CS/B.Sc (H), BT, MolBio, Genetics, MicroBio/SEM-4/CBB-402/2011

2011
COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

Time Allotted : 3 Hours   *Full Marks :* 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

GROUP – A
(Multiple Choice Type Questions)

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

    i)   Blastx is a tool of

        a)   EBI

        b)   NCBI

        c)   SWISS Institute

        d)   Cambridge Institute of Bioinformatics.

    ii) NCBI contains

        a)   Primary databases

        b)   Derivative databases

        c)   both Primary and Derivative databases

        d)   none of these.

iii) Genbank file ends with

a) //

b) \\

c) ENDML

d) > !

iv) The database that will help geneticists most is

a) RESS   b) KEGG

c) OMIM   d) BLISS.

v) Pairwise sequence alignment is associated with

a) The Baum-Welch Algorithm

b) The Viterbi Algorithm

c) Needleman-Wunch Algorithm

d) None of these.

vi) Data mining is all about automating the process of searching for

a) classification, clustering, pattern search and outlier detection

b) data warehousing

c) primer design

d) none of these.

vii) PDB is an example of

a) Protein 3d info database

b) Protein family info database

c) Nucleotide 3d info database

d) None of these.

viii) The low complexity region among the following is

    a)   'aaggtcctagtagtcga'

    b)   'aaaaaaaaaa'

    c)   'atttattaaagcgctgcat'

    d)   'nnnnnnna'

ix)   Dynamic programming is used to

    a)   develope the phylogenetic tree

    b)   detect the most suitable alignment

    c)   detect the conserved domain of different protein

    d)   none of these.

x)   .... indicates the proteins expressed by a genome.

    a)   Proteinex   b)   Trancriptome

    c)   Protease   d)   Proteome.

xi)   Protein primary structure contains

    a)   peptide bond

    b)   hydrogen bond

    c)   disulphide bridge

    d)   none of these.

xii)   In sequence of DP algorithm

         I.   Initialization

         II.   Trace back

         III.   Matrix fill

    a)   I, III, II   b)   III, I, II

    c)   I, II, III   d)   any one of these.

xiii) Match the items in the SET-I to those in SET-II :

SET-I :

    a)   Knowledge base software tools

    b)   BLAST

    c)   Genome mining project

    d)   FASTA

SET-II :

    i)   Borrow from neural networks

    ii)   NN promoter recognition tools

    iii)   Altschul method

    iv)   Pearson and Lipman

a)   a-i, b-ii, c-iii, d-iv   b)   a-iv, b-iii, c-ii, d-i

c)   a-i, b-iii, c-ii, d-iv   d)   a-ii, b-iii, c-i, d-iv.

xiv) FTP stands for

    a)   File Transaction Protocol

    b)   File Transfer Protocol

    c)   File Topological Probability

    d)   File Transfer Percentage.

GROUP – B

(Short Answer Type Questions)

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

2. What is dot plot ? How can we designate the conserved domain between two sequences by DOT PLOT ? Define orthology and paralogy. 1 + 2 + 1 + 1

3. Define pairwise, global alignment and local alignment. Mention the names of the algorithms for global and local alignments. 1 + 1 + 1 + 1 + 1

4. What is sequence alignment ? Define E-value and Z-score. What is stop codon ? 1 + 3 + 1

5. What do you know about International Nucleotide and Protein Sequence Database Collaborations ? $2\frac{1}{2} + 2\frac{1}{2}$

6. What is scoring system in an alignment ? Align : "AATGCCTAGATCGAG" & "ATGCCACTCGAG". Write down the score best score taking any value with your own. 1 + 4

GROUP – C

(Long Answer Type Questions)

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

7. State the differences between homology and similarity. Write down 6 emerging fields in bioinformatics. You have given a sequence of African elephant and Indian elephant, how would you draw the phylogenetic relation between them ? What would you expect the result ? What is relative mutability ?   3 + 3 + 4 + 2 + 3

8. What is Genomics ? Write down the steps of finding a nucleotide from NCBI namely "Heamoglobin A". Convert it into protein. What is NCBI ? Find the homologue of the protein from the databases. Then find the similarity of the corresponding protein with other species. What is pair-wise alignment and *k*-tupple ? Write the use of HSSP in Blast.
          1 + 2 + 1 + 3 + 1 + 2 + 3 + 2

9. Align the following two sequences by dynamic programming method :

GAATTCAGTTA (sequence#1)

GGATCGA (sequence#2)

Write down the score. 12 + 3

10. Write main features of SCOP and CATH. Define domain of a protein. Write short notes on protein databank. 8 + 2 + 5

11. What is the Needleman-Wunsch algorithm ? You have been given a protein along with the sequence of it. How would you identify the protein ? If not possible, then how would you get an idea of the functionality of the protein ? Give proper explanation. Write a short note on blast's working procedure. 5 + 5 + 5