Registration No. | | | | | | | | | |

M.E./M.Tech. Degree Examinations, January 2017

First Semester

BIOTECHNOLOGY

BY16102 – COMPUTATIONAL BIOLOGY

(Regulation 2016)

QP Code: 766996

Time: Three hours **Maximum : 100 marks**

Answer **ALL** questions

PART A - (10 X 2 = 20 Marks)

1. Determine the edit distance between TGCATAT and ATCCGAT.
2. Find the local alignment between S and T:

   S: ggtctgag

   T: aaacga

   Scoring function: match 2, mismatch/gap: -1
3. Calculate the score for the provided alignment using the given scoring function.

   CTGTCGCTGCACG

   –TGCTG–TG——

   Scoring function:

   Reward for match: 10

   Penalty for mismatch: 2

   Gap penalty:5
4. Distinguish between rooted and unrooted trees.
5. What are the three main methods of protein tertiary structure prediction?
6. What are the uses of Force field? Give an example.
7. Motivate the need for normalisation in the microarray data analysis.
8. Identify the top differentially expressed gene and indicate its fold change:

Gene	state1	state2
A	100	50
B	5	10
C	200	100
D	30	10
E	5	20

9. Write a one-line PERL statement to print if the sequence stored in $myseq contains a start codon.

10. Write PERL code to reverse-complement a DNA sequence stored in $myseq.

PART B - (5 X16 = 80 Marks)

11. (a) (i) What is the recurrence relation of dynamic programming? **(4)**

(ii) Construct the dynamic programming Matrix and hence find the global alignment between the following nucleotide sequences: **(12)**

Seq. 1: CTCGCAGC

Seq. 2: CATTCAC

Scoring function: match 10, mismatch -2, gap penalty -5.

(OR)

(b) (i) End-gap free alignment is a special case of global alignment which permits free gaps at the start and the end. **(12)**

S: CACTGTAC

T: GACACTTG

Scoring function:- match:+2, mismatch/gap: -1

(ii) Is it identical to the global alignment? (use the same scoring function) **(4)**

12. (a) (i) Evaluate the methods for protein function annotation using profiles and patterns. **(4)**

(ii) How will you convert sequence data to distance data? **(4)**

(iii) Deduce the neighbor-joining tree given the following distance matrix. **(8)**

A	0				
B	2	0			
C	7	7	0		
D	14	10	9	0	
E	15	11	10	3	0
	A	B	C	D	E

(OR)

(b) (i) Enumerate four features of ClustalW that facilitate the alignment of distant but homologous sequences. **(4)**

(ii) Compare and contrast parsimonious trees and ultrametric trees. **(4)**

(iii) Estimate the reliability (quality) of a tree. **(8)**

13. (a) (i) Propose the steps in homology modeling when the sequence identity between the target and template is <30%. **(4)**

(ii) How would you assess whether a lead compound is an effective inhibitor of a certain receptor? Discuss the methodology with the appropriate tools. **(12)**

(OR)

(b) (i) What is threading? **(4)**

(ii) Given an mammalian enzyme of unknown structure, how would you carry out biophysical studies of the enzyme in silico? **(12)**

14. (a) (i) Draw a schematic representation of an artificial neural network with a hidden layer. **(6)**

(ii) Design an ANN to predict protein secondary structure. **(10)**

(OR)

(b) (i) What is a profile HMM? **(4)**

(ii) Determine the sequence signals that could be used in gene prediction strategies. **(4)**

(iii) Design an HMM representation for gene prediction. **(8)**

15. (a) Write a Perl program to scan for a given motif in multiple Fasta files. **(16)**

(OR)

(b) (i) Given a list of oligos, write Perl code to print them sorted by length. **(8)**

(ii) Write a Perl program to accept a protein sequence and count the charged residues in it. **(8)**