Indraprastha Institute of Information Technology Delhi (IIITD)
Department of Computational Biotechnology
BIO213 – Introduction to Quantitative Biology
MID-SEM EXAM (March 05, 2023)

_____

Time duration: 1 hour   **Total marks:** 60

Question 1. Differentiate between **any 4** of the following: (8 marks)

1) Primary and secondary databases
2) BLOSUM and PAM substitution matrices
3) blastx and tblastx
4) Cladogram and Phylogram
5) Linear and Affine gap penalty

You can evaluate this question on your own. Make sure that the major differences have been mentioned. 2 marks each.

Question 2. Align the given two reads to make a contiguous sequence (contig) using dynamic programming approach.   (10 marks)

*R1*: LKSDSHDLKK *R2*: ASHLKSK *Scoring scheme*: Match= +4, Mismatch= -1, Gap= -5

		L	K	S	D	S	H	D	L	K	K
	0	0	0	0	0	0	0	0	0	0	0
A	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
S	0	-1	-2	3	-2	3	-2	-2	-2	-2	-2
H	0	-1	-2	-2	2	-2	7	2	-3	-3	-3
L	0	4	-1	-3	-3	1	2	6	6	1	-4
K	0	-1	8	3	-2	-4	0	1	5	10	5
S	0	-1	3	12	7	2	-3	-1	0	5	9
K	0	-1	3	7	11	6	1	-4	-2	4	9

```
---LKSDSHDLKK
ASHLKSK------
```
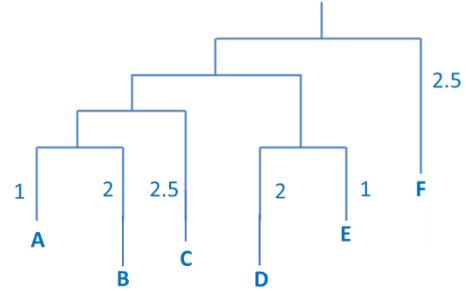
8 marks for the correct matrix, 1 mark for the correct traceback and 1 mark for showing the alignment.

Question 3. Draw the phylogenetic trees that correspond to the following Newick format. Comment whether the trees obtained are the same or different. (5 marks)

i. (((A,B)C)(D,E))F
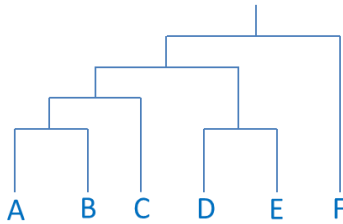
ii. (((A:1,B:2)C:2.5)(D:2,E:1))F:2.5

The (ii) tree gives information about the branch length as well. Therefore, along with the evolutionary relationship, we also get evolutionary distances between the different taxa.

2 marks for each tree and 1 mark for the comment

Question 4. Match the following. (5 marks)

1. Protein data bank   i. Nucleotide database
2. GenBank   ii. Dot plot
3. FASTA   iii. Mutations and crossover
4. Jackknifing   iv. Biomolecular structural database
5. Genetic algorithm   v. Tree Evaluation

1-iv, 2-i, 3-ii, 4-v, 5-iii (1 mark for each of the correct answers)

Question 5. Get the substitution value for P→Q. (8 marks)

PQPP
QQQP
QQPP
QPPP
QQQP

	P	Q
P	13	14
Q		13

$P(P) = \underline{10/20 = 0.5}$; $P(Q) = \underline{10/20 = 0.5}$

$e_{PQ} = \underline{0.5 * 0.5 = 0.25}$

$q_{PQ} = \underline{14} / \underline{40} = \underline{0.35}$

$lod = \log_2(q_{PQ}/e_{PQ}) = \underline{\log2 (0.35/0.25)}$
$= \log2(1.4)$
$= 0.485$

Answer is log is also acceptable.

3 marks for the matrix, 1 mark for each value in the right panel.

Question 6. How many peptides can be translated from the following DNA sequence? Mark the corresponding region(s) to show their location as well. (3 marks)

CTAATGCTCAGACCGCATTAGTT
GATTACGAGTCTGGCGTAATGAA

Two peptides are possible.

1 mark for the correct number and 2 marks for locating the regions in the DNA sequence.
Deduct 1 mark if stop codons have not been shown.

Question 7. (i) What sites in the following alignment would be informative for a parsimony analysis? (2 marks)

```
      1----------10   11---------20   21---------30
1     GAATGCTGAT     ATTCCATAAG     AAAGTACTCG
2     GGATGGTGAT     ACTTCGTAAG     AAAGTACTCG
3     GGATGATGAT     ACTTCATAAG     AAGGTACTTG
4     GGATGCTGAC     ACTTCATAAG     AAAGTACTTG
5     GGATGCTGAC     ACTCCGTAAG     AATGTACTCG
```

Informative sites: 10, 14, 16, 29
Deduct marks if wrong sites have also been given along with the correct answer.

(ii) How many rooted trees can be constructed to describe the possible relationships between 5 taxa?
105 (1 mark)

(iii) Draw any one possible rooted tree for five taxa whose nucleotides at the position under consideration are A, A, G, G, A. Label each of the internal node with the most likely candidate for the inferred ancestral sequence. What are the minimum number of substitutions required by your tree topology? (5 marks)
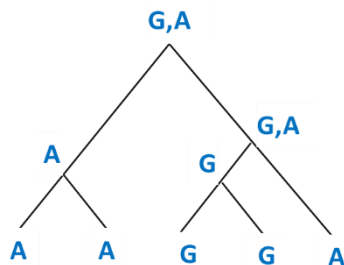
Minimum number of substitutions will be 1 (with A at the root node).
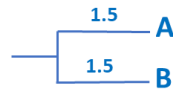
This is one of the numerous possibilities of making the tree.

Question 8. Use **unweighted** UPGMA to reconstruct a phylogenetic tree using the following distance matrix. (10 marks)

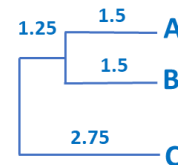| Species	A	B	C	D
B	3			
C	6	5		
D	9	9	10	
E	12	11	13	9

I.

	A	B	C	D
B	3	-	-	-
C	6	5	-	-
D	9	9	10	-
E	12	11	13	9

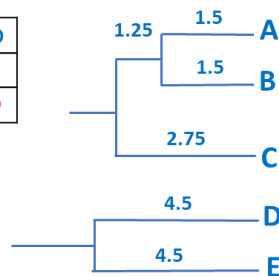**II.** d(AB,C)=[d(AC)+d(BC)]/2 = [6+5]/2 = 5.5
d(AB,D)=[d(AD)+d(BD)]/2 = [9+9]/2 = 9
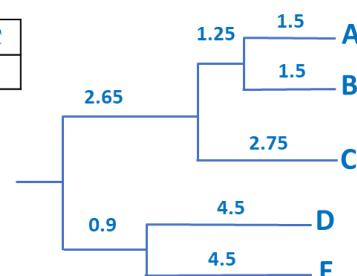d(AB,E)=[d(AE)+d(BE)]/2 = [12+11]/2 = 11.5

	AB	C	D
C	5.5	-	-
D	9	10	-
E	11.5	13	9

III. d(ABC,D)=[d(A,D)+ d(B,D)+ d(C,D)]/3 = [9+9+10]/3 = 9.3
d(ABC,E)=[d(A,E)+ d(B,E)+ d(C,E)]/3 = [12+11+13]/3 = 12
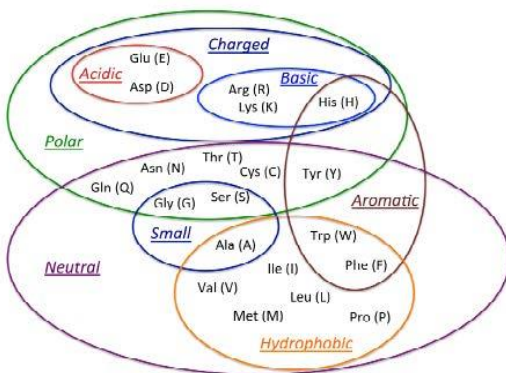
	ABC	D
D	9.3	-
E	12	9

IV. d(ABC,DE)=[d(A,D)+d(B,D)+d(C,D)+d(A,E)+d(B,E)+d(C,E))]/6
= [9.5+12.25]/2 = 10.875

	ABC
DE	10.7

Step I is for 1 mark, all subsequent steps carry 2 marks each.
Correct tree with incorrect branch lengths will be 7 marks.

Question 9. The provided image shows the categorization of different amino acids on the basis on their physicochemical properties. (3 marks)

Using the given information, state whether the following statements are CORRECT or INCORRECT.

(i) Substitution of Pro with Ala in a protein molecule will not make a significant structural difference as both are hydrophobic in nature. Incorrect.

(ii) Log odd ratio of Glu→Asp is expected be higher in comparison to Glu→Arg. Correct

(iii) Arg, Lys and His are positively charged amino acids. Correct