you <u>DO NOT</u> save the changes you've made (if you do all of the DNA sequence data will be lost).

--Now go to PAUP*. Open the data set called 'primate-mtDNA.nex.'
--Perform a heuristic search and examine the resulting cladograms.
--Now go to the 'data' menu and select 'show other', then 'user type,' and 'all.' This will display the two step matrices included with this data set. Examine the step matrices and note how they are different.

--Go to the 'data' menu and select 'set character types.' A dialogue box will appear. First push the 'all' button to pick all the characters, and then select one of the step matrices from the 'user defined' choices.
--Perform a heuristic search and compare the results to you previous cladograms. Are there differences? Why or why not?
--Repeat for the other step matrix. Again, are there differences?

Frequency of Base Changes

MacClade can also chart how often a particular base in a molecular sequence changes across a given cladogram. This sort of knowledge might be useful for developing a weighting scheme for the characters (for example, should all third positions be down-weighted?).
--Open the data set 'primate mtDNA.'
--Go to the 'chart' menu in the tree window and select 'character steps/etc.' again. Select the middle option in the upper row, which will chart how many base changes occur at each particular site (character). Then hit chart.
Note that there are some areas of the sequence that change relatively infrequently. Remember, this is charting the number of steps required for each site based on the currently selected cladogram.

--Now go back to the 'chart' menu and select 'chart options.'
--Push the 'changes' button, and then 'restrict.' Then push 'transversions.' The chart that results will show you the distributions of the transversions along the sequence. Again this is based on the currently selected cladogram.

--Go back to chart options and select the upper right hand chart. This will cause MacClade to chart the changes by codon position. Is the distribution similar to what you expected?

As you view these different charts be sure to try some of the different buttons on the right-hand side of the chart. They change how the results are displayed (*e.g.*, in table

form, graph form, etc.) If you click on the 'X' button, the data will be displayed in text form.

--Now go to the 'characters' menu and choose 'inclusion sets' and then 'inclusion set list.'
--A new window will appear. Click on the number 4 to choose 'only 2nds,' which obviously includes only second positions, then push 'use.'
--Go back to the chart window and chart how many of the 2nd position changes are transitions and transversions. Does there seem to be a bias toward one type of change?

--Go back to the inclusion set window and choose the 'none' option from the 'inclusions sets' pull down menu.
--Now go the chart menu and select 'states.' MacClade will then show you how many of each nucleotide is present in the current data set. Note the low number of 'G's.' Can you think of a situation where this type of information could be useful?

Practice with Character State Reconstructions

There are often different ways to optimize transformations among character states on a given cladogram. Two common techniques used to determine character state optimizations are Accelerated Transformation (ACCTRAN) and Delayed Transformation (DELTRAN). Under ACCTRAN changes among states happen earlier on a cladogram rather than later, and the numbers or reversals are increased. If DELTRAN is used, state changes occur later on the cladogram and independent gains increase.
What two biological processes do these different models of character evolution correspond two?
--Open the DELTRAN/ACCTRAN example in your MacClade examples.
There are 64 equally parsimonious reconstructions of character evolution for this character. If we are interested in the number of changes from wrinkled (state 2) to smooth (state 0) texture, then we might choose to ignore most of these resolutions and examine just the DELTRAN and ACCTRAN resolutions, hoping that these would provide extreme values for the number of wrinkled to smooth changes.

--Under “trace”, click on “resolving options.” This provides you with the option of choosing ACCTRAN of DELTRAN.
--Choose DELTRAN and examine the reconstructions, then repeat for ACCTRAN. How do the resolutions differ?
--Now set the ‘resolving options’ back to ‘show all most parsimonious states’ and then go to the ‘trace’ menu and select ‘Show MPRs Mode,’ which will display in turn each of the most parsimonious reconstructions for the characters of interest. Now use the “apple-R” keystroke to pass your way through each of the possible character state reconstructions.
--Using the “go to” option in the ‘show MRPs’ menu, go to tree reconstruction number 64. How many changes from wrinkled to smooth occur?
--Which reconstruction would you prefer, ACCTRAN, DELTRAN, or one of the equivocal cycling options (remember, they’re all equally parsimonious)? There is no right or wrong answer, but try to think in terms of your own project and decide what reconstruction you might prefer given a particular hypothesis.

Patterns of Correlated Character Evolution

The concentrated changes test of Maddison (1990) is designed to test the association of changes in a binary character with some other binary variable within a clade of interest. It can test whether changes (from 0 to 1) in one (dependent) character are more concentrated than expected by chance on branches having a shared character state for another (independent) character.
--Open the file “Concentrated Changes Example.”
--Note that there are two characters included. The cladogram upon which these characters will be mapped has already been created, and we’ll assume it is robust.
--Go to the data editor and using “State Names and Symbols” option in the ‘characters’ menu, rename each of the two characters and states as follows:
Character 1 = dependent character = Fruit Type; state 0 = ancestral, fleshy; 1 = derived, dry
Character 2 = independent character = Ecology; state 0 = tropical rainforest; 1 = open savanna

--Count the number of times dry fruits have evolved from fleshy according to this cladogram.
--Now examine their ecology and note that all instances of dry fruits occur in clades that inhabit the open savanna. This may suggest something about the causal influences of ecology or fruit type, but we must test this relationship for statistical significance. Are the changes in fruit type from fleshy to dry "overly" concentrated in clades inhabiting the savanna?
--To test this, we must first establish a null hypothesis. Write that down here:

Our question now becomes: What is the probability, given the null hypothesis, that three gains of dry fruits from fleshy fruits would occur in clades inhabiting savannas?

--Go to the tools window (in the bottom corner of the tree window) and click on the "test correlation tool" (the tree with a 'c' at its base). Make sure that you are tracing the Ecology character, as the character currently traced will be viewed as the independent character. Then click on the lowest branch on the cladogram.
--In the dialog box that comes up (called the Correlation test Parameters dialog box) add in the number of gains and losses of dry fruits.
--Now click on "exact count." This will use the formulae presented by Maddison (1990). The exact count algorithms become computationally challenging (especially for these computers) as one increases the number of changes in the dependent variable of interest (in fact, changes in the dependent character of more than 5-6 can take a long time). Doing simulations is often a necessary option. The simulations generate changes randomly on the clade selected and count the number of gains and losses on the branches with the specified dependent and independent variables across the entire cladogram. Only those that correspond to the previously specified number of gains and losses will be examined for the branch distributions of gains and losses
--Now click "calculate."
--The next dialog box that will appear is the correlation test results dialog box, which will allow you to ask what the probability of having more, as many, or fewer than the

indicated number of gains and losses along the distinguished branches (independent variable branches).
--Find out what the probability of having as many or more than three gains and zero losses. What can we say about the evolution of dry fruits in savannas?

What are three assumptions of Maddison's Test?

Random Data

Random data is useful for determining whether observed values for a particular measurement are statistically different from those obtained from a random model. Of course, the details of the model you choose will influence your results (as usual).
MacClade can generate random data to fill all or part of a particular data matrix.
Open MacClade and create a new data matrix with 15 taxa and 100 characters.
Select all of the character cells and then select 'fill random' from the utilities menu.
A dialog box will appear showing the frequencies of each character state. You can adjust these and the number of states by moving the black bar and purple boxes. However, be aware that your frequencies must sum to 1.
The three buttons on the bottom can be used to make your frequencies add up to 1. The first makes the frequencies equal. The second makes them sum to 1 by adding or subtracting a constant amount to each. The third makes them sum to 1 by multiplying or dividing each by a constant amount (thus retaining their relative proportions).
Generate some random data. Make a chart of the frequency of each state and see if MacClade gave you the results you asked of it.

MacClade will also randomly shuffle existing data in a data set. Select a portion of your data set and then select 'shuffle' from the utilities menu. The data will be scrambled and reassigned randomly to the cells.

Finally, MacClade will let you randomly evolve characters in reference to a particular cladogram.
--Go to the tree window, and construct a cladogram you like.
--Then go to the characters menu and select 'evolve characters.' A dialogue box will appear asking you to select the probabilities of change for your characters states, and the number of characters you want to evolve (they will be added to then end of your data set. Again, the probabilities must add up to 1 (you'll recognize the norming tools from before. MacClade will

save a text file giving you some details of the results it came up with. Be sure to check this box in the dialogue box (save the file to the desktop) and see how closely the actual values meet your expectations. Trace the characters and look at hw they change.

The program uses the probabilities to come up with values for each branch. By default, each branch is only one step long (thus MacClade gives each branch one 'chance' to change). However, you can modify the branch lengths (and thus give some branches several 'chances' to change). You can do this by in the tool window. Hold down the option key and select the 'set evolve segments' tool (it looks like a ruler). While still holding the option key, click a few branches and make them longer than 1 step. You can display your changes by selecting 'show evolve segments' from the display menu.

Generate some characters with your new branch lengths. When you do this, be sure to check the 'consider branch segments' option in the dialogue box. Now trace all changes on the cladogram. Does there seem to be a concentration on the longer branches?

Random Cladograms

MacClade can generate random cladograms, and can do so in several ways.

Go to the tree window and select 'Create trees' and then 'random' from the tree menu. A dialogue box will appear.

There will be three options available. "Equiprobable trees" generates cladograms under a model where it is equally probable that any dichotomous, rooted cladogram will be chosen. Generating random cladograms in this manner is a good way to get a feel for the overall distribution of treelengths for your data set without generating all of the possible cladograms.

"Random joining/splitting" adds taxa to the cladogram in a random order. This simulates a case where all species have an equal chance of speciating, and is biased towards producing asymmetric cladograms

"Random partition' divides up taxa into partitions, with the each taxon having a 0.5 chance of falling into one of the partitions. This method is biased towards producing highly symmetric cladograms. The fourth option (that probably isn't highlighted) in the dialogue box can be used to randomly resolve polytomies (if your cladogram has them).

MacClade will also save the details of the cladograms it generate in a text file, allowing you compare observed vs. expected frequencies of cladograms with different degrees of symmetry. Generate 500 cladograms using each method and save the report for each run. Then compare the reports to see how the results differ. The expected values are based on equations in Maddison

and Slatkin (1991; Evolution 45: 1184-1197), a great reference when thinking about random cladograms.