Phylogenetics Lab
BIOL 01104 Fall 2019
Instructor: Dr. Spielman

This activity will make use of another program on the *rstudio.cloud* website to examine phylogenies, character matrices, and making phylogenies using parsimony. To set up this program, follow the steps to create a new project by clicking **Project->New Project from Git Repo**. Enter this URL when prompted for the "URL of your Git repository": https://github.com/spielmanlab/biol01104_phylo_lab .

Once this project has opened in RStudio, rename it to something memorable. Execute the **LAB_SETUP.R** file by clicking the file name and then the "Source" button. You will wait several minutes for the setup to complete. You may see a yellow bar pop up prompting you about installing "BiocManager." *This is not an error and you can safely ignore it!*

You know that setup is complete once a plain blue text arrow appears at the end of the red/blue status text that appears during installation. Once setup finished, click the file named **LAB_LAUNCH.R** and launch the program with the Green "Run App" button.

Note: All questions here are for thinking, not for submitting! Even so, it is very important that you understand the answers.

1. Under the "Character Matrix" tab, you will see a *sequence alignment* showing DNA sequences for several *orthologs*. Along the left side of this matrix you will see the names of twelve primate species. The numbers along the top represent the 898 positions along this gene. Each cell in the matrix has a color-coded "character state". **Identify the following:**
    ○ What does each row of the matrix represent?

    ○ What does each column of the matrix represent?

    ○ What does each cell in the matrix represent and what are the different letters?

    ○ How many different *character states* are in the tree? Hint: It's not 4! What do the unexpected state(s) represent?

2. You can see that at some positions or *columns* (for example, the first three), all the species have the same character state, in this case the same nucleotide. However, at other columns, character states differ among species. **Do you think the states the following positions will help to determine relationships among these primates? Why or why not? In addition, which of these positions do you think is the <u>most informative</u>?**

   ○ Position 4

   ○ Position 5

   ○ Position 7

   ○ Position 10

3. Using the "Phylogeny" tab, we will try to find the most parsimonious tree possible. **For now, only use the TOP SECTION of this tab which contains the button "Update Tree!".** Every time you click the "Update Tree!" button, a new guess for the best tree will appear on the left telling you the *parsimony score*, i.e. the *tree length* which represents the total number of evolutionary changes which occur along the branches of the tree. The tree shown on the right represents the best tree you have found so far. To search for the best tree, you can toggle the following options:

   ○ **How to update tree?** specifies how each new tree guess should be performed.
     i.   "New random tree" means guess an entirely new random tree
     ii. "Move on branch only" means, compared to the most recent guess, only a single branch will be moved elsewhere in the tree. This means the next tree guess will be very similar to the previous one
   ○ **Force tree improvement** means, when clicked, only trees that are *better* (i.e. have a lower parsimony score) than the previous guess will be considered. If the next tree guess is *worse* that the previous guess, it will not be considered in the tree search.
   ○ **Select outgroup** will change which species is used as the outgroup to draw the tree. Notably, the choice of outgroup does NOT affect the parsimony score (i.e changing the outgroup will not help to find the best tree). Changing the outgroup will simply change how the tree is drawn. Importantly, we know from fossil and molecular data that *Ring-tailed lemur* is the proper the outgroup for this data, but it is still useful to see how other outgroups look.
   Update at least 20 times to try and find the best tree you can find!

4. Once you have gotten the parsimony score as lower as you can, confer with your lab group and figure out which person has the **lowest parsimony score**. That person has found the **most parsimonious tree** among all group members.

5. Given that there are over **315 billion (!)** different ways to draw a tree for 12 species, it is pretty unlikely anyone found the best tree. On the bottom of the "Phylogeny" tab, you will see a button "Reveal the most parsimonious tree." Click this button, and set "Ring-tailed lemur" as the outgroup. **What is the score of the most parsimonious tree? How does this compare to your score?**

6. Compare your best tree to the most parsimonious tree. **Which branching patterns are similar? Different?**

7. Compare your most parsimonious tree to your lab group and/or the whole class. **Does everyone else's tree match up with yours, or are there any differences? What does this tell you about the "most parsimonious tree" for this data?**