Lab 15A: Biogeographic Reconstruction on Phylogenies

Today we're going to be looking at programs that compare trees between two associated groups of objects to deduce their common history. This could be a comparison of host & parasite, organism & gene or area & organism trees. The different relationships can be analogized like this:

Host	Organism	Area
Parasite	Gene	Organism
Host switch	Horizontal transfer	Dispersal
Cospeciation	Orthology	Vicariance
Parasite speciation on one host	Gene duplication or allelic divergence	Sympatric speciation (kind of)
Parasite extinction	Gene loss or fixation	Extinction

In all three cases comparisons can be made between the two trees to see how often dispersal or vicariance (or their analogous events) best explains the situation. We are going to try several different approaches that use different criteria to determine the relationship between areas and associated organisms.

Today we will use the programs *Treemap* and *Lagrange* in addition to *Mesquite*.

Treemap

COMPONENT by Rod Page is a good program for analyzing and comparing trees and can do some of these comparisons. However, it can only reconcile the trees. To reconcile trees is to add hypothesized extinct taxa to the organism/parasite tree based on the assumption that there is no dispersal/ host switching. Thus all differences between the trees are a consequence of an ancestral area/host having two organisms/parasites one of which has since gone extinct. The reconciled tree adds hypothesized extinct organisms/parasites.

There is another program *Treemap* from Rod Page's lab. It looks cool and allows you to do diverse comparisons of the trees. Both programs are available for free from http://taxonomy.zoology.gla.ac.uk/software/index.html. We are going to use *Treemap* to explore comparisons between the host-parasite trees that they provide as examples. We will look at host parasite-data, but the same principles apply for biogeography.

Open *Treemap* and use it to open the **HAFFNER88.NEX** example file. This is a file with a phylogeny of gophers and their associated lice. This file contains two trees, one for the host and one for the parasite, and a description of which hosts are associated with which parasites.

You will see four windows.

The first window is the **Tanglegram**, which shows the parasite tree on the right, the host tree on the left and arrows connecting the associated hosts and parasites. This is basically a graphical representation of the data in the input file. You can click on the

nodes to switch the branches around and try to untangle the intersecting lines. This will not change the topology of the trees or the information that you are looking at, only the appearance. Pull down the **view** menu and select **phylogram**. Then pull down the **view** menu again and select **Internal labels**.

The second window is called the **Reconstruction Window**. This is where the program does its real work. This window can be difficult to read. It shows the two trees overlaid. The parasite tree is black and the host tree is grey with parasites below their associated hosts. Circles at nodes of the parasite tree represent cospeciation and squares represent speciations of a parasite on a single host. Initially it will show the reconciled parasite tree, which assumes no host switching. Thus you will see that for some of the branches, there are two parasites on the branch. This indicates that for this reconstruction there were two species of parasites living on the host at that time.

The third window is the **Branch lengths** window. It shows a graph comparing the distances of branches shared by the host and the parasite. If the molecular clock holds and your reconstruction is correct, then these points should fall on a straight line through the origin. Why? As you can see the only two points plotted seam to fit a straight line.

The fourth window is the **Histogram Window**, but you will have to run an analysis for this to show anything.

This file does not actually have branch length data, so it is better to look at coalescence times. In the **Branch lengths** window pull down the **View** menu and select **Plot coalescence times**. The plot will now change to show a comparison of the "age" of the nodes shared by the parasite and the host. These are the nodes where they cospeciated. As you can see, you now have many more points, and, although three of them distinctly fall on a line, the other two do not.

Reconstructions

In the **Reconstruction** Window click the square next to the node labeled **13**. This will make the parasite tree change, so that the clade with *cheriei* and *costaricensis* has undergone a host switch. (Although this may be hard to tell, because of crappy graphics. It is experimental.)

Look at the coalescence time graph in the **Branch lengths** window again. It has changed to reflect the fact that you removed a cospeciation node, but this does not lead to any improvement. Click the square in the **Reconstruction** window again to return the tree to its default state.

You can try clicking different combinations of parasites and nodes to get a reasonable parasite tree. It works well on Macs, but I can't figure out how to get this to work on a PC. (Once again experimental).

There is also a reconstruction that was saved with this file. Pull down the menu at the top of the **Reconstruction** window that says **None** and select **Pagel_1990**. How does this one look? What about its coalescence times?

You can also search for the "best" reconstructions. This program defines "best" as the tree with the greatest number of cospeciation events. While in the **Reconstruction** window, go to the **Reconstruction** menu and select **Heuristic Search**. The program will search for the best tree. How do the coalescence times look now?

This time do an **Exact Search** for the "best" tree. When you're done, pull down the reconstruction menu labeled **None**. You will find six trees labeled best. Look through these trees. How do they look? Do any of them have completely consistent coalescence times? What assumptions may be violated that could explain this? Which node seams to be particularly problematic? What might be going on here if none of the assumptions are violated?

Randomization

One way to test if your pattern is significant is to randomize your data, and see how often you get results with as many cospeciations as you got from your actual data. If you rarely get that many cospeciations in the best reconstruction, then your results are probably significant.

Pull down the **Randomisation** (they're Scottish) menu and select **Parasite tree.** Type **100** for the number of trees and hit **OK**. This will generate 100 randomizations of the parasite tree and count the maximum number of cospeciations on each one.

Go to the **Histogram** window. You will see a distribution of the results from your randomization. How many times did you get as many or more cospeciations than you found in the real data? Is this a significant result? What if you randomize the host tree or both trees?

Just for fun

Close this file and open **HAFFNER94.NEX**.

Try to rearrange the **Tanglegram** so that it makes since. You can't get it perfect, but you can improve it.

Look at the **Branch Lengths** window. This data set has actual branch lengths, and as you can see the graph is a lot messier. Can you improve it?

Question 1. How significant is the number of cospeciations on this tree? (This is not related to the previous question.)

Brook's Parsimony Type II

Last week we used Type I Brook's Parsimony to construct an area cladogram using two organism trees as evidence. This week we will use Type II Brook's Parsimony to reconstruct the history of parasites on the host tree. This is done just like Type I. Branches are labeled on the parasite tree (or on the host tree it works either way around); a matrix is then constructed for the host taxa, in which each branch on the parasite tree is counted as a character and the host taxon has state 1 it if it has any parasites descended from that branch of the tree and 0 if it does not. These characters are then reconstructed on the taxon tree using parsimony. If a parasite tree branch is reconstructed as being in state 1 on a branch of the host tree, then that means that that branch of parasites had already speciated by that time in the host history. In this way Brook's parsimony assumes that all the discordance between the host and parasite trees comes from ancestral speciation and independent loss of parasites without any host switching.

Open the file **hafner88_Mesquite** in *Mesquite*. This is the exact same data as we were just looking at. You will find a parasite tree with the branches already labeled, an empty matrix for the hosts and an association matrix saying which parasites are