## M. Sc. (BIOTECHNOLOGY) EXAMINATION, 2022

(1st Year, 2nd Smester)

**SUBJECT: BIOINFORNATICS** 

Paper: MSBT 235

Time: Two hours Full Marks: 40

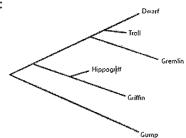
## Answer all the questions.

- 1. Answer any ten questions:  $1 \times 10 = 10$ 
  - a) What is the primary and secondary database?
  - b) What makes BLAST more useful than FASTA for database search?
  - c) What is the basic structure of a GenBank and an EMBL entry?
  - d) Write the name of the scoring model used to build similarity between nucleotide sequences?
  - e) For what purpose BLOSUM matrices are used?
  - f) What is the FASTA format?
  - g) Which one has been used to search all NCBI associated databases?
    - i) BLAST
- ii) CLUSAL W
- iii) ENTREZ iv) PDB
- h) If you want to BLAST the non-redundant database using a new protein sequence as query, which is the BEST search program to use?
  - i) blastp ii) blastn iii) tblastx iv) blastx
- i) Submission to GenBank include
  - i) Sequin and Banklt ii) Sequin
  - iii) Banklt
- iv) Banktl and sequeen
- j) Name two different methods used in pairwise sequence alignment.
- K) What is T-Coffee algorithm?
- 1) Define "Sum of pairs" method for scoring the multiple sequence alignments
- 2. Answer *any four* questions:  $2.5 \times 4=10$ 
  - a) Write the two most commonly used scoring matrices.
  - b) Write down the differences between CLUSTAL W and MUSCLE in multiple sequence alignments?
  - c) What are the phylogenetic tree construction methods that use direct sequence information?
  - d) What are Cladogram and Phylogram in phylogenetic analysis? Explain with schematic diagrams What do you mean bootstrap method?
  - e) What do you mean by identity and similarity between two protein sequences?
  - f) What is the Progressive alignment method in multiple sequence alignments? What are the major problems with it?
- 3. Answer *any three* of the following:  $4 \times 3 = 12$ 
  - a) What is DNA sequencing chromatogram? Explain how to interpret it? 2+2

- b) What is Homologs, Paralogs, Orthologs. Explains them with examples. 2+2
- c) What do you mean wire-frame diagram,
  Balls and sticks, space-filling representation,
  Ribbon diagrams of the protein structure? 4
- d) Define the Total Score, Max Score, Query coverage and E-value, Identity in the context of sequence alignment 4
- e) Write the notes for the two most popular methods for protein structure classification. 2+2
- 4. Answer *any one* of the following:  $8 \times 1=8$ 
  - a) Describe the two important methods used in secondary structure prediction of Proteins

4+4=8

b) Answer the following questions about this tree:



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