

Lab 02: Introduction to *Mesquite*

Tree Functions

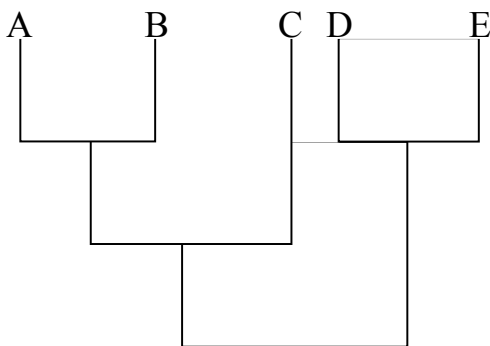
Last lab we had a brief introduction to the data matrix functions of *Mesquite* and learned about the Nexus file format, which allows *Mesquite* to be integrated with many other phylogenetics programs. Today we will learn some basic tree functions of *Mesquite*, which we will use a lot as the semester goes on.

Editing Trees

Today we'll be working with the **Amblygnathus.nex** file that you downloaded from the web site. This is a data set from the *MacClade* examples containing characters for a group of North American ground beetles.

Before we open up this file in Mesquite, let's take a look at the NEXUS file by opening it in a text editor. The file begins with the familiar #NEXUS and has the blocks that we worked with last lab. BEGIN DATA indicates your **data block** and this is in place of the **taxa block** and **character block**. Many programs use a **data block** instead of a taxa block and a character block. Most programs, but not all, are flexible and can use either format. The data block contains all the same information as the other two blocks, except it does not need taxalabels as they are described in the matrix

You will also see BEGIN TREES, which is the start of the **trees block** that contains your phylogenetic tree information. A nexus file will only contain a tree block, when it is necessary to import a tree into a program. Tree blocks often start with a **translate** command, which is required for a number of programs, but not all. It is a list of consecutive numbers followed by the names of the taxa that those numbers will represent. This is followed by a **tree** line (in **Amblygnathus.nex** there are 21 different trees), which describes the tree in **Newick format** using the numbers assigned in the translate command for the names of the taxa. Newick format is a commonly used way of representing tree topologies as text. Put simply, monophyletic clades are surrounded by parentheses and sister clades are separated by commas. Such that the tree



would be written as (((A,B),C),(D,E)). Newick format can also contain information about branch lengths (after colons) and node names (after closed parentheses).

Open the **Amblygnathus.nex** file in *Mesquite*. You will see that there are 29 taxa and 28 characters associated with this file. There are also 21 trees. Select View Trees. A new tab will appear with a tree of 29 taxa in it. On the top left side there is a menu button



that allows you to scroll through the 21 different trees stored in this file. Scroll through these trees to see some of the differences in topology. In the left-hand column of the window are several tools for modifying these trees. I want you to explore these different functions, so I know this is somewhat tedious, but please do the following to Tree #1:

1. Redraw “corvinus”, “cephalotes”, “gigas”, “reichardt”, “lucidus”, and “janthinus” as a triangle.
2. Color this triangle clade red.
3. Name this clade “Red-wings”
4. Remove the clade containing “iripennis”, “mexicanus”, “interior”, “evansi”, “woodruffi”, and “puncticollis”
5. Create a polytomy for “fricki”, “panamensis”, “tikal”, “whiteheadi”, and “subtinctus”
6. Reroot the tree with “bicolor”
7. Select the clade composed of “seriatoporus-g”, “ruficollis-g”, “ellipticus-g”, and “amaroides-g” [the clade should be dark black while the surrounding parts of the tree are greyed out]

Question #1: Send me a screen capture of what Tree #1 looks like after completing 1–7

That's enough for now, but many of the other tools can make tree manipulation more efficient and allow you to do some things that will be useful for later labs so I encourage you to explore those options as well.

Viewing Results as Text

All these pictures are very nice, but often you want a more utilitarian output. Every window in *Mesquite* has a text tab at the top of it, which displays the same information that you see in graphic form as text. Hit the **Text** tab at the top of this **Tree Window**. The text tab is particularly good for likelihood, because you can see the reconstructions as a table. Another useful feature is that you get the tree in Newick format. What does Newick format mean again? Why don't you tell me.

Question #2: Describe what Newick format is. Describe what FASTA format is. Describe what NEXUS format is.

Ancestral Character Reconstruction Using Parsimony

Discrete Character States

I've already mentioned that *Mesquite* is not a program for building trees or for creating alignments of sequence data. The whole purpose of *Mesquite* is to analyze how characters change on trees. (As we saw last lab, it is also a wonderful tool for storing and examining your data matrices.) There are two very broad categories of data: discrete and continuous. A discrete character might be "Mandible indentation" and the states could be "present" or "absent". A continuous character might be something like body size. Let's look at some discrete characters.

You are still probably in the **Tree Window**. From the **Tree Window** go to **Analysis > Trace Character History**. Select **Parsimony Ancestral States** and hit **OK**. (We will discuss the different options here later on in the semester.) Neat! Now each of the morphological characters in the data matrix is mapped onto the tree with the ancestral states colored along the branches. You can go through the different characters by clicking on the arrows in your **Trace Character** box in your **Tree Window**. Scroll on over to Character #5 and look at the branch connecting "darlingtoni" and "bicolor". There should be several colors along this branch. Any ideas as to why that is? If you have no idea, discuss with your fellow lab mates or ask me. This is called an ambiguous reconstruction. [FYI: Character #5 of the *Amblygnathus* data set is the number of spines on the internal sac of the male aedeagus (ie. penis). Because beetles are awesome like that and have spines on their genitals.]

Question #3: What is an ambiguous reconstruction?

In *Mesquite* all the branches are shown with the colors of all the possible states that they could have under parsimony [again, we'll discuss parsimony soon!]. When you move the cursor over a branch, you will see the character state at the bottom of the **Trace Character** box. On my computer I had to move this box up to see its bottom.

When you change a character state in your matrix the results will automatically appear in your tree. Go to the **Character Matrix** window and change the character state of the character you are currently tracing in your tree window for one of the taxa. The character state will automatically change in the **tree window**. This is a general feature of *Mesquite*, whenever you change something in one window it will effect what is happening in all the windows that are related to it.

Making Pretty Trees

You can change the colors of the character states by double clicking on their colored square in the **Trace Characters** box. Select a color and the tree will change. *Mesquite* tends to use black and white as a default but it is often easier to visualize if you use different colors.

Let's look at different shapes of trees. Go to **Drawing > Tree form > diagonal tree**. The shape of the tree will change. Try some of the other shapes. You should do the rest of the lab in **Balls and sticks**, because it allows you to see equivocal states more clearly. There are many other options under the **Drawing** menu that allow you to change the appearance of your tree. Try changing the **line size** and **line style** to make your tree more appealing.

Question #4: Make your tree beautiful, take a screen shot, and send it to me.

Changing Parsimony Models

It is also possible to change the parsimony model under which a particular analysis is run. Go to **character 5**, which has several character states. Currently the model is set to **ordered**. Select **trace>parsimony model>unordered**. This changes the model for this particular character. Does it affect the character state reconstruction? You might notice that there are two other models: Dollo and irreversible. Unfortunately they don't work yet, but they are there so that you can set the model for *MacClade* or *PAUP**. Hopefully, in the near future, they will work in *Mesquite* too.

Question #5: In lecture we discussed ordered vs. unordered characters – what does this mean? How is this different from character polarity?

A Few Extra Things If You Have Time

You can also construct your own parsimony model by building a step matrix. Go to **Characters>New Character Model>Step (Cost) matrix**. Make up a name for the model and hit **OK**. Now you need to set the maximum state by selecting **Step matrix>Set maximum state...** The number of your maximum state depends on your character. Since you start numbering states at 0 the maximum state will be one less than the total number of states for a character. For example, if you're still on Character 5, your maximum state should be three. Now click on the individual cells with the text select tool to make changes to your step matrix. When you go back to the tree window you should find your step matrix under the list of parsimony models (**trace>parsimony model>**). Select it and look at the results.

Trace All Characters

It is also possible to create a table showing the character state reconstructions for each character on each branch. Select **Analysis>Trace all characters**. Then select **Stored Matrices** and hit **OK**, then **Parsimony Ancestral States** and hit **OK**. A **Trace All Characters** window will appear showing the reconstruction of character states for every node that you have selected.

Discrete Character State Reconstruction Using Likelihood

It is also possible to reconstruct ancestral states for discrete characters using likelihood. But we'll wait to do this another time.