RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B.Sc. (Special) Degree in Applied Biology
Third Year – Semester II Examination – February/March 2019

BIO 3204 – BIOINFORMATICS

Time: Two (02) hours

Answer <u>ALL</u> questions.

---

1. In 1990 an HIV-positive dentist in Florida, USA was suspected of having transmitted the HIV virus to his dental patients. Although a number of his patients were HIV-positive, it was unclear whether they had been infected by visiting the dentist. The Centers for Disease Control (CDC) in the USA sequenced the gp120 gene from viruses in the dentist, his HIV-positive patients, and a number of HIV-positive people (local reference) from the same region in Florida. RNA sequence data was analysed using Parsimony and likelihood methods. The resulting highest scoring maximum likelihood tree is given below. Answer the following questions based on the given phylogeny.
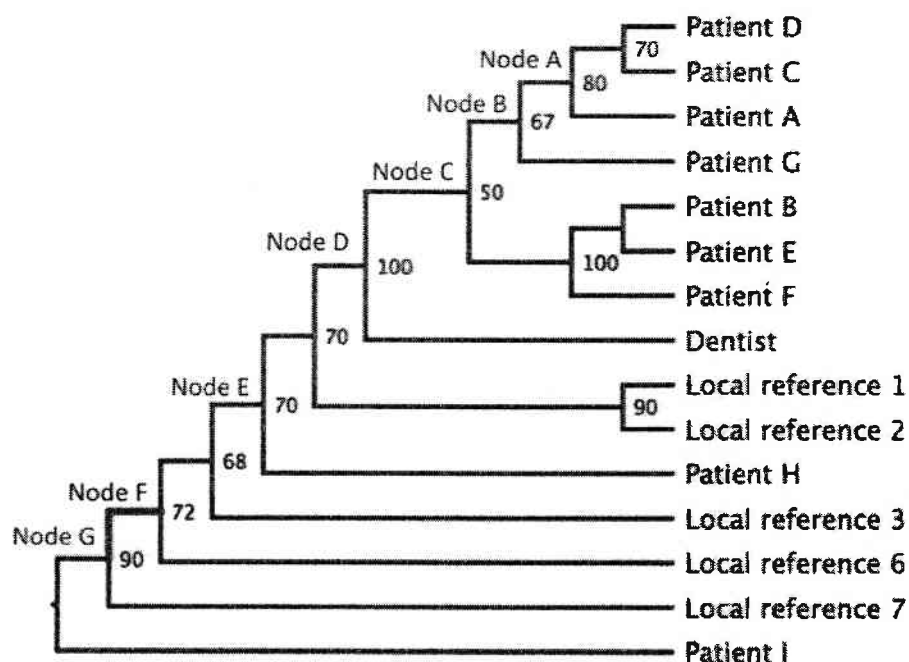
    a)   Using your knowledge in molecular phylogenetics and molecular biology, outline the most likely steps the scientists at the CDC must have taken to produce this molecular phylogeny.   **(20 marks)**

b) Based on the molecular phylogeny, is the dentist guilty of transmitting HIV to his patients? Justify your answer. **(20 marks)**

c) If you found the dentist guilty of transmitting the disease to his patients, which patients were most likely infected by him? Justify your answer. **(15 marks)**

d) What is meant by the numerical values shown at the nodes? **(15 marks)**

e) Name four nodes that you can have high confidence in. **(15 marks)**

f) State what is meant by the term 'highest scoring maximum likelihood tree' and write two disadvantages of the likelihood method. **(15 marks)**

(100 marks)

2. Explain the following.

   a) In homology searches, it is ideal to mask certain sequences.

   b) Algorithm used in a similarity search is a heuristic.

   c) Protein is better for similarity search than DNA.

   d) Naïve scoring methods are not ideal for a pairwise sequence alignment.

   e) If the information available about a query sequence is limited, the option to select during a similarity search is "somewhat similar sequences".

   (20 x 5 marks)

3. A group of scientists hypothesised that the genetic disorder "A" is caused by a specific mutation in gene x by which the coded protein becomes non-functional. If functional protein "X" converts amino acid "B" into products "X" and "Y", write concept paper to prove or disprove their hypothesis. **(100 marks)**

4. Write short notes on the following.

   a) Steps in comparative structure modelling

   b) Amino acid substitution matrices

   c) tRNA modelling

   d) Use of RMSD in assessing transformations in comparative modelling

   (25 x 4 marks)

---END---