COMPUTER SCIENCE TRIPOS Part II – 2025 – Paper 8

2 Bioinformatics (pl219)

(*a*) In genome alignment we need to detect *inexact* matching with respect to the reference genome in order to find variants. This allows characterising individual differences and disease-causing mutations. Briefly explain, with one example, what the Burrows–Wheeler Transform (BWT) is. State its advantage over naive search algorithms when finding exact genome alignments. Explain how it may be extended to detect variants. [8 marks]

(*b*) The polymerase chain reaction is used to amplify RNA sequences. There are short sequences, primers, that start the alignment. These should not form extensive pairings (for example, in stem-loop structures).

Discuss with one example (4 or 5 bases) how Ruth Nussinov's RNA algorithm can indicate whether an RNA sequence may be suitable. [7 marks]

(*c*) Experimental evidence shows that two proteins A and B bind each other.

(*i*) If experiments show that the two proteins interact in many species, should we expect that they have similar phylogenetic trees when you use the A and B sequences from these species? [3 marks]

(*ii*) Discuss the role of a scoring matrix in optimal tree phylogeny. [2 marks]