Roll No.											

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – MAY 2025

Date: 28-05-2025 **Time:** 01:00 pm – 04:00 pm

School: SOCSE	**Program:** CSE	
Course Code: CSE3069	**Course Name:** Introduction to bioinformatics	
Semester: IV	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**26**	**26**	**24**	**24**	**-**

Instructions:

(i) Read all questions carefully and answer accordingly.
(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks. **10Q x 2M=20M**

1.	Each of the 20 amino acids has a common core structure, outline the representation.	**2 Marks**	**L1**	**CO1**
2.	Define the importance of bioinformatics.	**2 Marks**	**L1**	**CO1**
3.	Translate reverse complement of: TACGGCTAAATTGCCATGCTAT.	**2 Marks**	**L2**	**CO1**
4.	Expand DDBJ and BLAST	**2 Marks**	**L1**	**CO2**
5.	Explain secondary database with an example	**2 Marks**	**L2**	**CO2**
6.	Briefly explain SAM file format	**2 Marks**	**L2**	**CO2**
7.	Briefly explain progressive alignment method (PAM).	**2 Marks**	**L2**	**CO3**
8.	Define match, mismatch and gaps in sequence alignment.	**2 Marks**	**L1**	**CO3**
9.	Differentiate sequence motif and structure motif.	**2 Marks**	**L2**	**CO4**
10.	Define phylogenetic.	**2 Marks**	**L1**	**CO4**

Part B

Answer the Questions. **Total Marks 80M**

11.	**a.**	With a neat diagram, Explain the basic unit of life that is responsible for all life's processes.	**10 Marks**	**L2**	**CO1**
	b.	Explain the four major biological macromolecules in detail.	**10 Marks**	**L2**	**CO1**
		Or			
12.	**a.**	Differentiate between DNA and RNA with their structure.	**10 Marks**	**L2**	**CO1**
	b.	Explain various types of omics in detail.	**10 Marks**	**L2**	**CO1**

13.	**a.**	Explain two substitution matrix with an example.	**10 Marks**	**L2**	**CO2**
	b.	Demonstrate the reconstruction of substring for the following sequence: 'AGTC', 'GTCA', 'TCAG', 'CAGT', 'AGTT', 'GTTA', 'TTAG', 'TAGC', 'AGCT', 'GCTA', 'CTAT', 'TATG', 'ATGC', 'TGCA', 'GCAT'.	**10 Marks**	**L3**	**CO2**
		Or			
14.	**a.**	With suitable examples, explain local and global alignment of sequences.	**10 Marks**	**L2**	**CO2**
	b.	Illustrate approximate pattern matching for DNA sequence SSS: " TGCATCGATCGTACGTAGCTAGCAT" Pattern P: " CGTAC" with maximum mismatches d: 3	**10 Marks**	**L3**	**CO2**

15.	**a.**	Calculate the alignment score of following alignment: S1: AGCTTTCGAACCTTACGATCGACTGGAGCTAGCTTAGC S2: AG--TACGATAGGACGAAGG---GATTACGATCTAGGT Scoring scheme : Match = +3, Mismatch = -2, Gap = -3	**10 Marks**	**L3**	**CO3**
	b.	Apply the Needleman-Wunsch Algorithm to the following sequences and present the global alignment: Seq1 = GCATGCG Seq2 = GATTACA Match: +1 , Mismatch: -1, Gap: -1	**10 Marks**	**L3**	**CO3**
		Or			
16.	**a.**	Illustrate 3 components of sequence alignment with suitable example.	**10 Marks**	**L3**	**CO3**
	b.	Apply the Smith-Waterman Algorithm to the following sequences and present the local alignment: Seq1 = ACCGTGA Seq2 = GTGAATA Match: +1, Mismatch: -1, Gap: -1	**10 Marks**	**L3**	**CO3**

17.	**a.**	Solve by identifying the GATTACA motif in the following sequences using GIBBS algorithm. A**CCATGAC**AG **GAGTATA**CCT CATGCTTACT (random selected sequence) CGG**AATGCAT**	**10 Marks**	**L3**	**CO4**
	b.	Interpret the information provided by a sequence logo with an example and its relevance in bioinformatics.	**10 Marks**	**L3**	**CO4**

Or					
18.	**a.**	Illustrate the significance of motifs in the study of genomic sequences.	**10 Marks**	**L3**	**CO4**
	b.	Apply the UPGMA algorithm to construct the phylogenetic tree, and interpret the evolutionary relationships between the sequences. Discuss the implications of these relationships in the context of their genetic similarities and differences. 1. GGTCAGTCAA 2. CCTAGGTTGA 3. TGGCATACTG 4. ATGCGTAGCT 5. GATTACAAGT	**10 Marks**	**L3**	**CO4**