

**Indraprastha Institute of Information Technology Delhi (IIITD)**

**Department of Computational Biotechnology**

**BIO213 – Introduction to Quantitative Biology**

**Quiz-2 (April 05, 2023)**

Total time: 40 mins

Total marks: 25

**Question 1.** In Prosite, which of the following patterns is wrongly matched with its significance?

- a. > – C-terminus (1 mark)
- b. { } – Any but not
- c. ( ) – Repeats
- d. Z – Any

**Question 2.** What are the four major factors that add to the complexity of fragment assembly problem? (2 marks)

Base call errors, chimeric fragments, contamination with vector, Unknown orientation, Repeated regions, Lack of coverage (any four – 0.5 mark each)

**Question 3.** Which of the fragments is wrongly placed in the following alignment layout for finding the target sequence, provided  $\epsilon = 0.3$ ? (4 mark)

```
TATAGCATCAT
  CGTC    CATGATCA
  ACGGATAG  GTCCA
-----
ACGTATAGCATGATCA
```

$$d_s(\text{TATAGCATCAT}, \text{ACGTATAGCATGATCA}) = 1$$

$$\epsilon |f| = 0.3 \times 11 = 3.3$$

Here,  $d_s(f, S) \leq |f|$ , therefore correct placement

$$d_s(\text{CGTC}, \text{ACGTATAGCATGATCA}) = 1$$

$$\epsilon |f| = 0.3 \times 4 = 1.2$$

Here,  $d_s(f, S) \leq |f|$ , therefore correct placement

$$d_s(\text{ACGGATAG}, \text{ACGTATAGCATGATCA}) = 1$$

$$\epsilon |f| = 0.3 \times 8 = 2.4$$

Here,  $d_s(f, S) \leq |f|$ , therefore correct placement

$$d_s(\text{GTCCA}, \text{ACGTATAGCATGATCA}) = 2$$

$$\epsilon |f| = 0.3 \times 5 = 1.5$$

Here,  $d_s(f, S) > |f|$ , therefore incorrect placement

(1 mark for each)

**Question 4.** The overall height of a logo position reflects how conserved the position is, and the \_\_\_\_\_ of each letter in a position reflects the \_\_\_\_\_ of the residue in the alignment. (1 mark)

- a. height, relative frequency
- b. width, relative frequency
- c. height, amplitude
- d. width, amplitude

**Question 5.** What does this representation mean - <A-x-[KM]-x(0,1)-V-{ED}? (2 mark)

- a. C terminal- Ala-any-[Lys or Met]-(any or none)-Val-{any except Glu or Asp}
- b. N terminal-Ala-any-[Lys or Met]-(any or none)-Val-{any except Glu or Asp}
- c. C terminal- Ala-any-[Lys and Met]-(any or none)-Val-{Glu or Asp}
- d. N terminal-Ala-any-[Lys and Met]-(any or none)-Val-{Glu or Asp}

**Question 6.** Map the restriction sites for enzyme A and B given the following results of a double digest experiment. (4 marks)

Enzyme A: 1400, 1000, 2100, 500

Enzyme B: 1300, 2500, 1200

Enzyme A+B: 500, 800, 1900, 200, 1000, 600

A+B: 500    800    1900    200    1000    600  
           a        b        c        d        e        f

A: 1400= b+f, 1000= e, 2100= c+d, 500=a

B: 1300= a+b, 1200= d+e, 2500=c+f

1000 = e	2100 = d+c	1400 = f+b	500 = a
1200 = e+d	2500 = c+f	1300 = b+a	

**Question 7.** Which of the following regular expressions would be matched by sequence MAVDEKC? (1 mark)

- a. {MG}-A-x(2)-[DE]-[ILK]-x
- b. M-A-x-[DE]-[ILK]-x
- c. {HS}-A-x(2)-[DE]-[ILK]-x
- d. M-A-x-[DE]-{LK}-C

**Question 8.** An invariant column in a multiple sequence alignment will have \_\_\_\_\_ entropy. (1 mark)

- a. Less
- b. High

**Question 9.** The maximum height of the entire stack of residues in case of DNA sequence logo is 2 bits. (or log<sub>2</sub>4) (2 mark)

**Question 10.** Calculate the contribution of each residue for making the sequence logo using the column information as given below: **(4 mark)**

G  
A  
A  
G  
G  
A  
A  
C  
A  
A

Give the value for: Entropy, Information index and contribution.

Entropy:  $-(0.3 \cdot \log_2 0.3 + 0.6 \cdot \log_2 0.6 + 0.1 \cdot \log_2 0.1) = 1.298$  (1.5 marks)

Information:  $\log_2 4 - \text{Entropy} = 2 - 1.298 = 0.702$  (1 mark)

Contribution: i)  $G = 0.3 \cdot \text{Information} = 0.3 \cdot 0.702 = 0.2106$  (0.5 mark)

ii)  $A = 0.6 \cdot 0.702 = 0.4212$  (0.5 mark)

iii)  $C = 0.1 \cdot 0.702 = 0.0702$  (0.5 mark)

**Question 10.** Viterbi algorithm is used to find the most likely sequence of states in an HMM model. **(1 mark)**

**Question 11.** Even if many pseudocounts are added in comparison to real sequence counts, the amino acid frequencies will not have any effect or influence. **(1 mark)**

- a. True
- b. False

**Question 12.** What is iterated in PSI-BLAST? **(1 mark)**

- a. PSSM
- b. Z-score
- c. E-value
- d. BLOSUM62