

COMPUTER SCIENCE TRIPOS Part II – 2014 – Paper 9

1 Bioinformatics (PL)

- (a) How can you be reasonably confident that an amino acid multiple sequence alignment is optimal? [6 marks]
- (b) From gene expression analysis we hypothesize that each DNA sequence of a dataset from a different animal species contains at least one similar short subsequence at an unknown position in the dataset. Explain the procedure to identify the short subsequence in the dataset and then cluster the species according to the similarity of the short subsequence. [8 marks]
- (c) Which algorithms might be useful in aligning very long DNA sequences, such as entire genomes and why? [6 marks]