Exercises – Algorithms for Genome Research

Universität Bielefeld, WS 2014, Dr. Pedro Feijao

Exercise List 10 — 16.01.2015

Discussion of exercises on: 23.01.2015

Exercise 1

Describe typical goals of a DNA microarray experiment.

Exercise 2

Given the following set of probes:

`GTACCGA, TAGACTG, GAGATGA, CAGCTAA`

(a) Give a lower bound for the deposition sequence length.

(b) Using the pattern $(\text{CGTA})^k$, find a deposition sequence using the Alphabet Leftmost Algorithm (find a leftmost embedding for each input sequence, remove unproductive steps).

(c) Using the supersequence above, try to find if it is possible to delete a position and re-embed the input sequences, to find a shorter supersequence. (*Hint: try positions with low productivity.*)

(d) Suppose that I want to find a supersequence starting with the prefix `CAGCTAA`. Give a lower bound on this supersequence length. Can this sequence be optimal?

Exercise 3

Finding the Shortest Common Supersequence (SCS) is NP-hard for 3 or more input sequences, but can be solved polynomially for only 2. This is similar for the Longest Common Subsequence problem (LCS). Assuming that you have an algorithm for finding a LCS of 2 strings, how can you use the LCS solution to build a SCS for the same 2 strings?