

Assignment #2

Part I. Drawing phylogenetic trees (using paper and pencil).

1. Draw the 15 unrooted binary topological trees that could describe the relationship between 5 taxa a, b, c, d, e. Group them naturally according to the relationship they display for the first 4 taxa a, b, c, d.
2. An unrooted tree has vertices v_1, v_2, \dots, v_9 with edges $\{v_1, v_2\}, \{v_1, v_6\}, \{v_1, v_9\}, \{v_2, v_7\}, \{v_2, v_8\}, \{v_3, v_9\}, \{v_4, v_9\}, \{v_5, v_9\}$.
 - a. Without drawing the tree, determine the degree of each vertex.
 - b. Use your answer to part (a) to determine the leaves of the tree.
 - c. Use your answer to part (a) to determine whether the tree is binary.
 - d. Draw the tree to check your work.
 - e. Write this tree in the Newick format.

Part 2. Git & GitHub

3. You need to know a little more about `git` than what we have covered in class. If you haven't used `git` before, check this [interactive tutorial](#) or read my [non-interactive one](#) or do both. Now create a `~/EEOB563/assignments` directory on Nova or your computer and initialize it as a Git repository. Add a README file to the directory and commit it to the repository (there are a few steps inbetween!). Create an empty GitHub repository and link your local repository to it. Push the changes you made. **Include the link to your GitHub repository in your assignment!**

Part 3. Unix and mafft

4. I shared with you several sequences sent to me by Jianzhi "George" Zhang, the author of the "The hearing gene Prestin unites echolocating bats and whales" paper. Use Unix commands to explore the file.
 - a. How many taxa are included in the dataset?
 - b. Are the sequenced aligned?
 - c. What is the length of the sequences?
5. Use `mafft` on the HPC-class (issue `module load mafft` first!) or [install it](#) on your computer. There are online servers that can run MAFFT, but do not use them or use them only after you tried all other options. One of our goals in this course is to become comfortable with the UNIX/command line interface. Use automatic option in mafft (`--auto`) to align the sequences and save the output in the clustal format.
 - a. What search strategy did mafft use for the alignment?
In addition to the alignment itself mafft program can output the guide tree it uses to build it (see [mafft website](#) for details).
You can view this tree with [FigTree](#) or [online](#).
 - b. What does this tree represent? Does it correspond to the species phylogeny you expected? Explain!

6. Alignment of aa sequences (extra point). Use Google to find out how translate DNA sequences into proteins. Translate DNA sequences we used in Q4 and 5 to amino acids and align them with mafft. Output a guide tree for the aa alignment. **a. What program/website did you use?**
b. Are the guide trees for DNA and aa alignments identical? What are the differences

Include your tree(s) (no alignment, please!) and your answers to the questions above in a single document and submit it on Canvas.

GOOD LUCK!