

COMPUTER SCIENCE TRIPOS Part II – 2015 – Paper 7

3 Bioinformatics (PL)

- (a) Compute the global alignment and the best score of the following sequences CGTGAA, GACTTAC with the following parameters: match score = +5, mismatch score = -3, gap penalty = -4. [4 marks]
- (b) If the sequences have different base composition or length, what parameter values would you choose in order to determine multiple alignment of the sequences? Justify your answer. [4 marks]
- (c) Discuss the main features of the Burrows-Wheeler transform (BWT) using the following example: T = GATTACA\$. Also explain the reversibility of BWT. [4 marks]
- (d) Discuss the use of clustering in the analysis of gene expression microarray data. [4 marks]
- (e) Discuss the UPGMA algorithm as applied to the following distance matrix of the species A, B, C and D .

$$\begin{array}{l} \textit{species} \\ B \\ C \\ D \end{array} \begin{array}{ccc} A & B & C \\ \left(\begin{array}{ccc} 2 & & \\ 4 & 4 & \\ 6 & 6 & 6 \end{array} \right) \end{array}$$

[4 marks]