Algorithms in Comparative Genomics

Bielefeld University, Summer Semester 2025 Dr. Marília D. V. Braga · Prof. Dr. Jens Stoye https://gi.cebitec.uni-bielefeld.de/teaching/2025summer/cg Exercise sheet 7, discussion: 13.06.2025

Exercise 1 (Inversion distance of circular genomes)

Given the canonical circular chromosomes

$$\mathbb{A} = (1 \ 3 \ \overline{5} \ 4 \ 6 \ 8 \ 7 \ 9 \ 2) \text{ and}$$
$$\mathbb{B} = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9).$$

- 1. Draw the relational or the breakpoint diagram of \mathbb{A} and \mathbb{B} .
- 2. Identify all cycles on the diagram, giving their lengths and their types (good / bad).
- 3. Identify all interleaving components on the diagram, giving their numbers of cycles and their types (oriented / unoriented).
- 4. What is the inversion distance $d_{INV}(\mathbb{A}, \mathbb{B})$?

Exercise 2 (Sorting a circular genome by inversions)

Consider canonical circular chromosomes $\mathbb{A} = (1 \ \overline{4} \ \overline{2} \ 3)$ and $\mathbb{B} = (1 \ 2 \ 3 \ 4)$.

- 1. Draw the relational or the breakpoint diagram of \mathbb{A} and \mathbb{B} .
- 2. Compute the inversion distance $d_{INV}(\mathbb{A}, \mathbb{B})$.
- 3. Draw the overlap graph of the candidate adjacencies, annotating each vertex with its corresponding score (the score of a vertex is the number of good vertices in the overlap graph if the inversion corresponding to that vertex is performed).
- 4. What are the unsafe inversions that can be identified on the overlap graph?
- 5. Find the first safe inversion sorting \mathbb{A} into \mathbb{B} . Indicate your choice by marking the corresponding vertex in the overlap graph.
- 6. Find subsequent safe inversions to completely sort \mathbb{A} into \mathbb{B} .

Exercise 3 (Canonical DCJ model with capping)

Consider the following canonical genomes:

- 1. Construct the relational graph $RG(\mathbb{A}, \mathbb{B})$.
- For each component C of the relational graph RG(A, B), give the type of C (cycle, AB-path, AA-path or BB-path).
- 3. Compute the DCJ distance $d_{DCJ}(\mathbb{A}, \mathbb{B})$.
- 4. Modify $RG(\mathbb{A}, \mathbb{B})$ into an optimal capped relational graph $RG'(\mathbb{A}, \mathbb{B})$ with the minimum number of caps $p = \max{\chi(\mathbb{A}), \chi(\mathbb{B})}$, where $\chi(.)$ is the number of linear chromosomes in a genome. Derive a pair of canonical circular genomes from the $RG'(\mathbb{A}, \mathbb{B})$, maximizing the number of circular chromosomes in each genome.
- 5. Modify the relational graph into an optimal capped relational graph $RG'(\mathbb{A}, \mathbb{B})$ and derive a pair of canonical co-tailed linear genomes from $RG''(\mathbb{A}, \mathbb{B})$. (Hint: it might require more than p caps.)