

Algorithms in Comparative Genomics

Universität Bielefeld, WS 2021/2022

Dr. Marília D. V. Braga · Leonard Bohnenkämper

<https://gi.cebitec.uni-bielefeld.de/teaching/2021winter/cg>

Exercise sheet 12, 27.1.2022

Exercise 1 (ILP for computing the family-free DCJ distance)

(4 pts)

The capped weighted multi-relational graph for the family-free model contains indel-edges for all genes in both genomes. The ILP for computing the FF DCJ distance is basically identical to the one for the DCJ distance of balanced genomes proposed by Shao *et al.*, except that for FF DCJ we need the following two modifications:

1. Add a constraint to enforce the selection of maximal matchings only.
2. Adapt the objective function to take the size of the matching and the weights into consideration.

Show how modifications 1. and 2. can be done.

Exercise 2 (Review: SCJ/breakpoint median)

(6 pts)

Given three canonical genomes $C_1^f = [1\ 2\ 3\ 4\ 5]$, $C_2^f = [1\ 2\ \bar{3}\ \bar{5}\ 4]$ and $C_3^f = [2\ \bar{3}\ \bar{4}\ 5\ 1]$:

1. Compute a general SCJ median M_{SCJ}^f of C_1^f , C_2^f and C_3^f .
2. Is there another SCJ median of C_1^f , C_2^f and C_3^f that is distinct from M_{SCJ}^f ?
(Justify your answer by giving a distinct median or explaining why it does not exist.)
3. Is M_{SCJ}^f also a breakpoint median of C_1^f , C_2^f and C_3^f ?

If *no*: Compute a breakpoint median of C_1^f , C_2^f and C_3^f .

If *yes*: Is there another breakpoint median of C_1^f , C_2^f and C_3^f that is distinct from M_{SCJ}^f ?
(Justify your answer by giving a distinct median or explaining why it does not exist.)

Exercise 3 (Review: DCJ halving)

(4 pts)

For duplicated genome $\mathbb{D} = (3\ 5\ \bar{4}\ 2\ \bar{5})\ (2\ 4)\ (1\ \bar{1}\ \bar{3})$:

1. Draw the natural graph $NG(\mathbb{D})$.
2. Compute the DCJ halving distance $h = h_{DCJ}(\mathbb{D})$.
3. Give a halving scenario with h optimal DCJ operations that transform \mathbb{D} into a perfectly duplicated genome.

Exercise 4 (Review: Inversion distance with component tree)

(6 pts)

Given canonical circular chromosomes $\begin{cases} \mathbb{A} = (0\ \bar{3}\ 1\ 2\ 4\ 6\ 5\ 7\ \bar{15}\ \bar{13}\ \bar{14}\ \bar{12}\ \bar{10}\ \bar{11}\ \bar{9}\ 8) \\ \mathbb{B} = (0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ 15) \end{cases}$

1. Draw the relational (or the breakpoint) diagram of \mathbb{A} and \mathbb{B} .
(You can use the Java program `InversionVisualization` provided on the course website.)
2. Based on the diagram, construct both the chained component tree $\Upsilon_{\blacksquare}(\mathbb{A}, \mathbb{B})$ and the component tree $\Upsilon_{\circ}(\mathbb{A}, \mathbb{B})$.
3. Find an optimal cover (i.e. a cover with minimum cost) for $\Upsilon_{\circ}(\mathbb{A}, \mathbb{B})$.
4. Compute the inversion distance $d_{INV}(\mathbb{A}, \mathbb{B})$.