Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
A Constituent Institution of Manipal University

PROGRAM ELECTIVE

VII SEMESTER B.TECH. END SEMESTER EXAMINATIONS, NOV 2017

SUBJECT: COMPUTATIONAL BIOLOGY [BIO 4018]

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data may be suitably 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.	Construct a phylogenetic tree from the given distance matrix by identifying the additive matrices and reconstructing the trees from them.	**4**

	<table><tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td></tr><tr><td>A</td><td>0</td><td>6</td><td>12</td><td>11</td></tr><tr><td>B</td><td>6</td><td>0</td><td>10</td><td>9</td></tr><tr><td>C</td><td>12</td><td>10</td><td>0</td><td>11</td></tr><tr><td>D</td><td>11</td><td>9</td><td>11</td><td>0</td></tr></table>	
3B.	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**
3C.	What is an edit distance and a tree distance? How do you fit them?	**2**
3D.	You were given two outputs of multiple sequence alignments. Which one do you suggest for phylogenetic analysis using maximum parsimony method? Why? <table><tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>S1</td><td>A</td><td>A</td><td>T</td><td>G</td><td>C</td></tr><tr><td>S2</td><td>A</td><td>T</td><td>T</td><td>G</td><td>C</td></tr><tr><td>S3</td><td>A</td><td>G</td><td>C</td><td>G</td><td>G</td></tr><tr><td>S4</td><td>G</td><td>C</td><td>A</td><td>G</td><td>A</td></tr></table> (a) <table><tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>S1</td><td>A</td><td>A</td><td>T</td><td>G</td><td>C</td></tr><tr><td>S2</td><td>A</td><td>T</td><td>T</td><td>G</td><td>A</td></tr><tr><td>S3</td><td>A</td><td>A</td><td>C</td><td>G</td><td>C</td></tr><tr><td>S4</td><td>G</td><td>T</td><td>C</td><td>G</td><td>A</td></tr></table> (b)	**2**
4A.	Can the machine learning methods used to predict the protein secondary structure be used for RNA secondary structure prediction? Give reasoning.	**2**
4B.	Predict the secondary structure of the given RNA sequence using Maximum Base Pair Matching Algorithm. G A G C C U U C U C	**8**
5A.	You were asked to induce multiple mutations in the sequence given in 4B 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?	**3**
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.	**7**