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

## BIO213 – Introduction to Quantitative Biology

## Quiz-1 (February 22, 2023)

**Total marks: 25**

1. Which of the following is an example of transition and transversion respectively?     (2 mark)

    a. A → C and A → T
    b. C → T and G → A
    c. G → C and C → T
    d. A → G and G → T

2. Vertical and horizontal frame shifts in a dot plot show the presence of     (1 mark)

    a. Palindromic sequences
    b. Microsatellites
    c. Insertion and deletions
    d. Repeating elements

3. Which sites in the following alignment would be informative for a parsimony analysis? How many sites are invariant?     (2 marks)

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

    Informative sites: 1, 4 and 5 (1.5 mark)
    Only one invariant site: 2 (0.5 mark)

4. Which of the following is the best suited for finding DNA binding domains in a set of protein sequences?     (1 mark)

    a. Global alignment
    b. Local alignment
    c. End-free alignment

5. Fill in the correct options:     (0.5 mark for each correct answer)

    1. PAM matrices are based on global alignments of closely related proteins.
    2. BLOSUM matrices are based on local alignments of related proteins.
    3. PAM1 is calculated from comparisons of sequences with no more than 1% divergence.
    4. BLOSUM 62 is a matrix calculated from comparison of sequences with 62% identity.
        *(Options: global, identity, divergence, local)*

6. Which of the following is not true for End-free alignment? (1 mark)

    a. The first row and column are initialized to zero for allowing indels/gaps at the ends without penalty
    b. The minimum value in a cell can be zero
    c. It is essentially used for building genome fragments out of shorter sequencing data
    d. Traceback starts with the best score in an end column or an end row.

7. State one major limitations of progressive multiple sequence alignment approach. (1 marks)

    The order of alignment dictates the final result

8. Consider the following multiple sequence alignment layout (2 marks)

    a.  KK        b.  KH
        KH            KR
        RK            KH
        HR            KR
        NK            KR
        KA            HK

    Which of the above (a or b) is more likely to have a higher log-odd ratio for the event K→R?

    a will have a higher log-odd ratio for the event K→R

9. Computational complexity of aligning $k$ sequences of length $n$ by dynamic programming and progressive alignment is given by (2 marks)

    a. $O(k^2n^2)$ and $O(2^k n^k)$, respectively.
    b. $O(2^k n^k)$ and $O(k^2n^2)$, respectively.
    c. $O(n^2)$ and $O(7n^k)$, respectively.
    d. $O(7n^k)$ and $O(n^2)$, respectively.

10. blastx is used for (1 mark)

    a. comparing amino acid sequence against the protein sequence database.
    b. comparing DNA sequence against the translated nucleotide sequence database.
    c. comparing amino acid sequence against the translated nucleotide sequence database.
    d. comparing DNA sequence against the protein sequence database.

11. Show all the possible sequence alignments corresponding to the paths shown with black arrows in the image below? (4 marks)

| | - | A | G | A | A | C | A | A | G | G | C | G | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **-** | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 | -8 | -9 | -10 | -11 | -12 |
| **A** | -1 | 1 | 0 | -1 | -2 | -3 | -4 | -5 | -6 | -7 | -8 | -9 | -10 |
| **C** | -2 | 0 | 0 | -1 | -2 | -1 | -2 | -3 | -4 | -5 | -6 | -7 | -8 |
| **A** | -3 | -1 | -1 | 1 | 0 | -1 | 0 | -1 | -2 | -3 | -4 | -5 | -6 |
| **A** | -4 | -2 | -2 | 0 | 2 | 1 | 0 | 1 | 0 | -1 | -2 | -3 | -4 |
| **G** | -5 | -3 | -1 | -1 | 1 | 1 | 0 | 0 | 2 | 1 | 0 | -1 | -2 |
| **A** | -6 | -4 | -2 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | -1 | 0 |
| **C** | -7 | -5 | -3 | -1 | -1 | 1 | 1 | 1 | 0 | 0 | 2 | 1 | 0 |
| **A** | -8 | -6 | -4 | -2 | 0 | 0 | 2 | 2 | 1 | 0 | 1 | 1 | 2 |
| **G** | -9 | -7 | -5 | -3 | -1 | -1 | 1 | 1 | 3 | 2 | 1 | 2 | 1 |
| **C** | -10 | -8 | -6 | -4 | -2 | 0 | 0 | 0 | 2 | 2 | 3 | 2 | 1 |
| **G** | -11 | -9 | -7 | -5 | -3 | -1 | -1 | -1 | 1 | 3 | 2 | 4 | 3 |
| **T** | -12 | -10 | -8 | -6 | -4 | -2 | 0 | 0 | 0 | 2 | 2 | 3 | 5 |

```
AGAACA-AGGCGT                  AGAACA-AGGCGT                 (2 marks each)
| || | | ||||    and    | || | || |||
ACAAGACA-GCGT                  ACAAGACAG-CGT
```

12. Find the score of the following alignment (Match = 4, Mismatch = -1, Gap opening penalty = -2, gap extension penalty = -1) (2 marks)

```
A T C G A T C G C G C A G
| |             | | | |
A T - C - - - A C G C A C
```

(4*6) + (-1*3) + (-2*2) + (-1*2) = 15

13. The optimal alignment of two similar sequences is usually that _____ number of matches and _____ the number of gaps. (2 mark)
   a. Minimize, maximize
   b. Maximize, minimize
   c. Degrade, upgrade
   d. Upgrade, degrade

14. Maximum likelihood method of phylogenetic tree construction outperforms the other alternative methods while working with shorter sequences. (1 mark)

15. Choose the correct option: (1 mark)

   a. Off-springs in Genetic algorithm are generated using mutations and crossover.
   b. Off-springs in Genetic algorithm are generated using Natural selection and crossover.
   c. Off-springs in Genetic algorithm are generated using Natural selection and mutations.