Roll No

PRESIDENCY UNIVERSITY
BENGALURU

SET B

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JAN 2024

Semester : Semester V - 2021
Course Code : CSE3069
Course Name : Introduction to Bioinformatics
Program : B.Tech.

Date : 08-JAN-2024
Time : 9:30AM - 12:30 PM
Max Marks : 100
Weightage : 50%

Instructions:
(i) Read all questions carefully and answer accordingly.
(ii) Question paper consists of 3 parts.
(iii) Scientific and non-programmable calculator are permitted.
(iv) Do not write any information on the question paper other than Roll Number.

PART A

ANSWER ALL THE QUESTIONS **5 X 2M = 10M**

1. Which of the following is a protein sequence database?
   a. PIR
   b. Genbank
   c. EMBL
   d. DDBJ
(CO1) [Knowledge]

2. What is the reverse complement of: GATTGATACGAATCTAGCT?
   a. AGGTAGATTCGTATCAATC
   b. AGCTAGATTGCTATCAATC
   c. AGCTACATTCGTATCAATC
   d. AGCTAGATTCGTATCAATC
(CO1) [Knowledge]

3. Enumerate the 2 commonly used substitution matrics.
(CO2) [Knowledge]

4. Which of the following are not part of PIR file format?
   a. P1
   b. N1
   c. R1
   d. D1
(CO3) [Knowledge]

5. Clustal W is a:
   a. Protien sequencing analysis tool
   b. Data retrieving tool
   c. Nucliec acid analysis tool
   d. Multiple sequence alignment tool
(CO4) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS **5 X 10M = 50M**

6. Discuss the below given concepts in detail
i. Substitution Matrix and its types (5 Marks)
ii. Types of Gap Penalties (5 Marks)

(CO1) [Comprehension]

7. Solve the following
i. Countd(Text, Pattern).
Pattern: GAGG
Text: TTTAGAGCCTTCAGAGGCTCGTCG
d(max mismatches): 2

ii. Reconstruct the substring using the following k-mers
'TTCC', 'CCAA', 'CAAG', 'TCCA', 'AGGT', 'CGAT', 'GTAC', 'TACG', 'AAGG', 'ACGA', 'CGAT', 'GATC', 'ATCG', 'ATCG', 'TCGA', 'GATC', 'TCGG', 'CGGA', 'GGAT', 'GGTA', 'GATC', 'ATCA', 'TCAG', 'CATC', 'CAGC', 'AGCA', 'GCAT', 'TCCA‘, 'ATCC'

(CO2) [Comprehension]

8. Explain the following concepts in detail with examples.
i. Types of Gap Penalties. (5 Marks)
ii. Substitution Matrix and its types. (5 Marks)

(CO2) [Comprehension]

9. Illustrate the concepts of (i) Sequence Alignment and its types (ii) Clustal W

(CO3) [Comprehension]

10. a) What is the significance of Motif? Explain in detail the 2 types of motifs.
b) Trace the sequence logo for the following sequences using the "Profile Analysis" method.

CCCATTGTTCTC
TTTCTGGTTCTC
TCAATTGTTTAG
CTCATTGTTGTC
TCCATTGTTCTC
CCTATTGTTCTC
TCCATTGTTCGT
CCAATTGTTTTG

(CO4) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS **2 X 20M = 40M**

11. Construct a Phylogenetic Tree for the following given sequences using UPGMA Method
ATCGTGGTACTG
CCGGAGAACTAG
AACGTGCTACTG
ATGGTGAAAGTG
CCGGAAAACTTG
TGGCCCTGTATC

(CO3) [Application]

12. A. What is the significance of Motif? Explain the 2 types of motifs in detail. (10 Marks)

B. Using "Profile Analysis" method, construct the sequence logo for the given sequences. (10 Marks)
TCAATTGTTTAG
TCCATTGTTCGT
TCCATTGTTCTC
TTTCTGGTTCTC
CCCATTGTTCTC
CCTATTGTTCTC
CTCATTGTTGTC
CCAATTGTTTTG

(CO4) [Application]