

# Algorithms in Comparative Genomics

Universität Bielefeld, WS 2020/2021

Dr. Marília D. V. Braga · Leonard Bohnenkämper

<https://gi.cebitec.uni-bielefeld.de/teaching/2020winter/cg>

## Exercise sheet 13, 4.2.2021

### Exercise 1 (Bounds for SCJ distance)

(4 pts)

Theoretical bounds for the SCJ distance with respect to the DCJ distance are

$$d_{\text{DCJ}}(\mathbb{G}_1, \mathbb{G}_2) \leq d_{\text{SCJ}}(\mathbb{G}_1, \mathbb{G}_2) \leq 4d_{\text{DCJ}}(\mathbb{G}_1, \mathbb{G}_2).$$

Give examples of pairs of mutually distinct genomes showing that these bounds are tight.

### Exercise 2 (SCJ/breakpoint median)

(6 pts)

Given three canonical genomes  $\mathbb{C}_1 = [1\ 2\ 3\ 4\ 5]$ ,  $\mathbb{C}_2 = [1\ 2\ \bar{3}\ \bar{5}\ 4]$  and  $\mathbb{C}_3 = [2\ \bar{3}\ \bar{4}\ 5\ 1]$ :

1. Compute a general SCJ median  $\mathbb{M}_{\text{SCJ}}$  of  $\mathbb{C}_1$ ,  $\mathbb{C}_2$  and  $\mathbb{C}_3$ .
2. Is there another SCJ median of  $\mathbb{C}_1$ ,  $\mathbb{C}_2$  and  $\mathbb{C}_3$  that is distinct from  $\mathbb{M}_{\text{SCJ}}$ ?  
(Justify your answer by giving a distinct median or explaining why it does not exist.)
3. Is  $\mathbb{M}_{\text{SCJ}}$  also a breakpoint median of  $\mathbb{C}_1$ ,  $\mathbb{C}_2$  and  $\mathbb{C}_3$ ?

If *no*: Compute a breakpoint median of  $\mathbb{C}_1$ ,  $\mathbb{C}_2$  and  $\mathbb{C}_3$ .

If *yes*: Is there another breakpoint median of  $\mathbb{C}_1$ ,  $\mathbb{C}_2$  and  $\mathbb{C}_3$  that is distinct from  $\mathbb{M}_{\text{SCJ}}$ ?  
(Justify your answer by giving a distinct median or explaining why it does not exist.)

### Exercise 3 (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  $h = h_{\text{DCJ}}(\mathbb{D})$ .
3. Give a halving scenario with  $h$  optimal DCJ operations that transform  $\mathbb{D}$  into a perfectly duplicated genome  $2 \cdot \mathbb{H}$ .

### Exercise 4 (Inversion distance with component tree)

(6 pts)

Given canonical circular chromosomes

$$\mathbb{A} = (0\ \bar{3}\ 1\ 2\ 4\ 6\ 5\ 7\ \bar{15}\ \bar{13}\ \bar{14}\ \bar{12}\ \bar{10}\ \bar{11}\ 9\ 8)$$

$$\mathbb{B} = (0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ 15)$$

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: enter the values for chromosome  $\mathbb{A}$ , without the first value (0) and assume that the first vertex of the outputted diagram is  $0^h$  and the last vertex is  $0^t$ .)
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  $d_{\text{INV}}(\mathbb{A}, \mathbb{B})$ .