ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU -27
M.Sc (BIG DATA ANALYTICS) – III SEMESTER
SEMESTER EXAMINATION: OCTOBER 2022
(Examination conducted in December 2022)
BDADE3621 - BIOINFORMATICS

Time: 2 ½ Hours   Max Marks: 70

This paper contains 3 printed pages and 3 parts

PART-A

Answer ALL questions   **10×1 = 10**

1. NR_12345 identifier belongs to _____ database
   a. RefSeq
   b. GenBank
   c. SwissProt
   d. UniProt

2. _____ is present only in eukaryotes.
   a. Cell membrane
   b. Ribosome
   c. Cell wall
   d. Endoplasmic reticulum

3. What is a 'Query coverage' in BLAST?
   a. Percentage of the query covered in the subject
   b. Percentage of the query covered in the database
   c. Percentage of the sequences in the database
   d. Percentage of the bases in the database

4. The 'Emulsion-PCR' method is used in_____
   a. Illumina sequencers
   b. Roche 454 sequencers
   c. Pacific Biosciences sequencers
   d. All the above

5. Which among the following is known as a 'third generation sequencer'?
   a. Roche 454

   b. Illumina Genome Analyzer

   c. Pacific Biosciences sequencers

   d. Illumina HiSeq sequencers

6. A 'read' in an NGS experiment indicates_____
   a. A raw sequence from a next generation sequencer

   b. Complete DNA sequence

   c. Method of reading the bases in the DNA

   d. Mapped reference sequences

7. De-novo assembly is generally used when
   a. Raw data is not of good quality

   b. The reference genome is available

   c. No reference genome is available

   d. There are not enough reads

8. Which of these is NOT a protein structure prediction method?
   a. Ab initio

   b. In silico

   c. Threading

   d. Homology modeling

9. Which of the databases stores 3D protein structures?
   a. Protein Data Bank

   b. Protein Data Browser

   c. Primary Data Bank

   d. Primary Data Base

10. The blast program that will help in matching a query nucleotide sequence with a subject protein sequence?
    a. BLASTN

b. BLASTP

c. Both A and B

d. BLASTX

PART-B

Answer any SIX questions   **6×5 = 30**

11. Define genomics, transcriptomics, and proteomics along with their applications. Write a note on various types of high-throughput data that are generated in these fields

12. What are the various applications of next generation sequencing in research and healthcare?

13. Write a detailed note on the usage of any 5 bioinformatics databases/tools in biological research.

14. Differentiate between RefSeq and GenBank databases.

15. What are the applications of multiple sequence alignment? Write a detailed note on various features of the ClustalOmega tool.

16. What is central dogma? How is it used by the cells to propagate?

17. Write a detailed note on various protein structure prediction methods.

18. What are the major differences between bacterial and human cells? Write the major functions of at least 3 cell organelles that are present in humans.

PART-C

Answer any THREE questions   **3×10 = 30**

19. Suggest a broad strategy for developing a computational method: There are 100 DNA samples from normal individuals and another 100 from patients with a genetic disorder. How would you compare the two sets of genomic DNA and find any existing difference that is associated with the specific genetic disorder?

20. What is reference assembly? Discuss in detail the methodology of Illumina and Roche technologies.

21. Discuss the importance of sequence analysis. List various features of the BLAST tool and important parameters for choosing the best alignment.

22. Discuss in detail the different steps of homology modeling.