

**PRESIDENCY UNIVERSITY  
BENGALURU****SET A****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. For the alignment rules Match=+1, Mismatch=-1 and Gap=0, what is the alignment score for S1= CGA-  
ACTCTT-GTTC-TGCA and S2= CCTCTCACTGAGACTCAGCC
  - a. -2
  - b. 0
  - c. -1
  - d. -3

(CO1) [Knowledge]
2. The \*(asterisk) in PIR file format is used for what purpose?
  - a. To mark the beginning of the sequence
  - b. To indicate unkonwn sequence
  - c. To mark the end of the sequence
  - d. To indicate the details of the sequence

(CO1) [Knowledge]
3. List any 2 applications of k-mers.

(CO2) [Knowledge]
4. All are nucleotide sequence databases, except
  - a. DDBJ
  - b. Swissprot
  - c. Genbank
  - d. EMBL

(CO3) [Knowledge]
5. Characterizing molecular component is:
  - a. Bioinformatics
  - b. Proteomics
  - c. Genomics
  - d. None of the above

(CO4) [Knowledge]

## PART B

ANSWER ALL THE QUESTIONS

5 X 10M = 50M

6. Alice wants to understand the working of BLAST algorithm. Explain the same in detail.  
(CO1) [Comprehension]
7. i. Identify all the starting positions where Pattern appears as a substring of Text with at most 'd' mismatches.  
**Pattern:** ATTCTGGA  
**Substring:** CGCCCGAATCCAGAACGCATTCCCATATTTTCG  
GGACCACTGGCCTCCACGGTACGGACGTCAATCAAAT  
**d:** 3
- ii. Reconstruct the substring from the given k-mers  
AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT  
(CO2) [Comprehension]
8. Describe the following in detail  
(i) Global and Local Alignments with examples. (5 Marks)  
(ii) FASTQ & FASTA with examples. (5 Marks)  
(CO2) [Comprehension]
9. Describe the following in detail (i) Global and Local Alignments with examples (ii) Clustal W  
(CO3) [Comprehension]
10. a) Define Phylogenetic and Phylogenetic Tree? Explain the elements of Phylogenetic tree.  
b) Trace the phylogenetic tree using the UPGMA method for the following sequences  
i. ATCGATCCTA  
ii. ATATACTCAT  
iii. CTCGATCCAT  
iv. ATATTTCACT  
v. ATCCATCCTC  
(CO4) [Comprehension]

## PART C

ANSWER ALL THE QUESTIONS

2 X 20M = 40M

11. A. Find All starting positions where Pattern appears as a substring of Text with at most 'd' mismatches. (10 Marks)  
**Pattern:** CGCCTGGA  
**Text:**  
CGCCCGAATCCAGAACGCATTCCCATATTTTCGGGACCACTGGCCTCCACGGTACGGACGTCAATCAA  
**d = 3**
- B. Reconstruct the substring from the given k-mers. (10 Marks)  
TGC GCC ATG GAT TGT CAT ATG CCA GGA AAT GTT TAA ATG TGG GGG  
(CO3) [Application]
12. Find the **GATTACA** motif in the following sequences using **Gibbs Sampling**  
a. ACCATGACAG  
b. GAGTATACCT  
c. CATGCTTACT  
d. CGGAATGCAT  
(CO4) [Application]

