Ain Shams University
Faculty of Science
Department of Mathematics
Computer Science Division
Date: 21/May/2023
Total Marks: 105

Level: 4th Year
Course Code: Comp 402
Course Name: Bioinformatics
Programs: CS (Special & Double)
Time Allowed: 3 hours
No. of Page Sides: 2

Question 1 [21 pts]

1. [8 pts] What is meant by: the Hamming distance between two strings, the profile matrix, motif consensus and scoring function of the motif?

2. [7 pts] Given the following sequences: CTACG ($n = 5$), GCACT ($n = 5$), ACTACG ($n = 6$) and CACTAG ($n = 6$). Use the motif profile to calculate the consensus string (of length $k = 5$) of these sequences and its score.

3. [6 pts] Given a long text string DNA, a short pattern string p, and an integer k. Write an algorithm to find the first occurrence of a string p' in DNA such that $d_H(p, p') \leq k$. What is the complexity of your algorithm?

Question 2 [21 pts]

1. [6 pts] Mention three needs for Bioinformatics?

2. [8 pts] Write the mathematical formulation for the "Change Problem" of converting an amount of money into given denominations. Write the recursive relation to solve this problem.

3. [7 pts] Write a dynamic-programming algorithm for solving the above problem. Discuss its running time and the memory needed by this algorithm. Modify this algorithm to return not only the smallest number of coins but also the correct combination of coins.

Question 3 [21 pts]

1. [6 pts] Describe shortly, the flow of genetic information according to the "Central Dogma of Molecular Biology".

2. [5 pts] Put (✓) or (✗) for each statement. Give reasons or examples for the correct statements or correct the false ones:

   (a) The gene is the organism's complete set of DNA.   ( )

   (b) DNA is a complex organic molecule made up of amino acid subunits.   ( )

   (c) Almost every cell in an organism contains the same genome.   ( )

   (d) One of the complications of the Motif finding problem is the very long length of the motif sequence.   ( )

   (e) To get the edit distance between two sequences; both sequences must be equal in length.   ( )

3. [5 pts] Devise an algorithm to count the number of breakpoints in a signed permutation.

4. [5 pts] Give an example to show that the breakpoint reversal sorting algorithm can enter an infinite loop.

See the other page.

Question 4 [21 pts]

Given two sequences; the source sequence $v = GTGACTT$ and the target sequence $w = GAATT$, and assume that the insertion penalty is (-1) deletion penalty is (-2) and the score of the match = 1 and mismatch is given by the following scoring matrix:

	A	C	G	T
A	2	-1	-2	-1
C		2	-1	
G			3	-2
T				2

1. [5 pts] Write one matrix that has both the alignment score and the backtracking to find the global alignment between v and w (Consider elements of v as rows and elements of w as columns).
2. [1 pts] What is the score of the optimal global alignment between v and w.
3. [1 pts] How many optimal global alignments between v and w.
4. [3 pts] List all optimal global alignments between v and w.
5. [5 pts] Add a constraint on the alignment so that indels are not allowed at the beginning of the alignment. What are those changes (either in the recurrence relation or in the scoring system) required to obtain optimal alignments that follow such constraint?
6. [4 pts] Apply the modified algorithm and construct the matrix that shows the score and the backtracking of the optimal alignments between v and w that follow such constraint.
7. [2 pts] List all constrained optimal alignments between v and w.

Question 5 [21 pts]

Assume that the permutation $\pi = +1 -3 -4 -2 +6 +7 -5 +8$ and the identity permutation $\theta = +1 +2 +3 +4 +5 +6 +7 +8$.

1. [1 pts] What is the lower bound distance between π and θ?
2. [3 pts] Apply the greedy algorithm to transform π into θ. Report all intermediate permutations and report the total number of reversal operations that were performed.
3. [3 pts] Apply the breakpoint reversal sorting algorithm to transform π into θ. At each step choose the reversal that reduces the number of breakpoints. Report all intermediate permutations, number of breakpoints in each permutation, and the total number of reversal operations that were performed at the end.
4. [4 pts] Construct the breakpoint graph from π and θ; using red (or gray) edges for π and blue edges for θ. Report the number of red-blue alternating cycles in the graph.
5. [10 pts] Apply the 2-break sorting algorithm to transform the graph into the trivial breakpoint graph. Draw all intermediate breakpoints graphs. Under each graph report (1) the number of cycles and (2) the permutation corresponding to the one with red edges.

Good luck!!!