2 Bioinformatics (PL)

(*a*) Discuss how to efficiently cluster a set of gene expression data. [5 marks]

(*b*) Explain with one example how you would detect CG islands in a genome. [5 marks]

(*c*) Discuss the use of the Burrows-Wheeler transform in genome assembly and its algorithmic complexity. [6 marks]

(*d*) Compare the advantages and disadvantages of having long versus short k-mers in genome assembly. [4 marks]