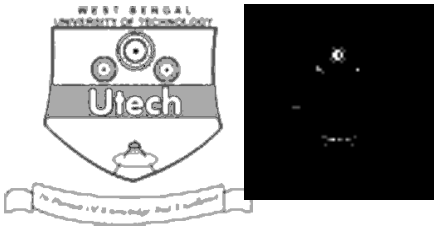
CS/M.Tech (BT)/SEM-2/MBT-201/09
BIOINFORMATICS (SEMESTER - 2)



1. .....
Signature of Invigilator

2. .....
Signature of the Officer-in-Charge

Reg. No.

Roll No. of the Candidate

CS/M.Tech (BT)/SEM-2/MBT-201/09
ENGINEERING & MANAGEMENT EXAMINATIONS, JULY – 2009
BIOINFORMATICS (SEMESTER - 2)

Time : 3 Hours] [Full Marks : 70

INSTRUCTIONS TO THE CANDIDATES :

1. This Booklet is a Question-cum-Answer Booklet. The Booklet consists of **32 pages**. The questions of this concerned subject commence from Page No. 3.
2. a) In **Group – A**, Questions are of Multiple Choice type. You have to write the correct choice in the box provided **against each question**.
   b) For **Groups – B** & **C** you have to answer the questions in the space provided marked 'Answer Sheet'. Questions of **Group – B** are Short answer type. Questions of **Group – C** are Long answer type. Write on both sides of the paper.
3. **Fill in your Roll No. in the box** provided as in your Admit Card before answering the questions.
4. Read the instructions given inside carefully before answering.
5. You should not forget to write the corresponding question numbers while answering.
6. Do not write your name or put any special mark in the booklet that may disclose your identity, which will render you liable to disqualification. Any candidate found copying will be subject to Disciplinary Action under the relevant rules.
7. **Use of Mobile Phone and Programmable Calculator is totally prohibited in the examination hall.**
8. You should return the booklet to the invigilator at the end of the examination and should not take any page of this booklet with you outside the examination hall, **which will lead to disqualification**.
9. Rough work, if necessary is to be done in this booklet only and cross it through.

No additional sheets are to be used and no loose paper will be provided

---

FOR OFFICE USE / EVALUATION ONLY

Marks Obtained

	Group – A										Group – B			Group – C				
Question Number																	***Total Marks***	***Examiner's Signature***
Marks Obtained																		

.....
Head-Examiner/Co-Ordinator/Scrutineer

51001 (01/07)

DO NOT WRITE ON THIS PAGE

CS/M.Tech (BT)/SEM-2/MBT-201/09

BIOINFORMATICS

SEMESTER - 2

Time : 3 Hours] [Full Marks : 70

GROUP – A

(Multiple Choice Type Questions)

1. Choose the correct alternatives for any *ten* of the following : 10 ∞ 1 = 10

i) The first bionformatic database was created by

a) Richard Durbin b) Steven Altschul

c) Dayhoff d) David Lipman. ☐

ii) Which of the following database can be used to access protein domain information ?

a) Prosite b) SANGER

c) DDBJ d) KEGG. ☐

iii) The Protein Data Bank (PDB)

a) functions primarily as the major repository of macromoleculer secondary structures

b) contains approximately as many structures as there are protein sequences in SwissProt/TrEMBL.

c) includes data on proteins, DNA-protein complexes as well as carbohydrates

d) is operated jointly by NCBI and EBI. ☐

iv) How dose the BLOSUM differ from PAM ?

a) It is best used for aligning very closely related proteins

b) It is based on local multiple alignments from distantly related proteins

c) It is based on global multiple alignment from closely related proteins

d) It combines local and global alignment information. ☐

v) Which of the following compares a protein query sequence against a translated nucleotide sequence library ?

a) FASTA b) TFASTA

c) FASTF d) FASTX. ☐

vi) As the *E* value of a BLAST search become smaller,

a) *K* value also becomes smaller

b) score tends to be larger

c) probability *p* tends to be larger

d) the extreme value distribution becomes less skewed. ☐

vii) Which one of the following is a not a distance based phylogenetic algorithm ?

a) UPGMA b) Maximum parsimony

c) Transformed method d) Neighbour joining. ☐

viii) An example of operational taxonomic unit OTU is

a) multiple sequence alignment

b) clade

c) protein sequence

d) node. ☐

ix) The main difference between Pfam-A and Pfam-B is that

a) Pfam-A is manually curated while Pfam-B is automatically curated

b) Pfam-A uses HMMs while Pfam-B does not

c) Pfam-A provides full length protein alignments while Pfam-B aligns protein fragments.

d) Pfam-A incorporates data from SMART an PROSITE while Pfam-B does not.

x) In a position-specific scoring matrix, the score for any given amino acid residue is assigned based on

a) a PAM or BLOSUM matrix

b) its background frequency of occurrence

c) the score of its neighbouring amino acid

d) its frequency of occurrence in an MSA.

xi) Which of the following sentences best describe the difference between global alignmnet and local alignment ?

a) Global alignment is usually used for DNA sequences, while local alignment is usually used for protein sequences

b) Global alignment has gaps, while local alignment does not have gaps

c) Global alignment finds the global maximum, while local alignment finds the local maximum

d) Global alignment aligns the whole sequence, while local alignment finds the best subsequence that aligns.

xii) The BLAST algorithm compiles a list of word. Words at or above a threshold value *T* are defined as

a) Hits and are used to scan the database for exact matches that may then be extended

b) Hits and are used to scan the database for exact or partial matches that may then be extended

c) Hits and are aligned to each other

d) Hits and are reported as raw score.

GROUP – B

(Short Answer Type Questions)

Answer any *three* of the following. 3 ∞ 5 = 15

2. How can you classify sequence databases ? Explain the structure of EMBL entries.

3. Describe progressive alignment for MSA in detail.

4. What is a phyogenetic tree ? Why are trees useful ?

5. Describe in words how BLAST algorithm works and explain different BLAST programs.

6. Provide detail structure of Transformed Distance method of phylogenetic analysis algorithm.

7. Define profile. Explain principle of Profile based searching method (PSI-BLAST).

GROUP – C

(Long Answer Type Questions)

Answer any *three* of the following. 3 ∞ 15 = 45

8. a) Compute the global alignment between the two sequences S1 = ACCGTT and S2 = AGTTCA, considering the following scoring parameters : + 1 for match, – 1 for mismatch and – 1 for a gap.

   b) What is the maximum similarity score between the two sequences S1 and S2 ?

   c) Find an alignment with this similarity score.

   d) Is the alignment you found unique, or are there multiple alignments achieving the maximum similarity score ? 10 + 1 + 2 + 2

9. a) Describe about Phylogenetic tree construction by using UPGMA method.

   b) Build the tree from the following distance matrix between species *A*, *B*, *C*, *D* using the UPGMA method :

	A	***B***	***C***	***D***
A	0			
B	0·26	0		
C	0·34	0·42	0	
D	0·29	0·44	0·44	0

   8 + 7

10. What is Scoring or Substitute matrix ? Describe Block Substitution Matrices (BLOSUM) in detail. 15

11. What is Dynamic Programming ? Discuss the Smith-Waterman algorithm in detail.

    5 + 10

12. Write short notes on any *three* of the following : 3 ∞ 5

a) Gribskov profile analysis

b) FASTA

c) Maximum Parsimony

d) Pfam

e) EMBOSS.

END