

2 Bioinformatics (pl219)

(a) Compute the local alignment between the following sequences: GATTACA, TATACG with the following rules: match score = +5, mismatch = -3, gap penalty = -4 and discuss how the alignment depends on the choices of match scores, mismatch and gap penalty. [5 marks]

(b) Discuss how a local alignment algorithm allows identification of internal sequence duplications. [3 marks]

(c) Define the UPGMA algorithm and state and justify its complexity. What is the output of the algorithm given the distance matrix of the species X_1, X_2, X_3, X_4 below?

$$\begin{pmatrix} \text{species} & X_1 & X_2 & X_3 \\ X_2 & 2 & & \\ X_3 & 4 & 4 & \\ X_4 & 6 & 6 & 6 \end{pmatrix}$$

[4 marks]

(d) Discuss a method to perform random access in DNA-based storage memory. [4 marks]

(e) Discuss with one example the complexity of the Gillespie algorithm and comment on the main differences with respect to a deterministic approach. [4 marks]