

## **Lab 01: Introduction to *Mesquite***

### **Data Matrices**

Today we will have a brief introduction to some of the features of *Mesquite*, and learn about the Nexus file format, which allows *Mesquite* to be integrated with many other phylogenetics programs. *Mesquite* was developed by Wayne and David Maddison as a tool for interpreting phylogenetic information. It is available for Mac, Windows and Unix at no charge from <http://mesquiteproject.org/mesquite/mesquite.html>. The strengths of *Mesquite* are creating and editing data matrices, examining the distribution of features on a phylogeny, and testing hypotheses about character evolution. This program will be a great way to explore the data you collect for your final project, especially if you are not quite yet comfortable with *R*. However, *Mesquite* is NOT the program you should use for any of your tree-building. You need to implement maximum parsimony, maximum likelihood, or Bayesian inference in another program such as *PAUP\** or *MrBayes* to build your trees.

*Mesquite* is modular, which means that the program is set up as a bunch of modules that all do different functions, such as draw a picture or do a parsimony analysis. Some modules use other modules and the modules are used in combination to perform an analysis. Thus *Mesquite* is very flexible and capable of doing analyses that it was never intended to do. Unfortunately it also means that *Mesquite* can be difficult to use, because it is not always clear where to find the appropriate command in its menus. Today we will learn some basic properties of *Mesquite*, which we will use a lot as the semester goes on.

### **Creating and Editing a Data Matrix**

*Note: Enter your own project data in this exercise if possible. If you don't have any character data yet, you can make it up.*

Open *Mesquite*; two windows will open for Macs (three windows for PCs). The Start Up Window will be open just as the program boots. The ***Mesquite* Log** gives a constant narration of what the program is doing. It is sometimes useful to look at this window, and it is the basic window for controlling files, but most of the work in *Mesquite* is done elsewhere. Closing this window is the fastest way to quit the program.

Select File>New. Choose a location for your file, where you'll be able to find it in the future and give it a name with a ".nex" file designation (eg. "Tracis\_temp.nex"). A New File Options window will appear. Check Make Taxa Block, give it a name (eg. "Beetles"), stay with 3 taxa, check Make Character Matrix, and hit OK. In the New Character Matrix window make up a name (eg. "Traci matrix"), and stick with two characters. The options in the Type of Data menu are pretty self explanatory, and we will study most of them later in the semester. For this lab we'll stick with Standard Categorical Data. Hit OK.

A Project window will open showing an empty **Character Matrix**. Here you can edit and add characters and character states. There are several tools along the left side that allow you

to manipulate the matrix. When you hold the cursor over each of the buttons, a description of what it does appears at the bottom of the window.

To the far left is a panel showing some of the other viewing options. Click the **Taxa** button. The **Taxa Block** can be used to edit information about taxa. Let's just change their

names. Click on the edit tool (the little cursor: ) and use it to select each of the taxon names, and name those taxa.

Now go back to the **Character Matrix** by clicking the tab at the top of the window, or the button on the far left. Now let's add some data to the matrix. Select the edit tool again and click on the various cells in the matrix. Since this is a categorical data matrix, all the characters must be consecutive integers starting with 0. So fill the matrix with a bunch of 0s and 1s; if you want to go really crazy, you can even add some 2s.

***Tools.** To the left of the data matrix are a series of tools.*

Move the mouse over each of them, and a text description will appear below the character matrix. Mesquite matrix editing makes a lot more sense when you know what these tools are. To get you acquainted with these tools, do the following:

1. Add three new characters
2. Add two new taxa
3. Fill in the empty cells of the matrix, but don't use the edit tool. Use the copy tool and the fill tool instead.

A matrix filled entirely with 0s and 1s is often not the easiest thing to interpret; we would be much better served if the characters and character states were more descriptive. At the bottom left are five small buttons next to a blue *i* that look like little windows. Select the one second from the right, the **Show State Names Editor Window**. Select the edit tool and use it to give names to some of the characters and their states. For example maybe character 1 is "common sense" and its states could be 0, "absent", and 1, "present". Now go back to the **Character Matrix** by clicking the tab at the top of the window. As you can see all the 0s and 1s have been replaced with actual words. One thing to note is that when adding or editing character states in the matrix you still need to type in 0 and 1, even if the states now have more interesting names.

*(add as many of your characters as you like, then continue)*

**Question #1.** Take a screen shot of Mesquite and your completely filled in data matrix.

*Notes:*

- \* on Macs, Command-Shift-4 lets you send an arbitrary chunk of the screen to a PNG image file on the Desktop;
- \* on Windows, it should be ALT+PRINTSCRN for whatever window is active (copies the image to the Copy/Paste buffer I think, so paste into e.g. Word or an email). PrintScr captures the whole screen.

If you are working with morphological data, *Mesquite* also has some options for visualizing your characters and their states. Make sure you are still in the Character Matrix. Besides the left-hand column options, you may have noticed the extensive menu bar across the top. These menu options will change based on which window you are in (eg. **Character Matrix** vs. **Taxa Matrix** vs. **Tree Window**). Go up to **Matrix** and then scroll down towards the bottom to select **Show Annotations Panel**. Now click on any of the character boxes, then click the small blue addition button. You can now make comments and even add images to help organize your data. This may be helpful for your term projects. There is also an additional package for *Mesquite* that you can download, called SILK.

**Question #2.** Add a linked image into Mesquite and take a screen shot.

Keep in mind that these images are not embedded within *Mesquite*, so if you move an image or rename it, *Mesquite* will not be able to find it again. We have only looked at a fraction of the tools available for editing the data matrix, and you might want to spend some time either exploring the other tools on your own, or going through the *Mesquite* tutorials.

## Nexus file basics

Let's save all of our work **File>Save File**.

Now navigate to the file and open it in a text editor (don't use Word); we're going to explore a Nexus file.

Nexus is a format used by many different phylogenetics programs. By having a common file format the results of one program can be viewed or analyzed in another (although this does not always work as well as one would hope). You can manipulate a Nexus file directly in a text editor, but often it is better to use a program like *Mesquite* and save the results in a Nexus file to be used in another program.

Every Nexus file starts out with “#Nexus” and then is followed by a brief description of the file surrounded by brackets. This is followed by the actual data of the Nexus file, which are organized into several **blocks**. Each block starts with a line “BEGIN *BLOCK NAME*,” and finishes with an “END;”. The lines between, which hold the actual data, are often indented. Each program creates and uses different blocks with information and commands that are particular to it, but there are several block types that are almost universally used and contain the most fundamental information for phylogenetic analysis.

The **taxa block** [ie. “BEGIN TAXA”] is pretty much just a list of taxa. It usually contains a name for the block, a command describing the number of taxa, and a list of the **taxalabels** (in other words the names of your taxa).

The **characters block** [ie. “BEGIN CHARACTERS”] contains your data matrix. First there is a **title** line and then a **dimensions** line, which can contain information about the number of characters and taxa. This is followed by the **format** line, which describes the data in the matrix. **Datatype** determines the basic class of data: discrete; continuous; protein; or DNA. After that there are several commands, which describe what symbols are allowed and what they represent. For example, does a “-“ mean unknown or missing. After the format line is the actual **matrix**. Each line of the matrix starts with the name of the taxon represented and is followed by

a series of symbols representing the character states for the various characters in the matrix. Each taxon line must have exactly the same number of characters in the same order. The characters block may also contain other information such as **charlabels** (names for the different characters) and **statelabels** (names for the different character states for each character) or **charstatelabels** (both types of info combined in one command).

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

If we had some trees to go along with these data, there would be a **trees block**, but we'll save that for next lab.

**Question #3.** Send me your NEXUS file with your morphological data.

I have not given you a detailed description of the Nexus format. If you want to change a Nexus file by hand, I would recommend that you use another Nexus file as a template.

Try to download some sequence data from Genbank (hopefully for your group). Download into FASTA format, and import it into Mesquite. I am not giving you step-by-step instructions, but will help you in class. The skill of downloading and importing data from different databases and formats is important, so figuring it out for yourself is useful. (You do not have to align it today, however). Rename the taxa to make sense.

**Question #4:** Send me the NEXUS file of the loaded molecular data.

There are more tools for you to explore with this molecular data. Try each of them out and see what they do to your matrix. At the bottom left again there are five small buttons next to a blue *i* that look like little windows. Select the one that looks like a little pixelated bird. This is **Bird's Eye View** and is a useful tool for quickly visualizing your molecular alignment. We didn't previously align the sequences today, so they should look pretty messy. Use some of the tools in the left-hand column to manually align a small portion of these sequences. If you have no idea what I'm talking about, just ask and I'll be happy to help.

**Question #5:** Send me a screen capture of the manually aligned data.