

Lab 04: Introduction to PAUP*

Today we will be learning about some of the basic features of PAUP* (Phylogenetic Analysis Using Parsimony [*and other methods]), a phylogenetic program developed by David Swofford. Supposedly it is pronounced “pop star”. PAUP* can infer phylogenies using distance, parsimony, and likelihood. Today we will run these types of analyses using a sample mtDNA data set. We will learn how to use more features in PAUP* in later labs.

Exercise 1: Basic PAUP

Before we get started, a word about preparing files for use in PAUP. PAUP takes a nexus (.nex) file as input. You have now worked with this file format several times. You can edit these files with a text editor (such as Notepad or TextWrangler).

Open PAUP*.

To load data, type “execute <filename>”. Today we’ll be using the file “primate-mtDNA.nex” or if you have your own data aligned and in a Nexus file, please use that. You may have to include the full path for the file, depending on where you saved it.

On a Mac: you can drag-and-drop a file into the terminal, and this will paste the text of the full path [Careful as spaces in file names are not recognized]

On Windows: the Windows Explorer file viewing program will show you the whole path just like a web browser shows you a URL. (You may have to turn this address bar on, view the View menu I think)

PAUP will tell you a few things about the file (how many taxa and characters it has in it, what nucleotide ambiguity codons are being used). For instance, right now “R” is treated as either A or G for analytical purposes.

Let’s find out a little bit about our file. First we’ll get a brief summary of the character status. Type:

`cstatus`

then press enter. (Menu equivalent **DATA** menu, **Show character status (Brief summary)**).

Question #1: What is the optimality criterion currently set to? Are these characters ordered or unordered?

Type:

`showdist`

This is useful because it tells you how many differences there are between any two taxa. Taxa that have 0 distance between them can’t be differentiated by PAUP. Since they have equally parsimonious arrangements, they can make your analysis run for a very long time without getting closer to an optimal tree. So it is good to know in advance how many you are dealing with.

Question #2: Enter each of the following commands into PAUP and tell me what they do.

showmatrix
tstatus

Here are some other optional commands that you can experiment with:

taxset

which can be used to create sets of taxa. It is used as:

taxset NewNameForSet = ListOfTaxa

Try typing:

paup> taxset homs = Homo_sapiens Pan Pongo

paup> taxset other = 5-7 9 10

charset

charset does the same thing except for characters. PAUP* automatically includes a few character sets:

Constant: all invariant characters

Gapped: all characters with a gap for at least one taxon.

Missambig: all characters with a gap or ambiguous character for at least one taxon.

Remainder: all characters not previously referenced in the command.

Uninf: all characters that are constant as well as autapomorphic.

Pos1: all characters defined by current CodonPosSet as first positions.

Pos2: all characters defined by current CodonPosSet as second positions.

Pos3: all characters defined by current CodonPosSet as third positions.

Noncoding: all characters defined by current CodonPosSet as non-protein-coding sites.

Try typing:

charset firsthalf = 1-400

include and exclude

can be used to exclude characters from the analysis and then put them back in.

Try:

exclude firsthalf

then type cstatus to see what's different.

delete and undelete

delete and undelete can be used to exclude taxa from the analysis and put them back in. Try:

paup> delete homs

Taxon-deletion status changed:

3 taxa deleted

Total number of taxa now deleted = 3

Number of nondeleted taxa = 9

?

keep in mind that you can always type a ? after a command to see brief

help about it:

delete ?

Before you move on to the rest of the exercises, include all and undelete all so that your results will reflect all twelve taxa and all 898 characters.

Exercise 2: Defining an Outgroup

It is easy to set an outgroup using PAUP. For most of the calculations, it is not important to set an outgroup before you do your analysis – Paup calculates unrooted trees and does not consider character polarity to be set by the outgroup. To set “Lemur catta” as the outgroup, type:

```
outgroup Lemur_catta /only
```

Outgroup status changed:

```
1 taxon transferred to outgroup
Total number of taxa now in outgroup = 1
Number of ingroup taxa = 11
```

This sets the Lemur as the only outgroup. If you don't type /only PAUP will add a new outgroup to an existing outgroup. Try it:

```
outgroup gorilla
```

Outgroup status changed:

```
1 taxon transferred to outgroup
Total number of taxa now in outgroup = 2
Number of ingroup taxa = 10
```

That's silly. Let's move the gorilla back to the ingroup:

```
ingroup gorilla
```

Outgroup status changed:

```
1 taxon transferred to ingroup
Total number of taxa now in outgroup = 1
Number of ingroup taxa = 11
```

And check what happened:

```
tstatus
```

Taxon-status summary:

```
Original data matrix contains 12 taxa
No taxa have been deleted
Designated outgroup taxa:
  Lemur catta
```

Exercise 3: Analysis

Now to do some actual analysis. For today, we'll just use a distance analysis. In other labs we'll learn more about what the settings mean and how to change them.

First, change your working directory to a folder on your desktop:

```
paup> cd /Users/"yourname"/Desktop/IB200_Lab4
```

I think creating a new folder each time you work with a program is important to keep track of which files were created on which day, etc. You'll end up with many files in this class and keeping on track of where they all are is key. Designating everything to the Desktop can become a headache very quickly.

Distance

Our first analysis will be a distance analysis. PAUP can run multiple types of distance analyses, but for today we'll use the default settings. First, we need to set the optimality criterion to distance:

```
paup> set criterion=distance
```

```
Optimality criterion set to distance.
```

Then, we can start the heuristic search:

```
paup> hs
```

Paup will now tell you about what it is doing, by reporting which options are set to what, and then after a few moments it will tell you what the results of the search were:

```
Heuristic search completed
  Total number of rearrangements tried = 546
  Score of best tree(s) found = 1.09002
  Number of trees retained = 1
  Time used = <1 sec (CPU time = 0.00 sec)
```

This means that Paup looked at 546 rearrangements, and found one best tree. If you're using your own data obviously you'll have different results. Now let's look at our tree:

```
paup> showtrees
```

Paup will show you the one tree that it found, in an ascii display. Then you can use savetrees to write your tree to file:

```
paup> savetrees file=distree.tre
```

A new tree file called "distree.tre" should appear in the folder you designated.

Exercise 4: Examine Your Tree

Copy the primate-mtDNA.nex file (or a Nexus file with your data) and paste it your folder for today. Open Mesquite and open this file.

To import the tree file, first go to **Taxa&Trees> Import File with Trees> Include Contents...** then choose “**distree.tre**”

Select **View Trees**.

Question #3: Take a screen shot of your tree in Mesquite and send it to me.

If you save the changes to your Nexus file and then open this in a text editor, you will notice that there is now a TREES block.

Enjoy your new-found PAUP* skills!