Register Number:……….……

M.TECH DEGREE EXAMINATIONS: JUNE 2018

(Regulation 2015)

Second Semester

BIOTECHNOLOGY

P15BTT203**:** Computational Biology

COURSE OUTCOMES

CO1: Understand the concepts of biological data and databases

CO2: Understand sequence alignment methods

CO3: Describe creation and biological motivation for preparing phylogenetic trees

CO4: Understand machine learning techniques as applied to biological data

CO5: Be able to use UNIX and program in Perl.

Time: Three Hours   **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1.   Assertion (A): Bootstrapping is a standard technique for inferring confidence values on   CO3   [K_2]

phylogenetic trees.

Reason (R): Bootstrapping do so by reconstructing many trees without altering the

input data and generate replicates

a)   Both A and R are Individually true and   b)   Both A and R are Individually true but

R is the correct explanation of A   R is not the correct explanation of A

c)   A is true but R is false   d)   A is false but R is true

2.   Identify the file format given below   CO2   [K_1]

P1:JMED

Protein X – Homo sapiens

MKALTARQQEVDSLIRD

a)   GDE   b)   FASTA

c)   NBRF   d)   GCG

3.   Which one of the following is a bibliography database ?   CO1   [K_1]

a)   GenBank   b)   PDB

c)   Scopus   d)   USPTO

4. Match the color with the amino acid group in a MSA output　　　CO3　[K$_2$]

List I	List II
A. Yellow	i. Negatively charged
B. Green	ii. Polar
C. Magenta	iii. Small non-polar
D. Red	iv. Hydrophobic

	A	B	C	D
a)	ii	i	iii	iv
b)	iii	iv	ii	i
c)	ii	iv	iii	i
d)	iii	i	ii	iv

5. If a BLAST search returns a match with an E-value of 1 e^-20, and 200 searches were done simultaneously, how many false negative are there?　　　CO4　[K$_3$]

   a) 1 e^-20　　　　　　　　　　　b) 10^-20

   c) 10^-20/200　　　　　　　　　　d) There is no way to calculate false negatives.

6. The average time to calculate the optimal alignment for two sequence m and n using a dynamic programming approach is　　　CO2　[K$_2$]

   a) *O(nm)*　　　　　　　　　　b) *O(n/m)*

   c) *O(n+m)*　　　　　　　　　　d) *O(n-m)*

7. Select the correct statement with respect to scoring matrices　　　CO2　[K$_2$]

   1. Likelihood of one residue being substituted by another in an alignment
   2. Obtained derived from statistical analysis of residue substitution data from sets of reliable alignments of highly unrelated sequences.
   3. Assignment of scores is based on the assumption that the frequencies of mutation are equal for all bases in a nucleotide matrix
   4. The conversion can be either to the inverse log base of 10

   a) 1,3　　　　　　　　　　　　b) 1,4

   c) 1,2　　　　　　　　　　　　d) 2,3

8. Which of the following is not a mechanism of genetic variation?　　　CO3　[K$_2$]

   a) Mutation within the coding sequence of genes　　　b) Mutation within the regulatory DNA of a gene

   c) Purifying selection　　　　　　d) Exon shuffling

9. PERL has the following features a.   CO5   [K_L]

     a. PERL works with HTML, XML

     b. PERL is not an object oriented language

     c. PERL requires a special interpreter that can't be integrated with others

     d. PERL is a Y2K compliant.

a) A and B are correct   b) A and C are correct

c) B and C are correct   d) A and D are correct

10. Arrange the following procedure in order to perform structural-based alignment to obtain sequence and structure based elements   CO4   [K_2]

     1. PSI-PRED and PSI-blast to obtain sequence and structural-based constraints

     2. MSA of representative sequence and MSA of input sequence.

     3. PSI-blast against sequence database

Clustering sequences of all representative sequences

a) 3-4-2-1   b) 4-3-2-1

c) 2-1-4-3   d) 1-2-3-4

PART B (10 x 2 = 20 Marks)

11.	List the advantages of star topology over bus topology.	CO1	[K_2]
12.	What does overclocking means ?	CO1	[K_3]
13.	List the common interpretations from a dot matrix plot.	CO2	[K_2]
14.	What does positive and negative indicate in a PAM scoring matrix?	CO2	[K_2]
15.	Illustrate the architecture of a CATH database.	CO4	[K_2]
16.	List any two tools in identifying a complete gene structure.	CO4	[K_2]
17.	Differentiate PHI and PSI-BLAST.	CO3	[K_2]
18.	How do you validate a modelled protein?	CO4	[K_2]
19.	How are hashes defined in PERL scripts?	CO5	[K_3]
20.	Enumerate the applications of system biology.	CO5	[K_2]

PART C (6 x 5 = 30 Marks)

21.	What is the most reliable network topology? Add a note on its importance.	CO1	[K_2]
22.	A lab working on a microbial consortium requires a deeper and divergent tree. What method would you suggest for construction of phylogeny? Justify your choice.	CO4	[K_5]
23.	How to calculate position-specific scoring matrix (PSSM)?	CO2	[K_3]
24.	Enumerate the steps involved in reconstructing a phylogenetic tree using neighbor joining method.	CO3	[K_2]

25. Explain the methods (any two) to evaluate a Multiple Sequence Alignment (MSA). List four applications of MSA.   CO4   [K$_3$]

26. Script a Unix code for the following   CO5   [K$_2$]

       Compare two protein files and display the difference.

Answer any FOUR Questions

PART D (4 x 10 = 40 Marks)

27. Given a set of multiple DNA sequences of similar length; explain how progressive methods of sequence alignment is employed to obtain an optimal alignment. What other approaches to obtain an optimal alignment?   CO3   [K$_2$]

28. Elaborate on the various steps involved in obtaining an optimal alignment of two sequences using XPARAL.   CO2   [K$_3$]

29. List the different types of gaps in a sequence alignment. Add note on the application of gaps in sequence alignment.   CO2   [K$_2$]

30. Consider that a scientist working on a media optimization problem. Four key parameters have been identified *viz.*, pH, temperature, concentration and water content. Considering the above parameters, construct a neural network model to obtain an optimal solution. Give a note on the mathematical function utilized in the hidden layer.   CO4   [K$_5$]

31. i) Script a PERL code to calculate the reverse complement of a strand of DNA   (5)   CO5   [K$_3$]

    ii) Illustrate the structure of UNIX file system. Add note on modifying file permissions.   (5)   CO5   [K$_2$]

****