

ASSIGNMENT 2

1. Construct a phylogenetic relationship for the given nucleotide sequences(Nucleotide.txt)

a. Write a script to generate a distance matrix csv file for the sequences present in the data file. Name the distance matrix file as '**Ndistance.txt**'

For example

seq1 = 'ATGCATGCAA'

seq2 = 'ATGCATGCTA'

Distance(seq1,seq2) = Mismatches/total length = 1/10 = 0.1

b. Write a script that uses 'Ndistance.txt' and generate phylogeny relationship between the organisms using UPGMA method.

2. Construct a phylogenetic relationship for the given protein sequences(Protein.txt)

a. Write a script to generate a distance matrix csv file for the sequences present in the data file. Name the distance matrix file as '**Pdistance.txt**'. Use 'BLOSUM62.txt' for getting score values.

For example

seq1 = 'AGYFKTP'

seq2 = 'GRKLYSK'

Score(AG) = 0, Score(GR) = -2 and so on

Distance(seq1,seq2) = Score(AG) + Score(GR) +....

b. Write a script that uses 'Pdistance.txt' and generate phylogeny relationship between the organisms using UPGMA method.