

Homework Assignment 4

Question 1: In class we worked with a whole number of different tree building approaches. For this week's homework I'd like you to explore their difference on our geographically diverse set of SARS-CoV-2 genomes that I have aligned for you in `data/sc2_subset.align.fasta`. In particular I'd like you to do the following

1. Create a UPGMA tree from that data using biopython
2. Create a likelihood tree from that data using RAxML under the JC69 and HKY85 mutational models.
3. Compare and contrast results between these trees. Do you notice any systematic differences? Please write a few sentence on what differences you do or do not see.

visual display tip— I've created a *metadata* file at `data/sc2_continents.csv` that has an assignment of each genome to broad geographic region. You could use that information to color branches in the tree according to geography. For instance check out this snippet of code