Algorithms in Comparative Genomics

Universität Bielefeld, SS 2024
Dr. Marília D. V. Braga · Dr. Roland Wittler · M.Sc. Leonard Bohnenkämper

https://gi.cebitec.uni-bielefeld.de/teaching/2024summer/cg

Exercise sheet 2, 21.04.2023

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

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

1. What is the breakpoint distance between $\mathbb{A}$ and $\mathbb{B}$?
2. What is the SCJ distance between $\mathbb{A}$ and $\mathbb{B}$?
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

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

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:

$$\mathbb{G}_1 = [1\ 2\ 3\ 4\ 5],\ \ \mathbb{G}_2 = [1\ 2\ \overline{3}\ 5\ 4],\ \ \mathbb{G}_3 = [2\ \overline{3}\ 1\ 4\ 5]\ \text{ and }\ \mathbb{G}_4 = [2\ 3\ \overline{1}\ 4\ 5].$$

Now let $\mathcal{S}^3 = \{\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3\}$ and $\mathcal{S}^4 = \mathcal{S}^3 \cup \{\mathbb{G}_4\}$.

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.)
3. Is $\mathbb{M}^3_{SCJ}$ also a breakpoint median of $\mathbb{G}_1$, $\mathbb{G}_2$ and $\mathbb{G}_3$?

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

   If *yes*: Is there another breakpoint median of $\mathbb{G}_1$, $\mathbb{G}_2$ and $\mathbb{G}_3$ that is distinct from $\mathbb{M}^3_{SCJ}$?
   (Justify your answer by giving a distinct median or explaining why it does not exist.)

Exercise 4 (Linear SCJ Median) (3 pts)
Find a polynomial time algorithm to calculate the linear SCJ median for k canonical genomes $\mathbb{G}_1 ... \mathbb{G}_k$, that is, the genome $\mathbb{M}_l$ consisting only of linear chromosomes that minimizes $m_l(\mathbb{M}_l) = \sum_{i=1}^{k} d_{SCJ}(\mathbb{M}_l, \mathbb{G}_i)$.