|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

 ****

**Presidency University**

**Bengaluru**

 **School Of Computer Science and Engineering & Information Science**

**Summer Term End-Term Examinations, August 2024**

**Date**: 9.8.2024

**Time**: 9.30 AM TO 12.30 PM

**Max Marks**: 100

**Weightage**: 50%

**Odd Semester**: 2023 - 24

**Course Code**: CSE3069

**Course Name**: INTRODUCTION TO BIOINFORMATICS

**Department: CSE**

 **Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Do not write any matter on the question paper other than roll number.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q.No** | **Questions** | **Marks** | **CO** | **RBT** |
| 1 | 1. To solve the String Reconstruction Problem and to "connect" a pair of k-mers a symbol is used,Mention which symbol with suitable reason.
 | 4 | CO1 | L1 |
| 1. Data has to be stored in a standardized plain text format for nucleotide sequence. Discuss which kind of common sequence file formats are available with neat example .
 | 6 | CO1 | L2 |
| 1. Demonstrate how DNA sequences plays a very important for life of all organism and mention its Six types
 | 10 | CO1 | L3 |
| OR |
| 2 | 1. Deposition of cDNA into the inert structure is called as what and explain the process.
 | 4 | CO1 | L1 |
| 1. Discuss in detail about four major types of Biological macromolecules .
 | 6 | CO1 | L2 |
| 1. Discuss in detail about four major types of Biological macromolecules .
 | 10 | CO1 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3 | 1. FASTA format is a text based format for representing which kind of sequences with an example
 | 4 | CO2 | L1 |
| 1. Guide tree is constructed by progressive alignment method ,discuss how it is done and give the details by elaborating on Clustal w and T Coffee.
 | 6 | CO2 | L2 |
| 1. Using Needle man wunsh ,find what is the aligned sequence score obtained in the matrix? Seq1=ATGCT, Seq2=AGCT. Consider the value for Match=1, Mismatch=-1, Gap Penalty= -2.
 | 10 | CO2 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4 | 1. *Omics* refers to the collective technologies used to explore the roles, relationships.List the types of omics.
 | 4 | CO2 | L1 |
| 1. Compare and contrast with DNA and RNA structures with neat sketches
 | 6 | CO2 | L2 |
| 1. For S1 = AATGC and S2 = GATG, find the alignment using Dynamic Programming algorithm for local Alignment using smith waterman. Given scores: Match = +2, Mismatch = -2 and Progressive Gap Penalty = -1.
 | 10 | CO2 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5 | 1. Identify the abbreviation of PSI BLAST and specify where it is used in protein structure.
 | 4 | CO3 | L1 |
| 1. Infer the final sequence from a common ancestor to compare and when we apply string reconstruction for the following reads?AAG ACC AGT ATG ATT CAT CCT CGG GAC GAT GGA GTC TCA TCG TGA TTC
 | 6 | CO3 | L2 |
| 1. Reconstruct the substring and graph for the following sequence: ‘TACG', ‘CAGC', 'CAAG', ‘TCAG', 'AGGT', 'CGAT', 'GATC', ‘TTCC', 'AAGG', 'ACGA', 'CGAT', 'GTAC', 'ATCG', 'ATCG', 'TCGA', 'GATC', 'TCGG', 'CGGA', ‘CATC', 'GGTA', 'GATC', 'ATCA', 'TCCA', ‘GGAT', 'CCAA', 'AGCA', 'GCAT', 'TCCA‘, 'ATCC'.
 | 10 | CO3 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 6 | 1. Identify the last step in the dynamic programming and explain how the process work.
 | 4 | CO3 | L1 |
| 1. What is Biological Database? Explain the Database Architecture and types of database in detail.
 | 6 | CO3 | L2 |
| 1. In a application of human genome an algorithm is required for comparing primary biological sequence information ,Explain the procedure followed in the algorithm of BLAST with detailed 12 steps.
 | 10 | CO3 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 7 | 1. Reproduce the Composition3(GGGTATCAA) in lexicographic orderz
 | 4 | CO4 | L1 |
| 1. Name the process of subjecting a DNA, RNA or peptide sequence to any of a wide range of analytical methods to understand its features, function, structure, or evolution and explain.
 | 6 | CO4 | L2 |
| 1. Discuss about accurate alignment of cDNA sequences to a genome, Explain how the Hidden Markov Model works with respect to accuracy in alignment.
 | 10 | CO4 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 8 | 1. Identify most widely used approach for multiple alignments and explain its advantage.
 | 4 | CO4 | L1 |
| 1. State any 2 database which is used as the search and retrieval tool of NCBI with explanation.
 | 6 | CO4 | L2 |
| 1. *Omics* refers to the collective technologies used to explore the roles, relationships. List and explain the types of omics.
 | 10 | CO4 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 9 | 1. Explain the terminologies:
2. Gene
3. mRNA
 | 4 | CO1 | L1 |
| 1. Infer the score for the following two sequences,Seq1:CCTCTCACTGAGACTCAGCC Seq2:CGA-ACTCTT-GATC-TGC considering the match =1,mismatch -1,gap =0.
 | 6 | CO1 | L2 |
| 1. Gap penalties are used to adjust alignment scores based on the number and length of gaps between two sequences . Explain the role of gap penalty in this and its types.
 | 10 | CO1 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 10 | 1. Bioinformatics is made an attempt to align the entire sequence, using as many characters as possible, up to both ends of each sequence. State which type of alignment is used with justification .
 | 4 | CO2 | L1 |
| 1. Biological macromolecules is correctly paired with one of its functions which is very necessary for all organisms for life ,List the 4 types of molecules.
 | 6 | CO2 | L2 |
| 1. Construct the string and graph for the following sequence:‘ATTCG', 'GTCAT', ‘ATGAC', 'TCATT‘, 'GACCT', ‘GATGA', ‘AGTCA', 'AAGTC', 'TTCGG', ‘TGACC', 'CGGAT', 'GGATG’, ‘TCGGA' , ‘CATTC‘.
 | 10 | CO2 | L3 |