

# Introduction to Biology

## Phylogenetic Analysis Assignment

- 1) Construct a phylogenetic relationship for the given nucleotide sequences (Nucleotide.txt).
  - a) Write a script (*q1a*) 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 (*q1b*) that uses '*Ndistance.txt*' and generate phylogenetic relationship between the organisms using UPGMA method.
- 2) Construct a phylogenetic relationship for the given protein sequences (Protein.txt).
  - a) Write a script (*q2a*) 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 for getting score values.
  - b) Write a script (*q2b*) that uses '*Pdistance.txt*' and generate phylogenetic relationship between the organisms using UPGMA method.

<http://www.slimsuite.unsw.edu.au/teaching/upgma/>

<https://www.ebi.ac.uk/Tools/psa/>