Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye
Dr. habil. Marília D. V. Braga

Algorithms in Genome Research

Winter 2025/2026

Exercises

Number 2, Discussion: 2025-November-14

1. Consider the following set of reads, assuming that you know already that all of them originate from the same DNA strand.

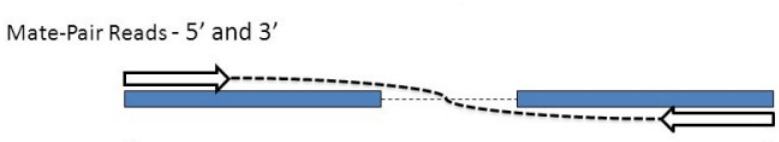
$$\begin{array}{ll} r_1 = \texttt{ATCCA} & r_6 = \texttt{GCAAG} \\ r_2 = \texttt{AGAGC} & r_7 = \texttt{AGATC} \\ r_3 = \texttt{AAGAT} & r_8 = \texttt{TAGAG} \\ r_4 = \texttt{GAGCA} & r_9 = \texttt{AGAGC} \\ r_5 = \texttt{CCATA} & r_{10} = \texttt{GAGCA} \end{array}$$

   (a) Build the corresponding overlap graph with a minimum overlap of 2.

   (b) Find a shortest common superstring for all given reads, with a corresponding layout and coverage. Find another common superstring (that could be slightly longer) whose layout gives a more uniform coverage.

2. Discuss the main experimental problems that make sequence assembly difficult in practice.

3. A mate pair is a pair of reads that originate from opposite ends of the same clone, with a good estimation of the distance between them:



   How can mate pairs help in the assembling process?

4. Let the following DNA sequence be a *reference genome*:

   `AATGAGGTCATCCTTGCTGGACTCTAGCAC`

   The following three sets of reads (a), (b) and (c) originate from three distinct *target genomes* that are closely related to the reference. Each target genome differs from the reference by a single structural variation (rearrangement or indel).

   Reconstruct the three target genomes by mapping the reads to the reference and identify the rearrangements. (Assume that there are no sequencing errors and recall that a read may come from any of the two complementary DNA strands.)

(a)

1	ACTCTAGCAC
2	AGTCCTGTACAG
3	CCTTGCTGTA
4	GCTGTACAGGAC
5	GGTCATCCTT
6	TGACCTCATT

(b)

1	AATGACAAGG
2	ACCCTGGACTCT
3	GGATGACCCTG
4	GTCATCCTTG
5	GTGCTAGAGT
6	TCCAGGGTCA

(c)

1	AATGAGGTCA
2	AGGTCATCGAC
3	AGTCGATGAC
4	CATCGACTCT
5	CTAGAGTCGAT
6	GTGCTAGAGT