

**M. SC. (BIOTECHNOLOGY) EXAMINATION, 2022**

( 1st Year, 2nd Semester)

**SUBJECT : BIOINFORMATICS****PAPER : MSBT 235**

Time : Two hours

Full Marks : 40

**Answer all the questions.**

1. Answer any ten questions:  $1 \times 10 = 10$
- What is the primary and secondary database?
  - What makes BLAST more useful than FASTA for database search?
  - What is the basic structure of a GenBank and an EMBL entry?
  - Write the name of the scoring model used to build similarity between nucleotide sequences?
  - For what purpose BLOSUM matrices are used?
  - What is the FASTA format?
  - Which one has been used to search all NCBI associated databases?
    - BLAST
    - CLUSAL W
    - ENTREZ
    - PDB
  - If you want to BLAST the non-redundant database using a new protein sequence as query, which is the BEST search program to use?
    - blastp
    - blastn
    - tblastx
    - blastx
  - Submission to GenBank include
    - Sequin and BankIt
    - Sequin
    - BankIt
    - BankIt and sequeen
  - Name two different methods used in pairwise sequence alignment.
  - What is T-Coffee algorithm?
  - Define "Sum of pairs" method for scoring the multiple sequence alignments
2. Answer **any four** questions:  $2.5 \times 4 = 10$
- Write the two most commonly used scoring matrices.
  - Write down the differences between CLUSTAL W and MUSCLE in multiple sequence alignments?
  - What are the phylogenetic tree construction methods that use direct sequence information?
  - What are Cladogram and Phylogram in phylogenetic analysis? Explain with schematic diagrams What do you mean bootstrap method?
  - What do you mean by identity and similarity between two protein sequences?
  - What is the Progressive alignment method in multiple sequence alignments? What are the major problems with it ?
3. Answer **any three** of the following:  $4 \times 3 = 12$
- What is DNA sequencing chromatogram? Explain how to interpret it?  $2+2$
  - What is Homologs, Paralogs, Orthologs. Explains them with examples.  $2+2$
  - What do you mean wire-frame diagram, Balls and sticks, space-filling representation, Ribbon diagrams of the protein structure?  $4$
  - Define the Total Score, Max Score, Query coverage and E-value, Identity in the context of sequence alignment  $4$
  - Write the notes for the two most popular methods for protein structure classification.  $2+2$
4. Answer **any one** of the following:  $8 \times 1 = 8$
- Describe the two important methods used in secondary structure prediction of Proteins  $4+4=8$
  - Answer the following questions about this tree:
 

```

graph LR
    Root --- Node1
    Node1 --- Dwarf
    Node1 --- Node2
    Node2 --- Troll
    Node2 --- Node3
    Node3 --- Gremlin
    Node3 --- Node4
    Node4 --- Hippogriff
    Node4 --- Griffin
    Node4 --- Gump
      
```

    - Which are the two most closely related species?
    - Which of these is probably the outgroup?
    - Which is most "distantly related to the Gremlin other than the outgroup species?
    - Write down the few basic characteristics of an outgroup.  $[1+1+2+4]$
  - Write a note on Position-Specific Scoring Matrices (PSSMs), Profiles and Hidden Markov Models (HMMs)  $8$