BT302: Bioinformatics Laboratory Assignment No. 4 31.08.2022

Note 1: Submit the assignment online through Moodle either in .doc or .pdf format. Your final report file should be named as "**YourName_BT302_Lab4_31082022**". Make sure that your name and roll numbers are written at the first page of your final report. Note that you can upload only one file; thus, put together all the answers in a single file.

Note 2: There are two parts of this assignment. In this first part, students are expected to answer the questions 1-6. In the second part, students are expected to write a script to achieve the intended results. Each student should choose only one of the parts.

----------PART 1----------

The purpose of this exercise is to perform multiple sequence alignment (MSA).

1. Download the protein sequence of Cas9 of *Streptococcus thermophilus* (strain ATCC BAA-491 / LMD-9) from UniProtKB database.
   (a) Use the program PSI-BLAST from the NCBI home page keeping all the search parameters as default except the database as UniProtKB. Perform the homology search using the downloaded protein sequence as query and run iterations until you get no new hit. How many iterations did you need to perform?

2. Pick one sequence homolog each having sequence identity of 100%, 90%, …, 10%, if available. In case of multiple sequence homologs having same sequence identity, you can choose any one of them randomly. How many homologous sequences did you choose? Download the protein sequences (in .fasta format) of these hits and note down the details (e.g. protein name, organism, sequence length) of these entries. *Note: you can copy and paste all the protein sequences with a proper naming in a single word file.*

3. Use the program ClustalW to perform multiple sequence alignment (MSA) of all the sequences which you have downloaded.
   (a) Note down the default set of parameters which are being used to carry out MSA.
   (b) Copy and paste the MSA in your report and download the alignment file in your computer for later use.
   (c) Note down all the amino acids which are conserved across all these sequences.
   (d) Generate the phylogeny tree; copy and paste the image in your report file.
   (e) Now, change the protein weight matrix for both pairwise and MSA options keeping all the other parameters as default. Do you find any difference in the alignment and/or tree?

4. Now, use the program Clustal Omega to perform multiple sequence alignment for the same set of protein sequences.
   (a) Note down the differences in the conservation of amino acids to that of the MSA obtained using the program ClustalW.
   (b) Do you see any difference in the phylogeny tree?

5. Now, use the program T-COFFEE to align these sequences. Note down if there is any difference either in the alignment or in tree.

6. Use the program ESPript to decorate for any of the multiple sequence alignments.

----------END of PART 1----------

----------PART 2----------

1. Write a code to align at least three user-input protein sequences using progressive method of multiple sequence alignment.

Note: Copy paste or attach your written code in the report and create a README file on how to use the code.

----------END of PART 2----------