Universität Bielefeld
Technische Fakultät

AG Genominformatik
Dr. Marília D. V. Braga
Dr. Daniel Dörr

Algorithms in Genome Research
Winter 2019/2020

Exercises

Number 3, Discussion: 2019 November 15

1. What are mate pairs and paired-end reads? What can be done with long reads that can not be done with paired-end reads?
2. What are the main differences between "traditional" (*de-novo*) genome assembly and comparative assembly?
3. What are the major steps in the comparative assembly strategy?
4. Let the following DNA sequence be a "reference genome":

   AATGAGGTCATCCTTGCTGGACTCTAGCAC

   The following three sets of "reads" (a), (b), (c) originate from three "target" genomes that are closely related to the reference.

   Consider the following conditions:

   - There are no sequencing errors.
   - Each target genome differs from the reference by a single structural variation (rearrangement).
   - A read may come from any of the two complementary DNA strands.

   Reconstruct the three target genomes by mapping the reads to the reference and identify the rearrangements.

   (a)

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

   (b)

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

(c)

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