

Algorithms in Genome Research

Winter 2013/2014

Exercises

Number 7, Discussion: 2013 December 20

1. Pacific Biosciences provide a new technology for sequencing : SMRT (Single Molecule Real Time). The size of the reads is in average 10 000 bp but the quality of the sequencing is low (85% accurate). What is the advantages and defaults of this technology? Can we use the techniques presented during the lectures to assembly these long but bad reads?
2. Compute the numbers of adjacencies, of 2-adjacencies, of breakpoints, of common intervals, of conserved intervals and the measures MAD and SAD for the following pair of genomes. What is the specificity of each of these measures?

0 1 2 3 4 5 6 7 8 9

0 -7 -6 3 1 2 5 4 8 9

3. Define the previous measures in the case of duplication.