COMP 5970/6970-004
Computational Biology: Genomics and Transcriptomics
Project 1

Haynes Heaton

Spring, 2022

From the multiple alignment found in the file msa.pir in Files on canvas, write a program to find the distance matrix and phylogenetic tree for these sequences. This file is in the PIR format which consists of a line with >sequence_name followed by a line with the sequence with dashes for indels. This continues in this fashion for each sequence in the multiple alignment. Output the distance matrix and a dot format https://en.wikipedia.org/wiki/DOT_(graph_description_language) representation of the graph and use graphviz (https://graphviz.org/) (or some other graph visualizer) to visualize this graph. Note: in dot you can give edges weights with something like `A - B [weight = 2]`.

Deliverables

- The code used
- Distance matrix of original sequences
- Distance matrix with internal nodes added
- Graph representation in dot format
- Graph visualization