

## **Lab 2a: Null Models of Character Distribution**

This lab is divided into two sections. In the first part we will learn how to detect phylogenetic signal in a discrete character using *Mesquite*. The basic principle will be to see if there are significantly less steps in the evolution of our character on our tree, than you would get for some random character that is not associated with the phylogeny. We will explore several different null models of character distribution, and try to get a feel for how tree shape effects the number of steps in the evolution of a character. If there is no signal, then the value of doing phylogenetic tests is questionable.

In the second part of the lab we will look at ways of detecting correlation between two discrete characters.

Open the **Dess200B.nx** file in *Mesquite*. Examine the first tree and the reconstruction of the character desiccation tolerance – make sure you understand why the parsimony reconstruction comes out as it does. Open the tree window. To keep things simple, resolve the polytomy in the seed plants (pick any resolution)

### **Null 2 - Calculate the distribution of steps under null trees.**

The first thing that we need to do is calculate the number of steps in this character on our tree. Select **Analysis>Trace character history** and then select **Parsimony ancestral states**. Record the number of steps in our character. Now we will compare our character to characters randomly distributed over random trees. Select **Analysis > New Bar & Line Chart For > Trees**. This will create a chart where the y-axis is number of trees.

Now you will be given several different options for how to generate trees, we want to simulate trees. We will deal with the different random tree null distributions more in a few weeks. The **Uniform speciation (Yule)** uses a random branching process, in which every branch has an equal probability of duplicating at any time. It only produces trees with the same number of taxa as you have in your character matrix. Pick the **Yule distribution** to start. 10.0 is cool for the tree depth.

Select **parsimony steps in character** (this will be your x-axis). Let's make 999 trees to start. Choose the only available character. You will see a histogram of how many steps in your character the simulated trees have. That looks nice and all, but you want real numbers. Click the **text** tab. Now you can see exactly how many simulations had each number of steps.

**Question 1.** What is your p-value for the number of steps on your actual tree? Don't forget to include your original tree in the calculation.

Save the outputs to file so you can compare the three different random tree generation algorithms. Select **File > Save window as text** and save the results to a new folder on your desktop. You can change your simulation method by going to **Chart > Tree Simulator**. Try **Equiprobable Trees** (all trees have the same probability) and **Uniform Speciation with sampling** (same as the yule process, except you create a tree with more taxa and then randomly sample from those), save your results as text and calculate a p-value. How does your simulation method affect your p-value? Which null distribution makes the most sense to use?

What is the minimum and maximum number of steps for this character? Create trees by hand showing the two extremes.

## **Null 1 – Calculate the distribution of steps for shuffled character**

Now we will compare our character to a character with its states randomly assigned to the tips of the same tree. Remember that a particular tree topology may inherently tend to generate more or less steps. Therefore a significant signal over random trees may reflect an effect of our phylogeny, rather than an effect of our character. Randomizing the data over the same tree will eliminate this potential complication.

First let's just create one character and look at it. In the **Character matrix** window create a new character using the **add character** tool. Now switch to the **select** tool, copy your original row and paste it onto your new character. Select your new character and then go to **matrix > alter/transform > shuffle states among taxa**. Trace your new character on the tree. You can do that with just the arrow. Make sure that you use your original tree and not one of the trees that you just created. How many steps does this character have?

So, we could do that a thousand times, or we could have *Mesquite* do it for us all at once. First delete the new character that you just made (**Edit > cut**), so that it doesn't confuse things. Go to **analysis > new bar and line chart for > characters**. What will this do? . Select **character value with current tree- parsimony character steps**. So, that's not really what we wanted, it just plotted the two characters from our matrix. To create a bunch of randomized characters select **Chart>Source of characters>Reshuffle character**. Choose your original character (not that it matters in this case). Let's do 999 characters again (well it was trees last time). Save this as text and calculate a p-value. How does this p-value compare to the random trees?

## **How does tree shape affect the distribution of steps for randomized traits?**

Manually create an extreme tree shape. The number of steps in the shuffled characters of your histogram will change automatically. Compare the parsimony steps for shuffled characters on this tree and calculate a p-value using the steps on our original tree. Do it again with a tree shape from the opposite extreme. How does this influence the distribution under the null? Do you have an intuition for the effect of tree shape?

**Question 2.** Describe one of your extreme shapes in words and provide the p-value.

## **Lab 2b: Correlated Changes in Discrete Characters**

Today we are going to look at several different statistical methods for determining if an apparent correlation between two discrete characters on a phylogeny is significant. All of these methods are available in *Mesquite*. These analyses can be done with other programs and there are other types of analyses that seek to answer the same question, but for now we'll stick with the programs that we will use on a regular basis.

The essential question at hand here is whether a character is more likely to change into a particular state in lineages that are in a particular state for another character, but they all take different approaches. Pagel's test looks at changes across the whole phylogeny, while Pairwise Comparisons breaks a phylogeny down into a number of comparisons between pairs of taxa. On the other hand, Pairwise Comparisons rely on looking for significance by counting changes in characters, while Pagel's test uses comparison of likelihood scores for the same purpose. Both these methods will be described in more detail below. A third method, Maddison's test, uses changes in number of characters, but makes the calculation over the whole phylogeny. This test is available in *MacClade*, but not *Mesquite*, so we won't be using it today.

We will do all these analyses on the **Concentrated Changes** example data set from the *Mesquite* examples. Open this file in *Mesquite*. Go to the **Show State Names Editor Window** and rename each of the two characters and states as follows: Character 1 = **dependent character, Fruit Type**, state 0 = **ancestral, fleshy**, state 1 = **derived, dry**; Character 2 = **independent character, Ecology**, state 0 = **tropical rainforest**, state 1 = **open savanna**. Save your work (but only this one time).

### **Comparing Reconstructions Visually**

The first thing that we will do is look at the distribution of the two characters against each other in one window, so that you might more clearly see their differences. Open the **Tree Window**. Go to **Tree>mirror tree window**. A new window will appear with the trees facing each other. Expand the window so that you can see what's going on.

First let's compare parsimony reconstructions of our two characters. Select **Mirror> left side >trace character history**, and select **Parsimony ancestral states**. Do the same for the right tree. You will see character 1 traced in both trees. Use the **Trace character** box for the right tree to switch to character 2. Now you can clearly see that Character 1 only switches from state 0 to state 1 in lineages that are in state 1 for character 2.

You can also modify the reconstructions for each of these trees. For example let's look at the likelihood reconstructions. You will see separate trace menus for each tree. For both trees go to **trace>Reconstruction Method>Likelihood Ancestral States**, and select **Current probability models**. It may be helpful if you set the **Tree Form** to **Balls and Sticks**.

You can set the Likelihood model to Mk2 by going to **trace>Probability Model>assym. 2 param**. Don't forget to do this for both trees. You can also use this method to compare different reconstructions of a single character. Why don't you compare the 2 parameter reconstruction of character 1 to the mk1 reconstruction?