BE7-R3: APPLIED BIOINFORMATICS

NOTE:

1. **Answer question 1 and any FOUR questions from 2 to 7.**
2. **Parts of the same question should be answered together and in the same sequence.**

Time: 3 Hours **Total Marks: 100**

1.

a) How is Gene organisation is different in Eukaryotes compared to prokaryokes?
b) Why BLAST algorithm is heuristic?
c) How GenBank format is different from FastA format?
d) What is the edit distance of the words CCT and ACGCTT? Align the sequences ATT and TTC.
e) Why higher order markov models are not generally used in Bioinformatics applications?
f) How can you find all open reading frames in a given sequence?
g) Imagine you want to calculate the shortest common super-sequence i.e. the shortest sequence to which all sequences can be aligned without mismatches. Which cost/weight scheme can you use? Give one example.

(7x4)

2.

a) Write the Needleman-Wunch algorithm for Global Alignment. Also mention the modifications necessary to solve local alignment problem.
b) What are three steps generally carried out for a typical fragment assembly algorithm? Explain?

(10+8)

3.

a) What is biological motivation of sequence analysis?
b) Assume you are given a set of DNA sequence belonging to different species. What strategy has to be adopted to find a common pattern in these sequences?
c) What are the basic assumptions made while developing PAM and BLOSUM scoring matrices?

(6+8+4)

4.

a) What do you mean by statistical significance? Explain the same in the context of BLAST searches.
b) How PSI-BLAST and PHI-BLAST are different from BLAST?
c) Why low complexity sequences are masked in BLAST searches. Mention name of the two programs used for this purpose.

(8+6+4)

5.

a) What are CG-islands? How will you proceed to find a predictive method for the same using First Order Markov Chain?
b) Explain the term “Staite transition probabilities” and “emmission probabilities” in the context of hidden Markov models.
c) Discuss the possible applications of Markov models in biological problems.

(8+4+6)

6.

a) What is homology-based approach for gene prediction?

b) What is DNA sequencing? Discuss DNA sequencing methodologies in brief.

(8+10)

7.

a) Let S_1 = AATTCGCGTA and
S_2 = TATCGCTACA
Obtain the optimal global alignment using dynamic programming method. Use any Scoring scheme of your choice.

b) Describe the Viterbi algorithm used in hidden Markov model.

(10+8)