

Phylogenetic Reconstruction

Lab 3
BIO 319: Fall, 2005
21 Sept

Overview

The purpose of this lab is to help you to learn to use tools to construct phylogenetic trees from aligned sequences. We will use the Phylip package of programs to do this. It is a command-line driven, freely available, Unix-based package written by Joe Felsenstein at the University of Washington. For this lab we will give you a data file containing a set of 28 protein sequences (the entire yeast Ras super-family) to start with. We will use ClustalW to align the protein sequences from these genes and then the next step will be to use Phylip to construct a phylogeny from the sequences, and then draw and print out a tree.

Some potentially helpful websites for this lab:

- EMBL's ClustalW alignment: <http://www.ebi.ac.uk/clustalw/>
- The Phylip website with documentation: <http://www.phylip.com/>

You should answer all the questions in this lab in your lab notebook. I have put the links above and links to other helpful resources at:

<http://www.cs.williams.edu/~stacia/courses/cs319>

1 Multiple Alignment of Ras Super-family

The first step will be to use ClustalW to align the amino acid sequences. You'll be using Unix for this lab, so open up Xwindows.

1. Create a directory for this lab in your home directory by typing the following command at the prompt:

```
-> mkdir lab3
```

Then cd into the directory (cd lab3).

2. Next copy the data file from my home directory:

```
-> cp ~stacia/yeast_ras_fam.fa .
```

3. Open Safari, go to the ClustalW website and create a multiple alignment of the sequences in this file.

BE SURE TO CHANGE TWO THINGS: the output format should be changed to "phylip" and the output order should be changed to "input."

4. Save the alignment file to your lab3 directory (ctrl-click on the link to the file).

2 Construct a Phylogeny From Aligned Sequences

This next part you will be doing from the unix command line, so return to your Xwindow. Make sure you are in the lab3 directory (use `pwd`).

1. First, rename the ClustalW file to something more reasonable:

```
-> mv clust<TAB> clustal.aln
```

where <TAB> is the tab character.

2. Next, start the protein sequence maximum likelihood (ML) program by typing `proml` at the prompt. It will tell you it can't find the input file "infile" but this is normal. Enter the name of your ClustalW file (`clustal.aln`) and hit return.
3. Now a menu should appear on the screen. Examine these options see if you can decipher what they mean (such as "Rate variation among sites?"). For this exercise, we'll leave the default settings, so type "Y" and hit return. **NOTE: This menu at the start of all Phylip programs is where you choose to *outgroup* a sequence by selecting the O option.**
4. Proml will now run ML on the 28 protein sequences. This takes about 10 to 15 minutes to run (and shows the progress it is making).

3 Print out the Phylogeny

There are some hoops we must jump through to get a picture of the phylogeny. An ASCII version of it will print to the file "outfile," but we'll get one we can print out using the program `drawtree`.

1. First, rename the output files since Phylip always puts the output in the files "outfiles" and "outtree."

```
-> mv outfile ml.out
-> mv outtree ml.tree
```

2. Then copy the font file to your directory:

```
-> cp /Applications/phylip3.65/src/font2 fontfile
```

3. Now run `drawtree` by typing the command at the prompt:

```
-> drawtree
```

4. Once again it will say it can't find the input file. Enter "ml.tree" and hit return, then type "Y" to accept the options and hit return again.
5. A new window will appear with our tree (granted, a bit ugly), and then you should save that tree by clicking the "File" box and choosing "Plot."

6. Now you'll have the file "plotfile" in your lab3 directory which can be printed. First though, rename the file:

```
-> mv plotfile ml.plot
```

7. Sent it to the printer with the lpr command:

```
-> lpr ml.plot
```

4 Compare Three Methods

In this section, we will compare three trees constructed using different methods but using the same dataset. The dataset is the sequences you extracted in last Wednesday's lab (the Ras2 sequences from different organisms).. The first tree will be the "guide tree" constructed by ClustalW, the second one will be one that you construct using `protpars`, the Phylip maximum parsimony program, and the third will be an ML tree (constructed as above).

If you need to, you can copy the Ras2 sequences from my directory:

```
-> cp ~stacia/ras2_all.fa .
```

The Clustal Guide Tree (Neighbor Joining)

1. The first step is to run the sequences through ClustalW to get the guide tree.
2. Save the aligned sequences (ctrl-click on the link).
3. To save the guide tree, you'll have to use the Mac "Grab" program which is located in the Utilities folder inside the Applications folder. Double-click on the Grab icon (scissors)..
4. Bring the window with the picture of the tree to the front and make sure you can see all of the tree on the screen.
5. Click on the Grab icon in the desktop, then select Capture Selection from the Capture menu at the top of the screen.
6. Then drag the mouse to select the area of the screen you want to capture. When you let go of the mouse, Grab captures what you selected.
7. You can save this file out of Grab as a Tiff file which can be dragged and dropped into the browser.
8. Alternatively, you can print the tree out of Grab.

The Maximum Parsimony Tree

1. Maximum parsimony is a program in Phylip, so you'll go back to your Xwindow and type `protpars` at the prompt. You will see the now-familiar error message about the missing file. Enter the name of the alignment file, then accept the default settings.
2. Then rename your outfile and treefile to `mp.out` and `mp.tree`.
3. Use the `drawtree` program to get a printout of your maximum parsimony tree.

The Maximum Likelihood Tree You will construct the ML tree for this dataset in the same way that you did with the Ras family dataset in Section 2.