

**Go to the Phylogenetics Activity website given on the course web page!**  
**You do NOT need to submit this worksheet, but it is VERY important that you understand the answers.**

**Part One: Understanding character matrices ("sequence alignments")**

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 the relationships among these primates? Why or why not? In addition, which of these positions do you think is the most informative?**
    - Position 4
    - Position 5
    - Position 7
    - Position 10

## Part Two: Reconstructing phylogenies using parsimony

1. 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:
  - **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* than 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 50 times to try and find the best tree you can find!***

2. Once you have gotten the parsimony score as low 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.
3. 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?**
4. Compare your best tree to the most parsimonious tree. **Which branching patterns are similar? Different?**
5. 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?**

**Part Three: Analyzing the evolutionary history of individual traits (sequence columns)**

Analyze the given column given below using the "Character History" tab, which traces character evolution over the most parsimonious tree. In a multiple sequence alignment, we consider the characters to be nucleotides (DNA sequence), and the specific character value to be either A, C, G, T, or - (missing or unknown).

1. **Column 7**

- a. How many and which species have the sequence "T"?
- b. Is the "T" a **derived** or **ancestral** character for this column?
- c. Did species gain T due to homology (shared ancestry), homoplasy (convergent evolution), or some combination? *Be very specific.*
- d. Do species with T form a monophyletic, paraphyletic, or polyphyletic group?

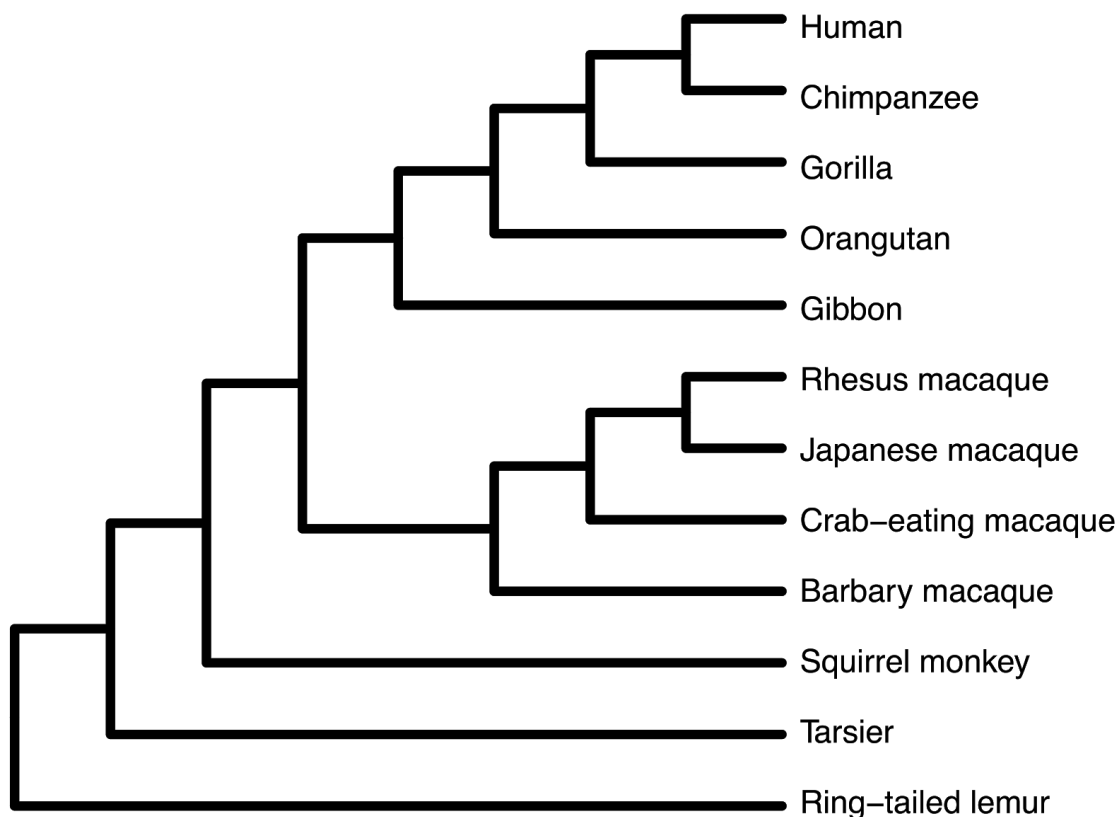
2. **Column 8**

- a. How many and which species have the sequence "T"?
- b. Is the "T" a **derived** or **ancestral** character for this column?
- c. Did species gain T due to homology (shared ancestry), homoplasy (convergent evolution), or some combination? *Be very specific.*
- d. Do species with T form a monophyletic, paraphyletic, or polyphyletic group?

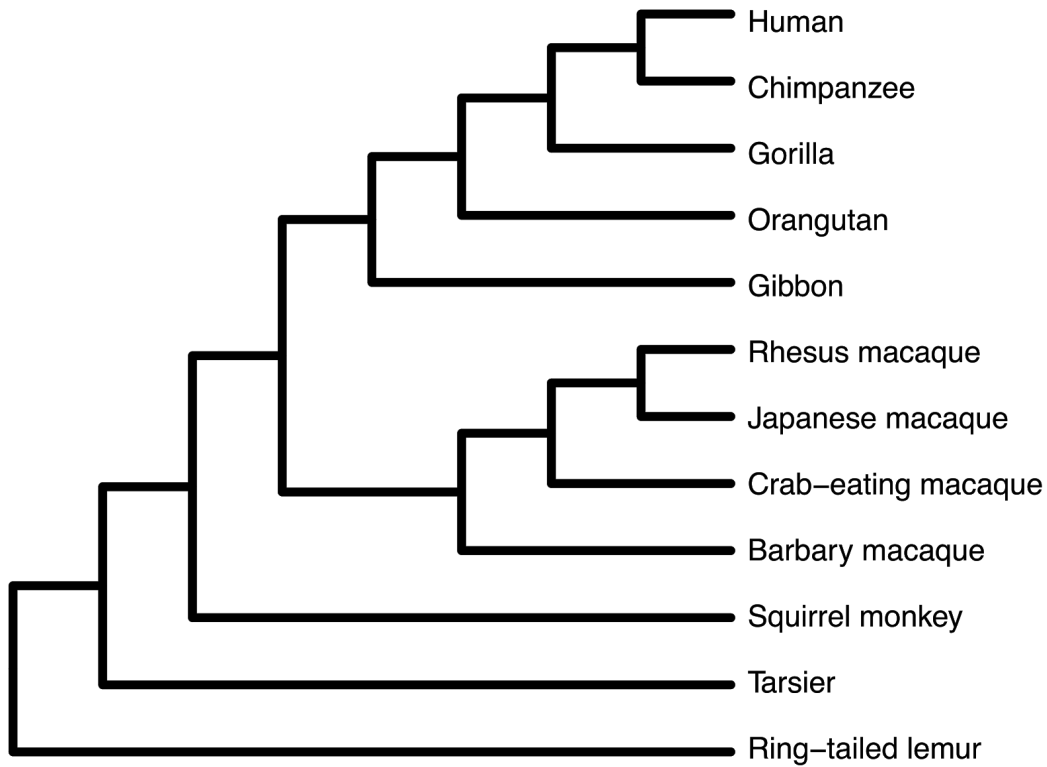
3. **Compare columns 20 and 21**

- a. How many different character values (nucleotides) does each column have and what are they?

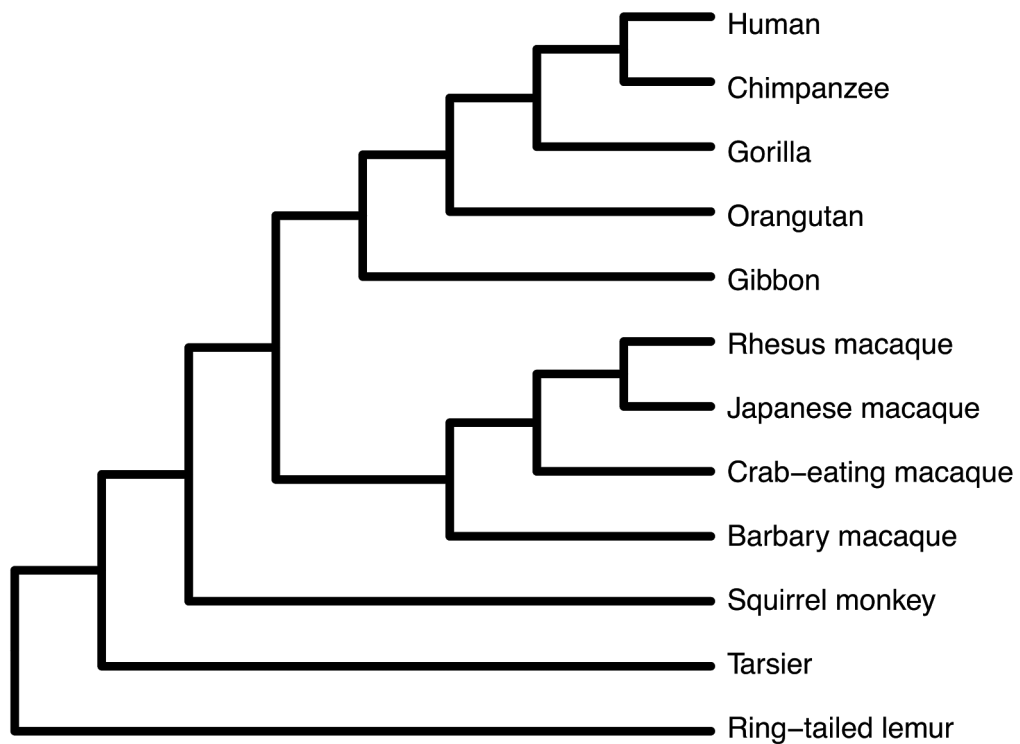
- b. Do columns 20 and 21 have the same amount of values or not?
- c. Which column appears to be evolving FASTER? Fast evolution is when there are more changes across the tree, and slow evolution is when there are fewer changes across the tree.
4. **Column 49. This question should be done by HAND.**  
This column has evolved to contain all 4 nucleotides in different species. It is therefore a very *variable* column in the character matrix. Determine the most parsimonious number of changes for the circumstance where the ancestor is each of the 4 nucleotides. In other words, first assume the ancestor is "A", then C, G, or T.  
How many changes would need to happen in the tree to produce the data we observe? On each tree, **map** changes when they occur and record your answer in the space for "score".  
**Ancestor is A. Score:**\_\_\_\_\_



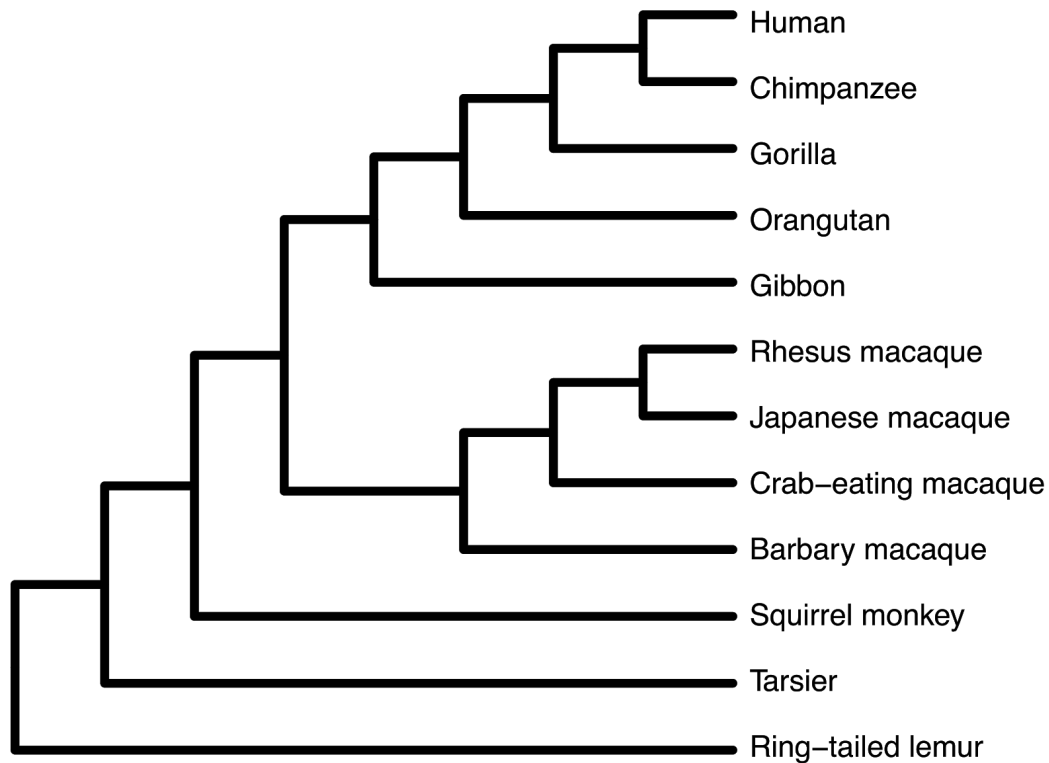
**Ancestor is C. Score:**\_\_\_\_\_



**Ancestor is G. Score:**\_\_\_\_\_



**Ancestor is T. Score:**\_\_\_\_\_



- b. Compare the parsimony scores for each ancestral case. Which character (A, C, G, T) is most likely the ancestral sequence for these species at column 49?  
[Hint: The answer may be more interesting than this!]

- c. Now, assume that **C** is the ancestor for column 49. What **WOULD** the phylogeny look like if all states were HOMOLOGOUS? Draw one possible tree that represents evolution with NO convergence using the data for column 49. Make sure the tree is fully bifurcating, i.e. no polytomies.