Assignment 2:
Trees

Important Notes:

- The maximum number of students in a group should not exceed 4 students.

- The assignment must be submitted as a compressed file containing both the Taskl.ipynb, the Task2.ipynb, and list_of_students.txt files, before midnight of the deadline date. The file must be named as assign2-student1ID-student2ID-student3ID-student4ID.zip, where student1ID is the ID of the first student, student2ID is the ID of the second student, student3ID is the ID of the third student, and student4ID is the ID of the fourth student.

- The list_of_students.txt file should list the names and IDs of all students in the group.

Required Tasks

Task1:   Find a published article from a scientific journal that contains a phylogeny. Refer to the phylogeny to answer the questions below. If you do not know where to begin to look, consider the journals *Systematic Biology*, *Molecular Biology and Evolution*, *Evolution*, *Molecular Phylogenetics and Evolution*, and *Genetics*.

1) Is the phylogeny rooted or not?

2) Does your phylogeny contain one or more outgroups? If so, identify them.

3) How many taxa are represented in your phylogeny? What biological taxonomic level is represented by a leaf on your phylogeny?

4) How many leaf nodes and internal nodes are in your phylogeny? How many edges?

5) Do the edges in the phylogeny have lengths? If so, what do these lengths represent?

6) If your phylogeny has edge lengths, is the phylogeny ultrametric or not?

7) Are the edges or nodes in your phylogeny labeled wih numbers? If so, what do these numbers represent?

8) Does your phylogeny have polytomies or is it a binary tree? If it has polytomies, are they hard or soft?

9) Find a clade with between 5 and 15 taxa in your phylogeny (or the entire phylogeny itself). What is the total number of possible rooted trees with this number of taxa? What about unrooted?

Task2:   Take at least 20 protein sequences which are at least 30% identical or better and:

1) Make a multiple sequence alignment. The sequences must be approximately the same length. Then:

2) Make two phylogenies, one using the UPGMA method and the other using the Neighbor Joining method.

3) Describe the resulting alignments and include graphic images of the alignments and phylogenies.

4) Mention if the trees seem reasonable biologically or taxonomically reasonable by comparison with standard taxonomies (NCBI Taxonomy Page).

5) Do the two trees have the same topology? See important notes, below!

6) Do the trees have the same branch lengths?

7) If the two trees do not have the same topology or branch lengths, describe the differences and indicate why you think the two trees differ.

8) Are the differences taxonomically or biologically significant?

9) Do the trees show evidence of paralogous evolution?

10)   Which nodes are orthologous and which are paralogous bifurcations?

11)   Do the trees show evidence of either gene conversion or horizontal gene transfer?

12)   Include your sequences in FASTA format (unaligned or aligned).