Exercises – Algorithms for Genome Research

Universität Bielefeld, WS 2014, Dr. Pedro Feijao

http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014winter/AlgoGR

Exercise List 6 — 5.12.2014

Discussion of exercises on: 12.12.2014

Exercise 1

(a) What are mate pairs? How do they simplify the assembly problem?

(b) Discuss the main experimental problems that make sequence assembly difficult.

Exercise 2 In the overlap phase of an assembly, prefix-suffix "local alignments" are seeked.

(a) Work out the details of a dynamic programming algorithm.

(b) What are the time and space complexities of the seed-based algorithm mentioned in class?

Exercise 3 Construct the overlap graph for the following set of reads, assuming no sequencing errors, i.e. only exact prefix-suffix matches are allowed, and considering only overlaps of size three or more. (Note that the orientation of the reads is unknown).

```
1 TCCCA
2 GGTAAT
3 TCTTAGT
4 ACCGAG
5 CCAGT
6 GGATTG
7 AATCT
```

(a) Compute a layout. How many contigs do you get?

(b) Assume that the first two reads `TCCCA` and `GGTAAT` from above form a mate pair in opposite relative direction, originating from a clone with approximate length 25bp. What do you learn about the relative location of the contigs?