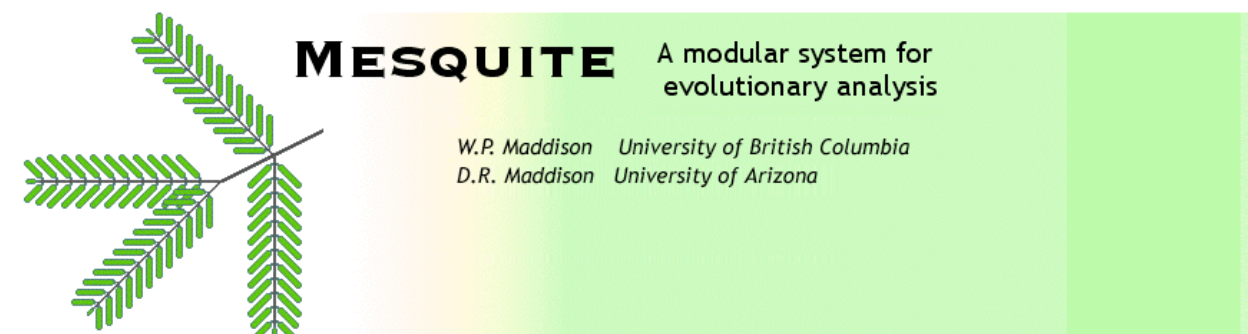


Lab Activity: Simulating Evolution; Turn in pages 5-8

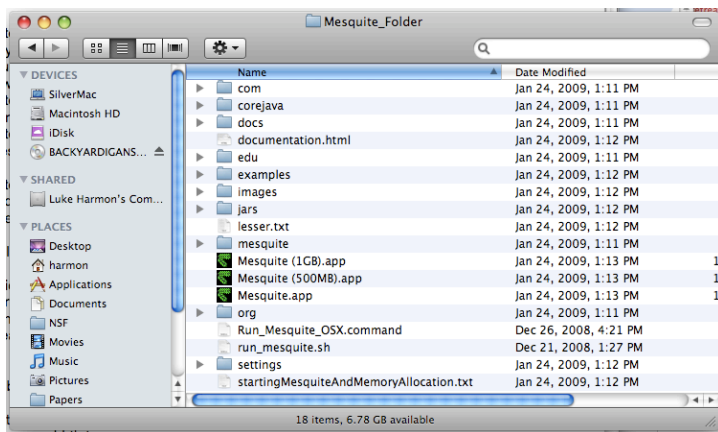
1. Basic introduction to Mesquite



About Mesquite

Mesquite is a free software package for analyzing phylogenetic data. Developed primarily by David and Wayne Maddison, Mesquite is a modular and immensely powerful way to analyze your data. Importantly, Mesquite is constantly being improved, and new analyses are constantly being added. The algorithms implemented by Mesquite are typically very stable and reliable. An additional benefit to Mesquite is that it has a graphical, menu-driven interface. It might be useful for you to know, though, that Mesquite can also be run through script files for more complex or cumbersome analyses.

Starting Mesquite and Creating a New Project

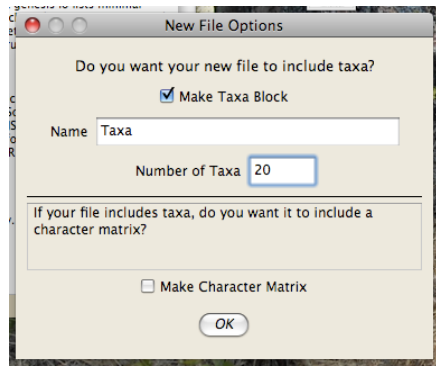


Mesquite is available on its website: <http://mesquiteproject.org/mesquite/mesquite.html>. It should already be installed on your computer. Look in the applications folder, for a subfolder named "Mesquite_Folder."

There are three different versions of Mesquite that you can run; two of them are high-memory versions useful for complex analyses. For this lab, you will just need the basic version, called Mesquite.app.

Double click on the application now to start Mesquite. It will take a short time to load the many modules that are a part of the program.

Mesquite should now be open on your desktop, but you can't do anything with it until you either load or start a new project. To do this, choose **File...new**. This will create a



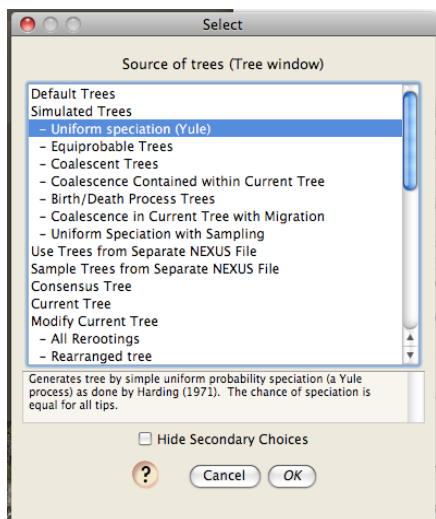
nexus file (*.nex), which is the format Mesquite uses to save projects. You will have to choose a name for this file, and save it somewhere on the computer. Make sure you pay attention to where you are saving the file! After choosing a name, you will have to decide on some options. First, you will be asked if you want your new file to include taxa. You do want to do this. Change the number of taxa for your file to 20, and click on *ok*. You do not need to create a character matrix, we will do this later.

You should now have a new project window open on your desktop.

2. Simulating trees in Mesquite

For the first part of the lab, we will be simulating some phylogenetic trees under both the pure-birth and birth-death models, and identifying some basic characteristics of those trees.

Simulating pure-birth trees

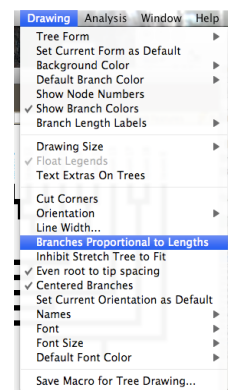


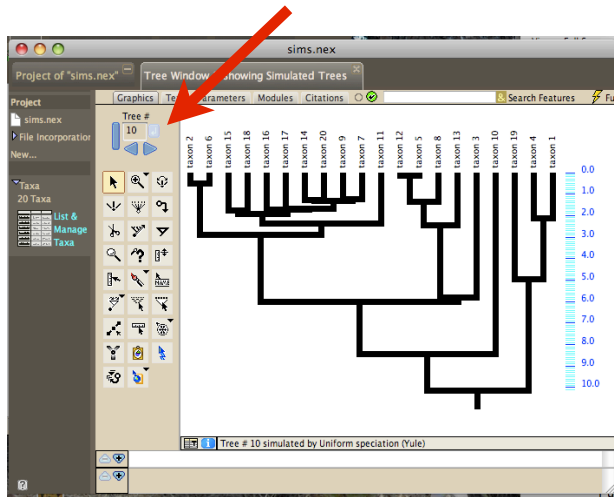
First, you will simulate some pure-birth trees. Remember, under a pure birth model, speciation rates are constant across lineages and through time, and there is no extinction.

To simulate pure-birth trees, choose **Taxa&Trees...New Tree Window**. In the window that appears, choose **Uniform speciation (Yule)** (see picture at left). Keep the default for the total tree depth, 10.0. This will open a new tree window. In this tree window you will see the results of your simulation: a phylogenetic tree grown under a pure-birth model with 20 tips. This tree has 20 tips because you set the number of taxa to 20 when you

created the new project. To grow trees of different size in Mesquite, you would have to change the number of taxa in your project.

You might not know it, but the branch lengths on this tree are not meaningful; this diagram is meant to portray only tree topology, not the timing of speciation events. To see the real branch lengths in the tree, choose **Drawing...Branches Proportional to Lengths**. Now you will be able to see the relative lengths of all the branches in your tree.





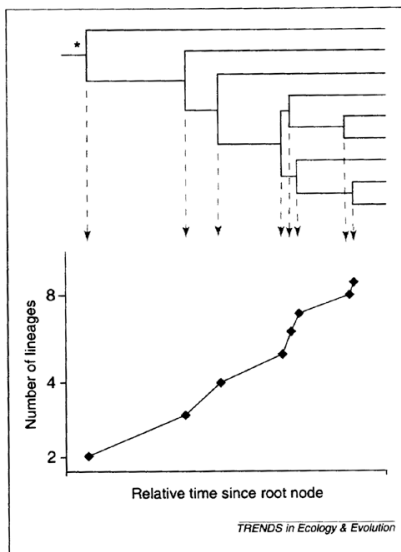
Finally, note the box in the upper left corner of the tree window, where it says "Tree # 1." This box has small arrows just below it. If you click on the arrow pointing to the right, you will go to Tree #2 (this action actually causes Mesquite to simulate a completely new pure-birth tree with 20 taxa). Click on this button a few times until you have created a dataset of 10 simulated trees. Note that you can go back to previously simulated trees using the left arrow.

Now, use this set of ten trees to complete **Activity 1**, in the assignment section of this lab.

Simulating birth-death trees

You can also simulate trees under a birth-death model in Mesquite. To do this, again choose **Taxa&Trees...New Tree Window**. This time, in the window that appears, choose **Birth-Death Process Trees**. In the next window, set the speciation rate to 0.6, and the extinction rate to 0.5. This will create a new tree window that shows trees simulated under the birth-death model. Notice that this tree does NOT show the extinct lineages, only the 20 lineages that survive to the present day and their relationships. Again, choose **Drawing...Branches Proportional to Lengths**, and cycle through the **Tree #** window to generate 10 random birth-death trees.

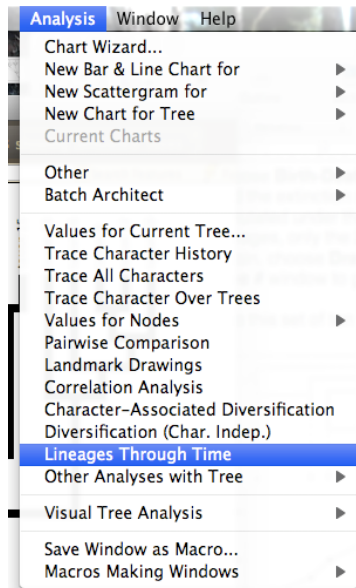
Use this set of ten trees to complete **Activity 2**.



3.Creating lineage-through-time plots

You will now look at some lineage-through-time (LTT) plots for your simulated trees. We haven't covered these in class yet, but they will come up in tomorrow's lecture, and it is worth having a look at some now.

LTT plots show the cumulative number of lineages in a tree as one moves from the root towards the present day. Since we are using reconstructed trees, this number will be strictly increasing. For reasons that will become clear soon, LTT plots are conventionally semilog, with the y-axis on a log scale. The figure to the left shows, for a simple example, how a LTT plot is constructed.



To create LTT plots in Mesquite, you first need to select a particular phylogenetic tree. Go to the tree window that contains your pure-birth trees (for me, it is called “Tree Window 2 showing Simulated Trees.” You can see the generating model at the bottom margin of the tree window). Use the **Tree #** selector to go back to Tree #1. To create the LTT plot, choose **Analysis...Lineages Through Time**. This will create a LTT plot for pure-birth tree #1.

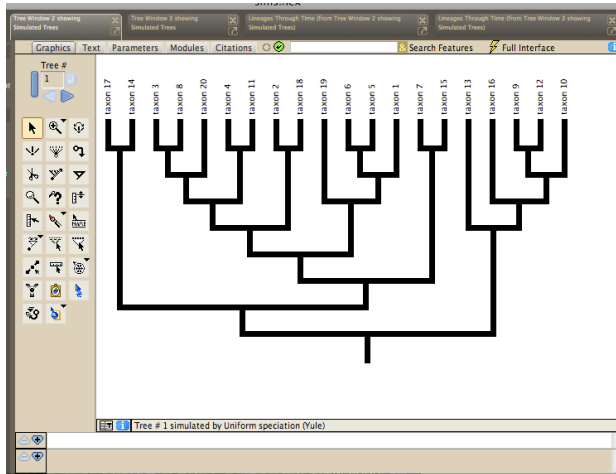
You can quickly switch to other trees to see their LTT plots. To do this, go back to the Tree Window with your pure-birth trees, and select Tree #2. Then, go back to the Lineages Through Time tab, and you will see the LTT plot for tree #2. Use this technique to look at the LTT plots for all ten of your pure-birth trees.

Now, you will want to create another window that will show LTT plots for your birth-death trees. To do this, go to the Tree Window that has your birth-death trees, and use the **Tree #** selector to go back to Tree #1. To create another LTT plot, again choose **Analysis...Lineages Through Time**. This will create a second LTT plot window for your birth-death tree #1. Again, you can switch trees by changing the Tree # selector in the Tree Window, and returning to the Lineages Through Time tab. The only tricky part is keeping track of which window corresponds with which set of trees! Again, have a look at the LTT plots for all 10 birth-death trees.

Use the LTT plots you have just made to complete **Activity 3**.

4. Activities - Turn in this part

NAME: _____



Activity 1: Tree balance across the root node of a pure-birth tree.

For each of your pure-birth trees, count the number of lineages on the right- and left-hand side of the root node. For example, in the tree pictured at left, there are 15 lineages on the left and 5 lineages on the right side of the root node. We would denote this tree as (5,15); remember the convention from class to write the smaller number first. Use the results from these counts to fill in **Table 1** below.

Table 1. Split counts at the root node for my pure-birth trees.

Split	(1,19)	(2,18)	(3,17)	(4,16)	(5,15)	(6,14)	(7,13)	(8,12)	(9,11)	(10,10)
Count										

Once you have finished, you will compile these counts for all students in the class. There will be a class master data sheet being passed around; add your counts to that sheet. When the class data is compiled, write it into **Table 2** below.

Table 2. Split counts at the root node for whole class pure-birth trees.

Split	(1,19)	(2,18)	(3,17)	(4,16)	(5,15)	(6,14)	(7,13)	(8,12)	(9,11)	(10,10)
Count										

In class, I gave you the expected frequencies for these splits under the ERM model. Is this an ERM model? Why or why not?

The ERM expectations are:

$$\Pr[(k,n-k)] = \begin{cases} \frac{2}{n-1} & k \neq n-k \\ \frac{1}{n-1} & k = n-k \end{cases} \quad (\text{Eq. 1})$$

Next, compare your results to the ERM expectation using a chi-squared test. With a chi-squared test, one compares observed count data with the expectation under some model; if the data deviate too much from the expectation, one obtains a high test statistic and rejects the null hypothesis that the data were generated under the model being considered. To carry out this test, fill out the table below. Instructions for each row are as follows:

- i. Expected proportion: use Eq. 1, above
- ii. Expected count = expected proportion * n, where n is the total number of trees considered by the entire class
- iii. Observed count: From table 2, above
- iv. $(O-E)^2/E$: E from ii, O from iii.

Table 3. Calculating the chi-squared test statistic for the pure-birth model.

Split	(1,19)	(2,18)	(3,17)	(4,16)	(5,15)	(6,14)	(7,13)	(8,12)	(9,11)	(10,10)
i. Expected proportion										
ii. Expected count										
iii. Observed count										
iv. $(O-E)^2/E$										

To calculate your test statistic, sum all the values in **row iv**.

Under the null hypothesis, this test statistic should follow a chi-squared distribution with $10-1 = 9$ degrees of freedom (because there are 10 possible outcomes). Compare your test statistic with this distribution to calculate a p-value. What do you conclude?

Activity 2: Tree balance across the root node of a birth-death tree.

Repeat the above activity with your birth-death trees.

Table 4. Split counts at the root node for my birth-death trees.

Split	(1,19)	(2,18)	(3,17)	(4,16)	(5,15)	(6,14)	(7,13)	(8,12)	(9,11)	(10,10)
Count										

Table 5. Split counts at the root node for whole class birth-death trees.

Split	(1,19)	(2,18)	(3,17)	(4,16)	(5,15)	(6,14)	(7,13)	(8,12)	(9,11)	(10,10)
Count										

Table 6. Calculating the chi-squared test statistic for the birth-death model.

Split	(1,19)	(2,18)	(3,17)	(4,16)	(5,15)	(6,14)	(7,13)	(8,12)	(9,11)	(10,10)
i. Expected proportion										
ii. Expected count										
iii. Observed count										
iv. $(O-E)^2/E$										

Test statistic:

P-value and conclusion:

Activity 3: Lineage-through-time plots

- What do you notice about the LTT plots for pure-birth trees?
- What do you notice about LTT plots for birth-death trees? Are there any systematic differences between these LTT plots and the ones from the pure-birth model?
- Some scientists call this phenomenon the “pull of the recent.” Offer an explanation for the pattern that you observe.