

Exercises – Algorithms for Genome Research

Universität Bielefeld, WS 2014, Dr. Pedro Feijao

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

Exercise List 7 — 12.12.2014

Discussion of exercises on: 19.12.2014

Exercise 1

Discuss the reasons why the traditional assemblers fail to assemble short-read data.

Exercise 2

The basic data structure used for short-read sequence assembly is the de-Bruijn graph. While it is conceptually easy, there are several challenges when you want to implement it in practice. Name a few of these challenges.

Exercise 3

Draw the 4-dimensional de-Bruijn graph (i.e. where vertices correspond to 4-grams) for the following set of reads. Can you assemble the data set into a single contig?

GTAAAT, AGACG, ACGTT, CACGG, ACTAG, TTAATG, TAATG, TGACC, GACCAGA, TAATG, AATGC, TGCAC, GCACG, ATGCA, GTAAATG, AAATG, TGCAC, GCACG, CACGG, TAATGA, AATGAC, CAGAC, AGACG, ACCAGA, ATAATG, TAATG, AATGA, GCACGG, ATAAT, CCAGA, ATGCA, ATAAT, ACCTT, ATGCAC, TGCAC, CGTTA, CGTTA, TTAATG, GACCA, ACCAG, CCAGA, CAGAC, ATGAC, GACGTT, ATGGA, ACGTT.