

Algorithms in Comparative Genomics

Universität Bielefeld, SS 2023

Dr. Marília D. V. Braga · Dr. Roland Wittler

<https://gi.cebitec.uni-bielefeld.de/teaching/2023summer/cg>

Exercise sheet 1, 14.04.2023

Exercise 1 (Algorithm for breakpoint distance)

(5 pts)

Devise a linear time algorithm for computing the breakpoint distance $d_{BP}(\mathbb{A}, \mathbb{B})$, where \mathbb{A} and \mathbb{B} are a pair of canonical genomes and can contain multiple linear or circular chromosomes.

Exercise 2 (Set of doubled genomes)

(3 pts)

For each of the following singular genomes, give the corresponding set of possible doubled genomes

1. $\mathbb{A} = \{ [1\bar{2}3] [45] \}$
2. $\mathbb{B} = \{ [\bar{1}\bar{3}2] (45) \}$
3. $\mathbb{C} = \{ [\bar{2}31] (\bar{4}5) (\bar{6}) \}$

Exercise 3 (Breakpoint double distance)

(6 pts)

For the following pairs with singular genome \mathbb{S} and duplicated genome \mathbb{D} , compute the breakpoint double distance $d_{BP}^2(\mathbb{S}, \mathbb{D})$ and give the doubled genome $\mathbb{B} \in 2\mathbb{S}$ and an optimal matching of the genes of \mathbb{B} and \mathbb{D} .

1. $\mathbb{S} = \{ [1234567] \}$
 $\mathbb{D} = \{ (12323) [156\bar{5}\bar{4}7] (467) \}$
2. $\mathbb{S} = \{ (1\bar{2}) (\bar{4}3) \}$
 $\mathbb{D} = \{ (\bar{4}341\bar{2}3) (12) \}$

Exercise 4 (Maximal matching)

(4 pts)

Given two annotated genomes

$\mathbb{A} = (13184778\bar{7}59)$ and

$\mathbb{B} = (21\bar{5}7856789\bar{5}3)$,

1. Give the number of markers per family in each genome.
2. Give the number of distinct pairs of singular genomes that can be obtained by fixing a maximal matching of the genes within each family.