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) Define the term "codon bias" in the genomic context.

b) Give an example for at least two types of repeats abundant in genomic sequences.

c) Does pairwise alignments signify any biological functions? If so what are they?

d) What inferences can be drawn from multiple sequence alignments?

e) How does the parameter "word size" dictate the sensitivity of BLAST searches?

f) Describe briefly two popular "gene finding software tools" based on markov models.

g) Derive the consensus sequence from the given alignment.

t g c a c t a t g g
t t c a t g a c g a
t a c a a g a t a g
t g c t t a a g a g
t g c t c c t g a t
a t c t c a a t t a
a a c a t c t g g a
t g c t t a a c c a
a g c a a g a a a c

(7x4)

2.

a) Describe in detail any one method of sequencing. Explain the steps generally adopted to annote the same.

b) Discuss briefly the salient features of GenBank.

c) Why are emputational techniques preferred to analyze sequences and structures?

(6+6+6)

3.

a) What to you understand by the term "distance" in string comparison problems? Illustrate with an example.

b) Explain the notion behind the Smith-Waterman algorithm for the local sequence alignment.

c) Given two strings i.e GAATTC and GATTA. Derive a Global similarity alignment for these strings using the scoring scheme S(a,a) = +2 ; s(a,b) = -1 if a ≠ b ; and gap penalty = 2.

(4+6+8)

4.

a) Write a short note on "scoring matrices" employed in BLAST searches.

b) How does the parameter "expect value" or "e value" play a role in controlling the BLAST search output?

c) List the names of different "Blast versions". How is PSI-BLAST distinct from PHI-BLAST?

(5+5+8)

5.

a) Define “markov chain” and its order.

b) Explain with a simple example, how the state probabilities and transition probabilities are computed in a markov model.

c) Assume that you are provided with the state transition probabilities for both positive model as well as negative model. How would you proceed to discriminate the same for test sequences?

(4+8+6)

6.

a) In a typical HMM model, it is possible to have many different state paths which can give rise to the same sequence. Explain how these probabilities are computed for all possible paths.

b) In the context of hidden markov model, when does one use the Baum–Welch algorithm, and the Viterbi algorithm, and why? Give biologically motivated examples.

(10+8)

7.

a) Describe any algorithmic method employed to obtain a multiple sequence alignment. Briefly mention about the computational complexity associated with it.

b) Explain the term “motifs” with an example.

c) What do you understand by secondary structure of proteins?

(10+4+4)