Phylogeny computer program exercise

Due Monday, 11/26

Get the program DNATREE by using the links on the course web page, which take you to the DNATREE web page which has links for downloading.

Use the program to simulate three trees with different rates of evolution. For each of these runs, use the same random number seed. This way all three will have the same true tree topology, but different branch lengths that are proportional to the rate of evolution. Simulate three data sets, each with one of these trees of 10 species and sequences of length 500 sites (this number of sites is a change from the default settings). For three different values of the rate of evolution, 0.1, 0.4, and 2.0, simulate the tree (and make sure to write it down), then simulate a data set, and then use parsimony to reconstruct the best estimate. Compare the estimate to the true tree.

To use parsimony to reconstruct the best tree, take advantage of the ability to rearrange the tree by hand in the last part of DNATREE. You can do this by using the T (“Try rearrangements”) command in the tree rearrangement menu, which is in the horizontal printed line below the tree. Do the T command on all the tip nodes one after another (using their numbers that appear in the tree on the screen) and on all the interior nodes as well. Look in the line of BETTER results for the best node to place that one at. Then use the R (“Rearrange”) command to move it to that node. Keep doing this on all the nodes in turn, until none show any better placements. Then write down the tree. If it disappears from the screen, the dot(.) command will make it reappear. Do not use the tree that first appears on the screen when you enter the last part of DNATREE – it is a completely arbitrary tree that is not a good estimate of anything. But do write down the number of changes that the arbitrary tree required, as well as the number of changes that were required on the most parsimonious tree that you could find.

IMPORTANT: You are asked to summarize the results by counting, for each internal branch in the unrooted version of the true tree (there should be 7 such branches), whether it is also present in the estimated tree. Report how many of the 7 branches are shared between the estimated tree and the true tree. A branch is counted as the same if it separates the species into the same two sets. Thus if a branch of one tree has species A, C, D, E connected to one of its ends, and species B, F, G, H, I, J to the other, it is considered the same branch as one which has the same two sets BFGHIJ and ACDE connected to its two ends (order of these two sets does not matter).
Note that the root of the tree is not estimated by the parsimony method (it is put in an arbitrary place). So the two branches coming from the rootmost fork of the tree are to be considered to really be one branch, and combined into one, and the branch coming from the root to that rootmost fork is also not to be counted. In effect, in your mind pull off the root and merge the two rootmost branches into one:

removing the root: \[ \begin{array}{c}
E \\
B \\
A \\
D \\
C \\
F
\end{array} \]
becomes: \[ \begin{array}{c}
E \\
B \\
A \\
D \\
C \\
F
\end{array} \]

Also, you can and should ignore all the branches that connect to tips. They are not interior branches and they are always present on all trees. An interior branch is one that does not have a tip species at either of its ends. After mentally removing the root, look at each interior branch of the true tree and ask whether you see the same branch in the parsimony tree.

In the tree shown above, the one on the right has two branches that I have marked with asterisks. The one with a single asterisk separates species ACD from BEF, so it has that partition of the set of species. The one marked with two asterisks separates the tip species F from the others. It is not to be counted when looking for differences, as it has a single tip species at one end.

You may or may not see a clear pattern from your run, but we will be summarizing the results in class and on the web site and when they are all put together we should be able to see whether the higher rate of evolution leads to greater or less accuracy of the estimate of the tree.

Report briefly (1-2 pages) on the results. Give enough material to make it clear what you did. I need to see the true tree and the reconstructed tree in each case, and have the number of branches differing between them. For each of the three cases I also need to know the number of changes (to evolve the DNA sequences) in the random tree that appears in the last phase of **DNATREE**, as well as the number of changes needed on the reconstructed tree. You may be able to put the trees into your document by cut-and-paste from the **DNATREE** screen, using a constant-width font such as Courier (if you do this, you may need to not have the **DNATREE** screen setting be of type ANSI).