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 2, 28.10.2021

Exercise 1 (Breakpoint and SCJ distances) (3 pts)

Given two canonical genomes
$\mathbb{A}^f_\triangleright = [3\ 4\ \overline{8}]\ \ [2\ 1\ 7\ 5\ 6]\ \ (11\ 12\ \overline{13}\ 9\ 10)\ \ [\overline{15}\ \overline{14}]$ and
$\mathbb{B}^f_\triangleright = [1\ 2\ 3\ 4\ 5]\ \ (6\ 7)\ \ [8\ 9\ 10\ 11\ 12\ 13\ 14\ 15]$,

1. What is the breakpoint distance between $\mathbb{A}^f_\triangleright$ and $\mathbb{B}^f_\triangleright$?
2. What is the SCJ distance between $\mathbb{A}^f_\triangleright$ and $\mathbb{B}^f_\triangleright$?
3. Explain the difference between the two distances (if any).

Exercise 2 (Bounds for SCJ distance) (4 pts)

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

$$d_{BP}(\mathbb{A}^f_\triangleright, \mathbb{B}^f_\triangleright) \leq d_{SCJ}(\mathbb{A}^f_\triangleright, \mathbb{B}^f_\triangleright) \leq 2d_{BP}(\mathbb{A}^f_\triangleright, \mathbb{B}^f_\triangleright)\,.$$

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

Exercise 3 (SCJ median) (8 pts)

Consider the following canonical genomes:

$$\mathbb{C}^f_{\triangleright 1} = [1\ 2\ 3\ 4\ 5],\ \ \mathbb{C}^f_{\triangleright 2} = [1\ 2\ \overline{3}\ 5\ 4],\ \ \mathbb{C}^f_{\triangleright 3} = [2\ \overline{3}\ 1\ 4\ 5]\ \text{ and }\ \mathbb{C}^f_{\triangleright 4} = [2\ 3\ \overline{1}\ 4\ 5].$$

Now let $\mathcal{S}^3 = \{\mathbb{C}^f_{\triangleright 1}, \mathbb{C}^f_{\triangleright 2}, \mathbb{C}^f_{\triangleright 3}\}$ and $\mathcal{S}^4 = \mathcal{S}^3 \cup \{\mathbb{C}^f_{\triangleright 4}\}$.

For each of the two sets $\mathcal{S}^3$ and $\mathcal{S}^4$:

1. Compute a general SCJ median $\mathbb{M}^k_{SCJ}$ of $\mathcal{S}^k$.
2. Is there another SCJ median of $\mathcal{S}^k$ that is distinct from $\mathbb{M}^k_{SCJ}$?
   (Justify your answer by giving a distinct median or explaining why it does not exist.)

Exercise 4 (Occurrences of adjacencies in canonical genomes) (3 pts)

Given a set of canonical genomes $\mathcal{A} = \{\mathbb{A}^f_{\triangleright 1}, ..., \mathbb{A}^f_{\triangleright n}\}$, prove that there cannot be a pair of conflicting adjacencies $m^x n^y$, $m^x p^w$ with both occuring in more than half of the genomes, that is $\phi(m^x n^y, \mathcal{A}) > \frac{|\mathcal{A}|}{2}$ and $\phi(m^x p^w, \mathcal{A}) > \frac{|\mathcal{A}|}{2}$ with $n^y \neq p^w$.