Math 340/Bio 373
Tutorial (of sorts) using Phylemon
Fall 2007

The link for this web resource is

http://phylemon.bioinfo.cipf.es/

Once you enter, you will have an opportunity to register. Do that. And I advise having it remember your password.

Then go to the Tools menu. You will see lots of software opportunities. We have not talked about tests for evolutionary models, but you could note their existence. We will focus on the phylogeny tools. (You can use MEGA to get NJ and UPGMA trees and Phylemon for likelihood, parsimony and MCMC.)

There are two ML packages – PhyML and one listed under the Phylip collection. Either is OK. Let's go with PhyML to start with.

First you need to load the data. You need to use a *.aln alignment file that you have already saved from a MEGA session. Browse to that file and click open. The file name will show up in the Phylemon server. There are all sorts of settings you can make – choose the evolutionary model is one of them, and DNA vs amino acid data. Leave all the other settings at the default – the sequence format from a MEGA file, for example, is interleaved, not sequential. Scroll down to the bottom of the page and hit run.

Shortly (after the little dog quits running back and forth) you will see three output files – a likelihood file, and statistics file and a tree file. The first gives the probability of a site given the model, for every site. The second is self-explanatory, when you open it, you will see. The third one is the tree file which describes the tree in the compact but not so useful Newick format.

To see an image of the tree, click on send results to ETE. Make any selections you like on the next screen and click on run.Then you can open and print if you like the .png file that shows up on the next page. The print size is miniscule and maybe you can discover a way to make it bigger. Alternatively you can save the .txt tree file as a .txt file and use MEGA (Phylogeny > Display Newick tree from file) to print it in MEGA.

Most of the other packages work pretty much the same way.
TreePuzzle – does quartet puzzling. Put in a .aln file just as for PhyML. Again you get three output files which you can look at to see what's there. As before the actual tree is in the third file. Once again, send it to ETE or save as .txt file and use MEGA.

For parsimony, use the Phylip package (DnaPars or ProtPars,whichever is appropriate.) Proceed as before. Your output tree file can just be sent to ETE for drawing, or you can use Consense to find a consensus tree and then send THAT to ETE.

Mr Bayes – this one is different and a bit trickier, involving more input from you than just “run”.

The format for the input is a nexus file, not an alignment file. So here’s what you do. Using your .meg file, export it as a Nexus file as follows:

File > Open data> (open a .meg file). You are now in MEGA’s sequence explorer. Then do
Data > Export Data and select Nexus *(PAUP 3.0) and hit OK. As I have learned, this conversion is imperfect. You will need to edit this file as follows:

Change gap= - to gap=. (so replace the dash with a period.) and delete matchchar=. Change datatype=nucleotide to datatype=dna. Now save it with the extension .nex. The header should now look something like

```
#NEXUS
[ Title trimmed brown primate]
begin data;
  dimensions ntax=5 nchar=896;
  format missing=? gap=.  datatype=dna interleave=yes;
  matrix

[!Domain=Data;]
```

(The stuff in [] is just commentary and titles and doesn’t really count.) Now save it with the extension .nex.

Go to Phylemon, click on MrBayes. Now you think you’re going to load your .nex file just like you loaded the .aln files before. But, mysteriously, this doesn’t work. What does work is to copy and paste your .nex file into the window in Phylemon where it says “or enter your data”. Then hit run. With any luck the program will read your data and you will see a bunch of stuff ending with

```
Successfully read matrix

   Exiting data block

   Reached end of file
```

If you don’t then there will be a list of things wrong and you will have to figure out how to fix them. Sigh. But let’s suppose you have arrived at this point safe and sound. Now you need to enter a mrbayes command, namely:

lset nst=6 rates=invgamma

and hit execute command. This sets the model parameters. (GTR model, in this case, with gamma distributed rates and a default proportion of invariable sites.) The next command is

mcmc ngen=10000 samplefreq=10

Of course, you can change these numbers if you like. In any case, ngen should be large and samplefreq something noticeably larger than 1,. Every so often you will see the summary statistics average standard deviation of split frequencies. If this number is small, say, less than .01 you are good to go. If not, you need to run the chain longer. To do that, type in the number of additional generations to run – maybe another 10000? You might have to do this more than once.

Next command is

sump burnin=250

This sets the burnin to 25% of the number of generations, which is reasonable. You can figure out what to put here based on your ngen number.

The final command is

sumt burnin=250

using the same number you used for the sump command. This will give you the cladogram and phylogram. Yay! Now type quit and you will get a list of output files, including tree.nw or something like that. As before, send to ETE for drawing. If you can't see the numbers on the tree, consult your phylemon cladogram and phylogram for the same information.

Whew!!

Help? Not much available – or, rather, too much. The help button just takes you to user manuals so you get lots of information. Possibly helpful. Or you e-mail me with your problem and we both agonize.