

Indraprastha Institute of Information Technology Delhi (IIITD)
Department of Computational Biotechnology

BIO213 – Introduction to Quantitative Biology

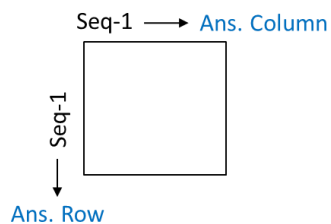
MID-SEM EXAM - March 4, 2021

Duration: 1 hour

Total marks: 50

Question 1. In dynamic programming, for not charging the gaps in the beginning of the first sequence, initialize first _____ with zeroes. (Row/Column) **(Marks 1)**

Answer:



Question 2. Get the substitution value for P → Q,

PQPP
QQQP
QQPP
QPPP
QQQP

	P	Q
P	13	14
Q		13

$$P(P) = \frac{10}{20} = 0.5, P(Q) = \frac{10}{20} = 0.5$$

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

$$q_{PQ} = \frac{14}{40} = 0.35$$

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

(Marks 10)

(1 mark per blank)

Question 3. Choose the correct options:

1. PAM matrices are based on _____ alignments of closely related proteins.
2. BLOSUM matrices are based on _____ alignments of related proteins.
3. PAM1 is calculated from comparisons of sequences with no more than 1% _____.
4. BLOSUM 62 is a matrix calculated from comparison of sequences with 62% _____.

Answer: 1) global, 2) local, 3) divergence, 4) identity

(Marks 4)

(1 mark per blank)

Question 4. List one advantage of MULTICONTIG over RECONSTRUCTION model of fragment assembly. (Marks 1)

Answer: MULTICONTIG allows reconstruction of multiple contigs, which is not possible in RECONSTRUCTION. RECONSTRUCTION is designed to return the shortest common superstring.

Question 5. Find all optimal global alignments between AAAG and ACG using the scoring scheme – Match = +1, Mismatch = -1 and Gap = -2. (Marks 5)

Answer:

		A	C	G
		0	-2	-4
A	-2	1	-1	-3
A	-4	-1	0	-2
A	-6	-3	-2	-1
G	-8	-5	-4	-1

1. -ACG
AAAG

2. AC-G
AAAG

3. A-CG
AAAG

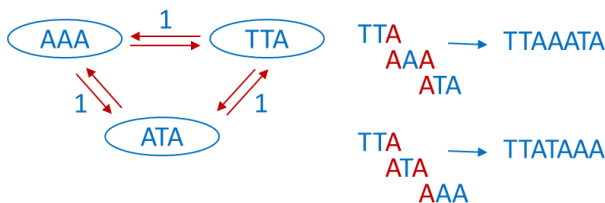
(Correct Matrix: 2 marks, 1 mark per alignment)

Question 6. State one of the major limitations of progressive multiple sequence alignment approach. (Marks 2)

Answer: The order in which sequences have been aligned matters. Once a wrong order is chosen, it cannot be corrected later.

Question 7. Construct an overlap graph for F = (AAA, TTA, ATA). Find a shortest common superstring for this collection. (Marks 5)

Answer:



(Correct graph: 2marks, 2 marks per superstring)

OR

Question 7. Given the following results of double digest experiments, try to find a solution,

Enzyme A: 12, 21

Enzyme B: 7, 4, 10, 12

Enzyme A+B: 8, 4, 12, 2, 7

(Marks 6)

Answer: A+B = 8 4 12 2 7
 a b c d e

A B
12 = a + b 7 = e
21 = c + d + e 4 = b
 10 = a + d
 12 = c

12 = a + b		21 = c + d + e		
8 = a	4 = b	12 = c	2 = d	7 = e

(You can use for own judgement here. There are many ways to solve this)

Question 8. What are the major factors that add to the complexity of the problem of fragment assembly?
(Marks 2)

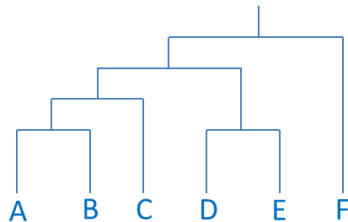
Answer: 1. Errors

- a. Base call errors
- b. Chimera
- c. Contamination
2. Unknown orientation
3. Repeated regions
4. Lack of coverage

(Even if a,b,c are missing give full marks)

Question 9. Draw the phylogenetic tree that corresponds to the one described in the standard Newick format as (((A,B)C)(D,E))F.
(Marks 5)

Answer:



(Award marks only if completely correct)

Question 10. Use UPGMA to reconstruct a phylogenetic tree using the following distance matrix:

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

(Marks 10)

Answer:

	A	B	C	D
A				
B	6			
C	9	9		
D	12	11	13	

6 marks

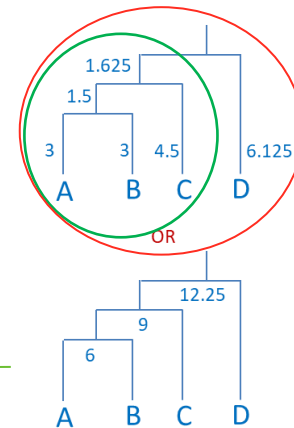
	AB	C	D
AB			
C	9		
D	11.5	13	

$$d(C,AB) = (d(C,A) + d(C,B))/2 \\ = (9+9)/2 = 9$$

$$d(D,AB) = (d(D,A) + d(D,B))/2 \\ = (12+11)/2 = 11.5$$

	ABC	D
ABC		
D	12.25	

$$d(D,ABC) = (d(D,AB) + d(D,C))/2 \\ = (11.5+13)/2 = 12.25$$



Branch lengths: 2 marks

So, correct matrices and a correct tree without branch lengths will be 8 marks

8 marks

OR

	A	B	C	D
A				
B	6			
C	9	9		
D	12	11	13	

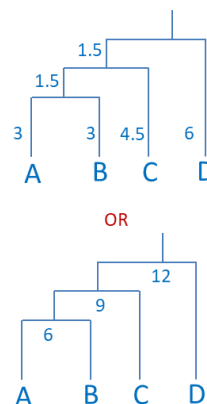
	AB	C	D
AB			
C	9		
D	11.5	13	

$$d(C,AB) = (d(C,A) + d(C,B))/2 \\ = (9+9)/2 = 9$$

$$d(D,AB) = (d(D,A) + d(D,B))/2 \\ = (12+11)/2 = 11.5$$

	ABC	D
ABC		
D	12	

$$d(D,ABC) = (d(D,A) + d(D,B) + d(A,C))/3 \\ = (12+11+13)/3 = 12$$



Question 11. What sites in the following alignment would be informative for a parsimony analysis? How many sites are invariant?

1	2	3	4	5
G	A	A	T	C
G	A	A	T	C
G	T	A	T	C
G	T	A	C	T
G	A	C	C	T
G	A	T	T	C

(Marks 4)

Answers: Invariant site: 1 (*1 mark*)
Informative site: 2, 4, 5 (*3 marks*)