

Bioinformatics Module
Sample Examination paper

David Gilbert

Answer 3 questions.

Appendices A,B,C are attached, and are for use in the following questions:

Appendix	Questions
A	1, 2
B	1
C	4

1

1.1 Sketch the composition of a nucleotide, and give the alphabets for sequences of DNA and RNA. [10]

1.2 Compute the complementary strand for the following sequence [10]

5' : atggtgcacctgactcctgaggagaagtctgcggttactgcacctgtggtag :3'

1.3 What is meant by the terms *transcription* and *translation* and *frame shift errors* in the context of protein synthesis. Illustrate your answer by considering the process of synthesising an amino-acid chain from the DNA sequence above, and referring to Appendices A and C. [20]

1.4 Define the recurrence relation for the *Edit distance* between two strings and give the order of complexity for a naïve algorithm based on this relation. [10]

1.5 Compute the dynamic programming table, alignments and associated sequence identities for the two strings WATER and WINE, where Symbol mis-match -5; gap insertion -1; match 5 [50]

2.1 Define: language, string, symbol, alphabet [10]

2.2 Explain the meaning of the symbols in the PROSITE pattern, referring to

Appendix A

[AC] -x-V-x (4) - {ED}

and show how it matches the sequence [30]

DEHSDVLPVLDVCSLKHVAYVFQALIYWIKAMNQQTTLDT

2.3 Distinguish between *pattern-driven* and *sequence-driven* pattern discovery. [20]

2.4 Give an algorithm for a *sequence-driven* discovery algorithm for patterns over amino-acid sequences, where a pattern is defined as one or more (sub)sequences which match an example and illustrate the working of your algorithm on the following sequences: [40]

s_1 = MERIAVLALEDKYGK

s_2 = MLEDKYGYLEDERIAVLAKV

$s_3 = \text{VKYGYKDEMERIAVLAL}$

3.

3.1 Define, and distinguish between: phylum, taxon and species [20]

3.2 Give an algorithm for Neighbour-Joining method, and show how it can be applied to the data in Table 1 to produce a tree. [40]

Table 1

	1	2	3	4
1	0	0.3	0.5	0.6
2		0	0.6	0.5
3			0	0.9
4				0

3.3 [40]

(i) What assumption is made by the Neighbour-Joining algorithm in terms of a *molecular clock*, and what test can be applied to tree data to determine whether Neighbour-joining tree reconstruction is likely to be correct?

(ii) Show how a rooted tree might be derived by Neighbour-Joining for the data in Table 1 and indicate what additional information would be required to achieve this

(iii) Name another tree reconstruction algorithm which can be used to reconstruct a tree, and give any assumptions that must be made about the data for correct tree reconstruction.

4

4.1. Define *directed graph*, *elementary path*, *elementary circuit* [15]

4.2 Show how a graph can be represented as an adjacency matrix. [15]

4.3 Give an algorithm for breadth-first search in a graph and indicate the advantages and disadvantages of this method compared to depth-first search. [30]

4.4 Apply your algorithm to perform a breadth-first search of the signalling pathway shown in Appendix D, indicating your result by annotating the nodes visited in order. *You should interpret a two-way arrow as pointing forwards from the current node that is being processed in the algorithm.* [40]