



RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B.Sc. Four Year Degree in Applied Sciences
Fourth Year - Semester I Examination – June/July 2018

COM 4202 – BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

Time: Two (02) hours

Answer ALL questions

The use of a non-programmable electronic calculator is permitted.

1.
 - a) Explain how to create BWT index. [15 marks]
 - b) Provide algorithm to get original sequence from BWT index. [20 marks]
 - c) Explain the meaning of the LF function. [10 marks]
 - d) how would you make LF function run in constant time? [20 marks]
 - e) Mention importance of making LF function run in constant time for aligning short DNA sequence with Genome with huge number of base pairs. [15 marks]
 - f) Provide algorithm to aligning short DNA sequence with Genome based on BWT. [20 marks]

2.
 - a) compare and contrast De Bruijn graph assembly over Overlap-Layout-Consensus Genome assemble technique. [15 marks]
 - b) Discuss following properties of node of graph with suitable examples.
 - i. balanced
 - ii. semi-balanced[10 marks]
 - c) What are the conditions needed satisfy by a directed connected graph to be Eulerian? [10 marks]
 - d) Consider following DNA sequences as segments of one DNA string:

ACAATTGT ACATGCCG TTGTACAT

- i. Represent above sequences with De Bruijn Graph of 4-mers nodes (Edge represent 5-mer).
- ii. Apply De Bruijn Graph Assembly method to get the original DNA string.

[60 marks]

3.

a) Briefly explain the three problems associated with Hidden Markov Model (HMM) with a suitable example.

[30 marks]

b) Explain how you would apply HMM for a selected DNA or DNA sequence related practical problem.

[30 marks]

c) Amino acid sequence is given. You need to predict secondary structure (alpha helices/ beta sheet/random coil) of the amino acid sequence using given pre-trained HMM model. Write a computer program or pseudo code to predict secondary structure of amino acid sequence. Clearly mention meaning of every variable you used.

[40 marks]

4.

a) Describe what pairwise sequence alignment is.

[20 marks]

b) Compare and contrast between the local and global alignments.

[40 marks]

c) In a homology search why is it necessary to filter the low complexity regions?

[20 marks]

d) Why is it important to have a heuristic algorithm in similarity search

[40 marks]

—END—