

Bio231 - Computational Biology

Assignment no 2

Assignment Due: March 21st, 2022

Instructions

- Read the section on plagiarism in your student handbook. Plagiarized assignments will be referred to the DC.
- Assignments must be submitted in the “Assignment” section on LMS as a master zip file containing individual files for MATLAB codes and word file for part 1. Code must be properly commented
- Use **MATLAB** to write code
- Use of built-in libraries and Functions are NOT allowed
- Assignments submitted on Dropbox, emailed assignments or late submissions will not be accepted.
- This is an individual assignment. Discussion is not allowed.

Part 1:

(10 + 10 + 10)

Question 1: Align the protein sequences THISLINE and ISALIGNED using dynamic programming (Repeated Matches). Use BLOSUM62 matrix to calculate the match/mismatch scores. The gap penalty is -4

Question 2: Compute the distance matrix and use UPGMA to identify the phylogenetic relationship between different sequences.

Sequence1: **AACTCGC**

Sequence2: **TACACGC**

Sequence3: **TATTCCG**

Sequence 4: **CGACCTA**

Sequence5: **GCAGCTA**

Sequence6: **TTATTCA**

Question 3: Perform a phylogenetic analysis using **ClustalW** or **MegaX** of **ATK1** gene and write a short note explaining how **ATK1** in humans is different from other species using the tree.

Part 2:

(10 + 10)

Question 4: Write a program to take input from user and generate a random DNA sequence of that length. Also find the GC content in the sequence.

Question 5: Write a program to Load 5 FASTA sequences and create a distance matrix.