

Lab 1: Introduction to Character Matrices

Introduction

Today we will learn the basics of putting together a character matrix. The main tools we'll be using will be MacClade and its descendant, Mesquite. These two programs make it easy to put together a data matrix with all the great characters you'll soon be coming up with. Both programs were developed by Wayne and David Madison, twin brothers who are both professors of phylogenetics. MacClade, the original program, is, as the name suggests, only for Macs. After many years of selling MacClade commercially, and updating it in their spare time, the Madisons decided to branch out and create Mesquite, a free, open-source, Java-based program that works on any platform. Neither program is used to actually infer phylogenies (although MacClade does have some very basic, outdated ways of doing this.) Instead, they are used for creating and editing data matrices, examining the distribution of characters on a phylogeny, and testing hypotheses about character evolution. Either program makes a great starting point, because they allow you to collect and organize characters and taxa (I mean, semaphoronts.) Another reason we're starting with these data-matrix programs is that they're a great way to explore the data you're about collect for your term paper project. Both MacClade and Mesquite can be found on the computers we'll be using today, by following the path: **Macintosh Hard Drive > Applications > 200A.**

MacClade

To open MacClade, go to **Macintosh Hard Drive > Applications > 200A > MacClade.** When you double-click on the multi-colored phylogeny icon, MacClade will load and after a few moments, a window will ask you what file you'd like to open. Today, we'll be using an example file that came with Mesquite, just so you all can see what a filled-in, working data matrix looks like. Follow this path to open the file: **Macintosh Hard Drive > Applications > 200A > Mesquite_Folder > examples > Ancestral_states > 01-DataMatrix.nex.**

MacClade is designed around two main windows, the tree window and the data editor window. Two other windows we will see today are the chart window and the character status window, although they are of less importance. The first window you'll see today is the tree window, which shows that this file already has a phylogenetic reconstruction stored with it. Of course, a newly-made data matrix won't have any phylogeny stored with it yet, but we'll get to that later. The tree is shown with scaled branch lengths and character reconstructions plotted on it. This can be changed using the display menu, which includes such options as "change tree type" and "change font." Feel free to explore some of these options and see how the way the tree looks changes.

Now, go to **Windows > Data Editor**, and switch to the data editor window, which is what we're focusing on today. The data editor window allows the entry, editing, and manipulation of phylogenetic data. Both qualitative and quantitative data can be entered, and rows, columns, and blocks of data can be manipulated. The data editor also has settings that make it easier to align molecular sequence data by hand, such as the ability to translate any given sequence into its corresponding amino acids. There are some tools to automatically align sequences, but they don't work very well. The data editor also includes options that allow weighting and ordering schemes to be designed and implemented. The character status window provides a brief summary of character weights, states, ordering, etc., and allows quick manipulations of these assumptions. MacClade saves files in nexus format, which can be read into programs that infer phylogenetic trees, such as PAUP*.

When you look at the data editor, you will see the names of taxa listed vertically at left, and the names of characters listed horizontally at the top. In the middle, you'll see a matrix filled in with numbers, such as 0, 1, and 2. These represent character states, such as "mirrors absent," "very slight mirrors," and "distinct mirrors." By convention, 0 usually represents "absent." You'll also see some cells have more than one characters state listed, such as "0&1." This means the character varies within the taxon.

MacClade includes a number of handy tools for working with your data, such as **Characters > Character list options > Show % missing** and **Characters > Character list options > Show % gaps**. These can be used to show the percent of a given character that is missing from a particular taxon, or to show the percent of that character coded as a gap. It also has tools for working with molecular sequences, such as **Set Codon Position** and **Translate Amino Acids**. You can also include and exclude characters, set parsimony models such as ordered parsimony, and set character weights. We're now done with MacClade, at least for the moment. Go ahead and close the file and quit.

Mesquite

Now, open up Mesquite: **Macintosh Hard Drive > Applications > 200A > Mesquite_Folder > Mesquite OSX (Java 1.5)**. At first, Mesquite may seem a little more difficult to use than MacClade, but it is really your friend. That said, it behooves you to save early and often, as the Java virtual machine can be a little, um, unstable. The computers in lab are already set up to run Mesquite, but if you want to use it on your own computer, you will first need to download two things: 1) a recent version of the Java virtual machine, which can be found at <http://java.sun.com/javase/downloads/index.jsp>, and 2) an installer for Mesquite itself, found at <http://mesquiteproject.org/mesquite/download/download.html>. For the sake of stability, it is best to update your Java virtual machine and Mesquite programs regularly – both programs

and their interaction have improved a lot in the time I've been using them and they will probably continue to get better.

Now, while you've been reading this, Mesquite is loading. Two windows will open. You never have to use the first two windows. The Mesquite Log gives a constant narration of what the program is doing. It is sometimes useful to look at this window, but not most of the time. The second window you'll see will be a small window that shows all the modules that Mesquite is loading. Mesquite is designed to be like a Mr. Potatohead, with different bits of software that can be stuck on to the central program for different purposes. A third dialog may pop up, with a message welcoming you to Mesquite and asking you to join their mailing list. Just click ok and ignore it. (On PCs, there is a fourth window that opens first, the command window is just black with letters. It is the platform on which Mesquite runs, and you don't need to interact with it.)

Which brings us to the forth window, the potato itself, aka the Mesquite Projects and Files window. This is the central window through which you can control your projects. Right now you will just see a green square labeled "current projects." Go to **File > Open File**, then open the same file you opened with MacClade before: **Macintosh Hard Drive > Applications > 200A > Mesquite_Folder > examples > Ancestral_states > 01-DataMatrix.nex**.

The character matrix should be the first window that comes up, but if not, you can always go to the projects and files window and open it there (just double-click). Take a look at the character matrix. It is much like the equivalent window in MacClade. 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. You'll see that Mesquite is showing the whole name of each character state. You can turn this on or off using **Matrix > Display > Show State Names**. You can enter and change the state names by going to **Matrix > Edit state names**. You can also view the cells color-coded by going to **Matrix > Color Cells** and choosing Character States, Annotation, Character Value, Excluded, Footnote Present, or Assigned Colors.

Mesquite has a lot of other windows, including a list of characters and taxa (and virtually anything else you can list) to allow you to keep tabs on your matrix. Mesquite also has a tree window, as well as (due to its modular nature) infinite potential for other types of analysis windows. But today we'll be sticking to the character matrix window.

Making a Data Matrix

Now, what really matters: making your own matrices. Your task, if you choose to accept it, is to make a morphological data matrix for an imaginary clade of animals, the Caminalcules. Working with a partner,

create a new data matrix in either MacClade or Mesquite. (See below if you need help. If saved properly, the .nex files created can be used in either program.) Add 10 taxa to the matrix – you can choose any subset of the Caminalcules that suit you, just follow the numbering from Sokal 1983.



Now, it is time to decide which characters you will use. Deciding what is a homology and what is a homoplasy can be pretty tricky, but you have to start somewhere. Add 10 characters to your matrix, then score each taxa for the characters you've chosen. Remember to try and make your characters independent; characters that are not independent can falsely inflate your confidence in the reconstructed phylogenetic hypothesis. For example, if you score Caminalcule #3 for “fingers present,” and Caminalcule #1 for “fingers absent,” you should not then go and score Caminalcule #3 for “fingernails present,” and Caminalcule #1 for “fingernails absent.” Instead, you should put a dash (gap) or question mark (unknown) for Caminalcule #1, since there is no way to know whether it would have fingernails if it had fingers.

Write a brief justification for each character, along with any special notes, and turn it in to me along with a print-out of your data matrix by next Tuesday. The best way to print your matrix is to open it in a text reader such as Notepad (PC) or TextEdit (Mac.)

The justification for a character should look something like this:

Character 1: *Fingernails* (0) = absent, (1) = present. If long, claw-like fingernails were present, this character was scored as 1. If round black dots were present, it was scored as zero. It was also scored as zero if fingers were present with no black markings of any kind. If fingers were not present or limbs were undivided, it was scored as missing.

To Make New Matrix in MacClade:

Open MacClade. In the first dialog box that opens, you have the option to make a new file by clicking the **New** button. A new empty matrix will appear. (With three taxa and a tree? If a tree pops up, just switch to the data editor window. You can use  to insert more taxa and  to insert more characters.

To Make a New Matrix in Mesquite:

Close any open projects. Go to the ‘File’ menu and choose ‘New.’ A dialog box will pop up, allowing you to name the new file and put it on the desktop. (The computers are set up so that you are not allowed to save stuff anywhere else.) Next, another window pops up – ‘New File Options.’ Make sure ‘Make Taxa Block’ is checked, give it a name, and add 10 taxa (don’t worry, you can change this later.)

A window pops up with a list of unnamed taxa on the left. For now, we’ll just ignore this window. From the menu labeled ‘Characters,’ choose ‘New Empty Matrix...’ Give it a name and make it 10 characters long. At last, the Character Matrix window appears.