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 terms "Transcription" and "Translation" in the context of protein synthesis.

b) Consider of DNA string 5' ATCCACGGCCATAGG 3'. Write all the 5' substrings ending with G.

c) Write down the consensus sequence for the following multiple alignment:

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

d) Why is dynamic programming technique not suitable for large-scale database homology search?

e) What types of errors do occur in DNA fragment assembly problem?

f) Does degree of dependence relate to order of a Markov Chain? Briefly explain.

g) What do you understand by the term "hidden" in a "hidden Markov model"?

(7x4)

2.

a) List the advantages of using Bioinformatics tools and techniques in analysing raw sequence data.

b) The restriction enzymes EcoRI precisely recognise the site GAATTC and cut the E.coli DNA into two fragments. Cuts are between the first G and the second A. Consider the DNA sequence ATCCGAATTCATTG
Write down the resulting fragment cut by the above-mentioned enzyme.

c) Explain a typical GenBank record and describe its main features.

(8+6+4)

3.

a) Give the basic dynamic programming algorithm for a pair-wise sequence alignment problem.

b) Compare the contrast the gene structure of a prokaryote and eukaryote.

c) How is Genome studied at different levels? Briefly discuss about the Human Genome project.

(6+6+6)

4.

a) Consider the problem of aligning two tRNA sequence of E.coli. These sequences are GGTGATTAGCT and GCTGATATAGCT. Derive a Global similarity alignment for the above two sequences using the scoring scheme S(a,a)=+1;s(a,b)=-1 if $a \neq b$; and gap=2.

b) Using dynamic programming method, discuss the algorithmic complexity of pairwise alignment problem. Why is such technique not preferred for multiple alignment sequence problem?

(12+6)

5.

a) Define a linear and affine gap penalty.
b) Explain the relevance and utility of multiple aligned sequences.
c) Which method is employed in popular "ClustalW" multiple alignment programs?

(4+4+10)

6.

a) Write a short note on PAM and Blosum scoring matrices.
b) How are homologous DNA sequences from the database retrieved using a typical BLASTN program for a given query nucleotide sequence?
c) How is PSI-BLAST different from BLASTN?

(4+10+4)

7.

a) Define the term "memoryless property" and "time homogeniety" property in connection with the Markov Chains.
b) Consider the following transition probability matrix. Identify the probable state, which is likely to be visited more often from the initial state "A"? Find out the error in the last row of this transition probability matrix.

\	A	C	G	T
A	0.2	0.1	0.6	0.1
C	0.1	0.7	0.1	0.1
G	0.2	0.2	0.5	0.1
T	0.1	0.3	0.2	0.5

c) For a biological important functional site the transition probability matrices for "+" model and "-" model are given below. Using these models suggest prediction methodology.

+	A	C	G	T
A	0.2	0.2	0.4	0.2
C	0.1	0.3	0.2	0.4
G	0.2	0.3	0.3	0.1
T	0.1	0.4	0.3	0.2

-	A	C	G	T
A	0.3	0.1	0.1	0.5
C	0.3	0.3	0.1	0.3
G	0.2	0.2	0.3	0.3
T	0.1	0.2	0.2	0.2

(4+4+10)