

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 2, 05.11.2020

Exercise 1 (Balanced genomes)

(3 pts)

Given two balanced genomes

$$\mathbb{B}_1 = (123185567778459) \quad \text{and}$$

$$\mathbb{B}_2 = (214\bar{5}7785678\bar{9}531),$$

1. Give the sets $\mathcal{F}_* = \mathcal{F}(\mathbb{B}_1) \cap \mathcal{F}(\mathbb{B}_2)$ (common families) and $\mathcal{G}_* = \mathcal{G}(\mathbb{B}_1) \cap \mathcal{G}(\mathbb{B}_2)$ (common genes).
2. Give a matching of genes of \mathbb{B}_1 and \mathbb{B}_2 and let \mathbb{C}_1 and \mathbb{C}_2 be the (canonical) matched genomes.
3. Give the sets $\alpha_* = \alpha(\mathbb{C}_1) \cap \alpha(\mathbb{C}_2)$ (common adj.) and $\gamma_* = \gamma(\mathbb{C}_1) \cap \gamma(\mathbb{C}_2)$ (common telomeres).

Exercise 2 (Breakpoint and SCJ distances)

(3 pts)

Given two canonical genomes

$$\mathbb{G}_1 = [34\bar{8}] [21756] (1112\bar{1}3910) [\bar{1}5\bar{1}4] \quad \text{and}$$

$$\mathbb{G}_2 = [12345] (67) [89101112131415],$$

1. What is the breakpoint distance between \mathbb{G}_1 and \mathbb{G}_2 ?
2. What is the SCJ distance between \mathbb{G}_1 and \mathbb{G}_2 ?
3. Explain the difference between the two distances (if any).

Exercise 3 (Bounds for SCJ distance)

(4 pts)

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

$$d_{BP}(\mathbb{G}_1, \mathbb{G}_2) \leq d_{SCJ}(\mathbb{G}_1, \mathbb{G}_2) \leq 2d_{BP}(\mathbb{G}_1, \mathbb{G}_2).$$

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

Exercise 4 (SCJ distance)

(6 pts)

1. Compute the SCJ distance between all pairs of the following three canonical circular genomes:

$$\mathbb{C}_{1c} = (123456789101112),$$

$$\mathbb{C}_{2c} = (\bar{1}\bar{2}\bar{3}\bar{4}\bar{5}\bar{6}\bar{7}\bar{8}\bar{9}\bar{10}\bar{11}\bar{12}) \quad \text{and}$$

$$\mathbb{C}_{3c} = (14\bar{3}5\bar{6}72\bar{9}\bar{8}\bar{10}1112).$$

2. Compute the SCJ distance between all pairs of the following three canonical linear genomes:

$$\mathbb{C}_{1L} = [123456789101112],$$

$$\mathbb{C}_{2L} = [\bar{1}\bar{2}\bar{3}\bar{4}\bar{5}\bar{6}\bar{7}\bar{8}\bar{9}\bar{10}\bar{11}\bar{12}] \quad \text{and}$$

$$\mathbb{C}_{3L} = [14\bar{3}5\bar{6}72\bar{9}\bar{8}\bar{10}1112].$$