

Algorithms in Genome Research
Winter 2018/2019

Exercises

Number 6 (b), Discussion: 2018 December 07

6. 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