(i) Printed Pages : 4 Roll No. .....

(ii) Questions : 9 Sub. Code : 2 5 9 5 0

Exam. Code : 0 4 3 7

M.Sc. Bio-Technology 3rd Semester

(2124)

BIOINFORMATICS

Paper–MBIO-305

Time Allowed : Three Hours] [Maximum Marks : 80

Note :— Attempt **FIVE** questions in all. Q. No. 1 is compulsory. Select **ONE** question from each unit.

1. (Compulsory Question) Answer any **EIGHT** of the following questions :

   (i) The CATH database offers a hierarchical classification of proteins. Which are the four levels in CATH ?

   (ii) What is the principal difference between homology modeling, fold-recognition and *ab initio* structure prediction ?

   (iii) Explain SP and REM TeEMBL.

   (iv) Define Hamming and Levenshtein Distance.

   (v) How phred score is used to calculate quality of a sequence ?

   (vi) What kind of programs you will use to find similarity between distantly related sequences ?

(vii) How bootstrapping is applied in phylogenetic analysis ?

(viii) Explain alignment score and e-value.

(ix) Explain Gap opening, Gap extension and Gap penalty.

(x) Discuss UCSC brower. 8×2=16

UNIT—I

2. (a) During BLAST analysis, what will you do, if you get very few or no matches for your query ? 6

(b) Explain primary protein sequence and structure databases with examples. 10

3. (a) How composite databases help in better understanding of protein sequences ? 8

(b) Explain various DNA databases, information stored in these databases and how can we access them. 8

UNIT—II

4. (a) Explain dot plot. Explain with suitable example how we can use it for sequence analysis. 7

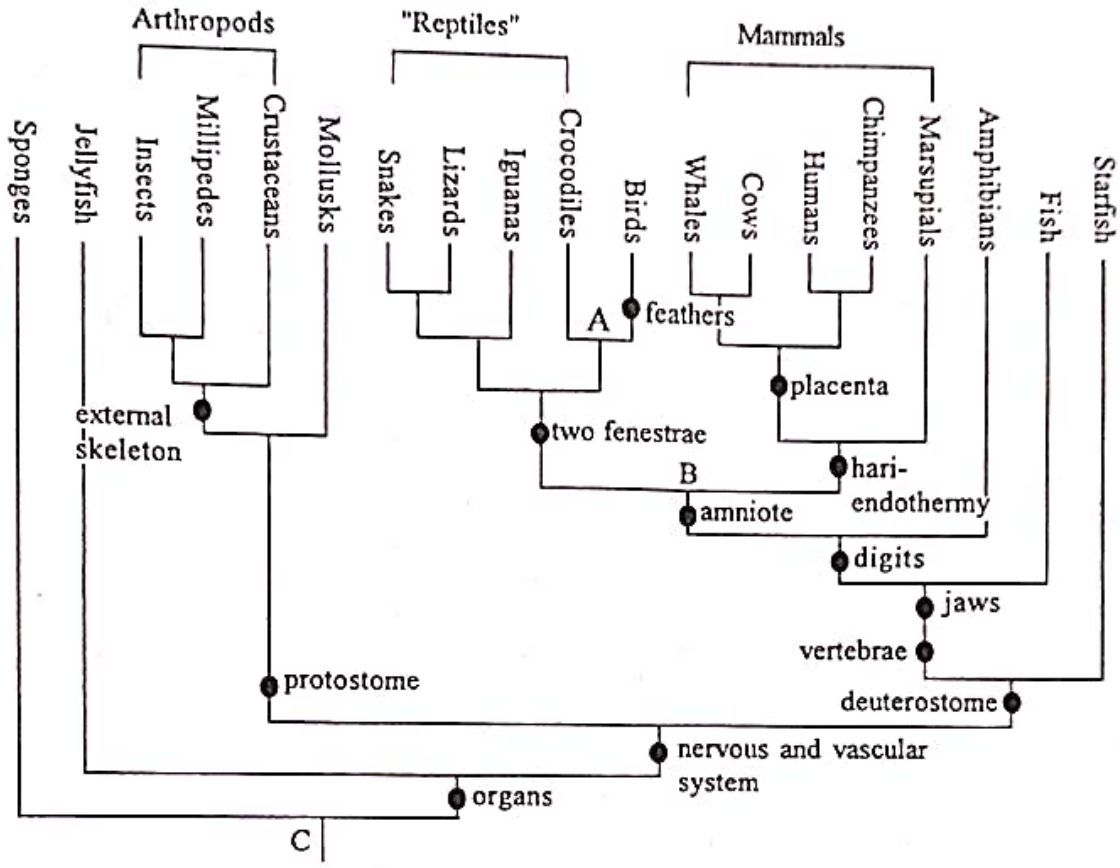
(b) Explain the matrices used for scoring the sequence alignment. Discuss, whether PAM is better or BLOSSUM and why. 9

5. (a) Discuss BLAST. How can we use BLAST to search protein and DNA databases ? 8

(b) Explain a global alignment-based algorithm. 8

UNIT—III

6. (a) Use the information in the figure to answer the following questions :



(a) What's the sister group to cows ?

(b) At the top of the tree, a bracket marks the groups that are considered to belong to the reptiles. Would you consider the reptile group, as labelled, to be a true clade ? If yes, why ? If no, why not ?'

(c) What represents the common ancestor of reptiles and mammals ?

(d) What group was used as an outgroup for this tree ?

(e) What does *A* represent ?

(f) What kind of tree is this ? 6

(b) How multiple sequence alignment is helpful in predicting the structural and regulatory aspects of proteins/genes ? 10

7. (a) Describe two programs that compute multiple sequence alignments. 6

(b) Draw the phylogenetic tree assuming four DNA sequences of at least 6 bases. Redraw the tree using maximum parsimony method. 10

UNIT—IV

8. (a) Explain the steps involved in assembly and analysis of NGS data. 8

(b) What is ENSEMBL ? How the ensembl data is useful for genome and transcriptome annotation ? 8

9. (a) How can you study the protein interaction by STRING database ? 8

(b) What is genome annotation ? Explain one method each of gene annotation based on comparative genomics and *ab initio* prediction. 8