MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent unit of MAHE, Manipal)

SEMESTER B.TECH. END SEMESTER EXAMINATIONS

NOVEMBER 2019

SUBJECT: COMPUTATIONAL BIOLOGY [BIO 4018]

Date of Exam: 23/11/2019

Time of Exam: 2.00 to 5.00 PM

Max. Marks: 50

Instructions to Candidates:

♦ Answer ALL the questions & missing data may be suitable assumed

1A.	You were asked to find a motif in the nucleotide sequences from NCBI. Which algorithm design technique would you suggest? Why?	2
1B.	What is the difference between object oriented and relational database? Which type do you think NCBI belong to? Justify.	2
1C.	Do you think supervised machine learning techniques are complex compared to the unsupervised ones? Why? Justify with respect to classification and clustering.	3
1D.	Both Hidden Markov Models and Neural Networks contain a hidden layer. Is this statement correct? Explain in detail.	3
2A.	What are frontend and backend of a database? Which one do you construct using SQL? Why SQL is preferred for this purpose?	2
2B.	Write an algorithm and explain it to find a chain of exons in a nucleotide sequence.	4
2C.	Exact method combines all the optimal pairwise alignments into a multiple alignment. Do you get a correct multiple alignment always? Justify.	2
2D.	Do you suggest Dynamic Programming for multiple sequence alignment? Why? Explain.	2
3A.	You were asked to cluster a set of sequences using NJ or UPGMA methods. Which one do you think is better with respect to distance calculation?	2
3B.	What is an edit distance and a tree distance? How do you fit them?	2
3C.	Construct a phylogenetic tree from the given distance matrix by identifying the additive matrices and reconstructing the trees from them.	4

		A	4	B	C	1	D										
	A	(C	6	12	2	1 1										
	B	B 6 0)	9										
	С	C 12		10	0		1 1										
	D	1	1	9	11	1	0										
	You were given two outputs of multiple sequence alignments. Which one do you su for phylogenetic analysis using maximum parsimony method? Why?									one do you sugge	st						
		1	2	3	4	5				1	2	3	4	5			
3D.	S 1	A	А	Т	G	С			S 1	A	A	Т	G	С			
	S2	A	Т	Т	G	С			S 2	А	Т	Т	G	А			2
	S 3	A	G	С	G	G			S 3	А	А	С	G	С			
	S4	G	С	Α	G	А			S 4	G	Т	С	G	А			
	(a) (b)																
	Dica	1100	tha	hior	rahi	001	alu	toring algorith	nm I	Iow	ia k	ioro	rahi	<u>aa1</u>	alustaring differen	+	
4A.	Discuss the hierarchical clustering algorithm. How is hierarchical clustering different from K-Means clustering?													ll	5		
4B.	How do you decide whether a cluster is good or bad based on the clustering principles? What are the two ways of computing distance between two clusters in hierarchical clustering?													? 1	3		
4C.	"Trimming the ends of sequences, obtained from multiple sequence alignment, will only affect the character-based method but not the distance-based methods for phylogenetic tree construction". Is this statement true? Justify.													y c	2		
5A.	You were asked to induce multiple mutations in the given sequence so that all the resulting sequences exhibit the same secondary structure as the wild type. Suggest any two such sequences with their secondary structures. Explain the concept of evolution that you consider in this case?												e y n	3			
	G A	GC	CU	JU	CU	С											
5B.	You were asked to predict the family and secondary structure of a protein sequence. Which computational technique do you suggest for each case? Discuss their working principles and methodology.												g	7			