

Bioinformatics Supervision

–Problem Sheet 1–

Supervisor: Sebastian Müller (Department of Plant Sciences)

Please hand in your work 24 hours prior to the supervision either to sm934@cam.ac.uk or at the Plant Sciences Department reception (make sure my name is on it). Feel free to team up with other group members, the main aim is to understand the material. If you hand in electronically, please name the file `group<x>_<crsid>_problemsheet<x>.<x>`

1. Why do we use dynamic programming algorithms for sequence alignment problems?
2. What is the difference between local and global alignment? What are the two good for?
3. Define the Longest Common Subsequence (LCS) problem between two strings and find a solution for the case of the two strings: ACGT and GGTTTAAGCCGT
4. Describe what needs to be taken into account for gaps in DNA sequence alignment.
5. 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. Show the alignment graph including backtracking pointers or bring it to supervision.
6. If the sequences have different base composition (such as GC content) or length, what parameter values would you choose in order to determine multiple alignment of the sequences? Justify your answer.
7. Copy the entire text from a FASTA file (<http://www.cs.ukzn.ac.za/~hughm/bio/data/DinosaurCollagen.fasta>), containing the Collagen protein sequences from a number of different species. Then enter it into an online tool for multiple sequence alignment from the European Bioinformatics Institute (<http://www.ebi.ac.uk/Tools/msa/kalign/>). What does the Phylogenetic tree look like? Can you change the alignment options so that the tyrannosaurus collagen is no longer paired with the newt collagen?
8. Discuss the limitations of the Nussinov algorithm for RNA folding prediction.
9. Discuss the main features of the Burrows-Wheeler transform (BWT) using the following example: T = GATATCA\$. Also explain the reversibility of BWT