Algorithms in Comparative Genomics

Bielefeld University, Summer Semester 2025
Dr. Marília D. V. Braga · Prof. Dr. Jens Stoye
https://gi.cebitec.uni-bielefeld.de/teaching/2025summer/cg

Exercise sheet 5, discussion: 30.05.2025

Exercise 1 (Occurrences of adjacencies in canonical genomes)
Given a set of canonical genomes $\mathcal{A} = \{\mathbb{A}_1, ..., \mathbb{A}_n\}$, prove that there cannot be a pair of conflicting adjacencies $\gamma_1\gamma_2$ and $\gamma_1\gamma_3$ with both occuring in more than half of the genomes, that is, we cannot have:

$$\phi(\gamma_1\gamma_2, \mathcal{A}) > \frac{|\mathcal{A}|}{2} \text{ and } \phi(\gamma_1\gamma_3, \mathcal{A}) > \frac{|\mathcal{A}|}{2} \text{ with } \gamma_2 \neq \gamma_3$$

.

Exercise 2 (SCJ Halving)
For each of the two duplicated genomes:

$\mathbb{D}_1 = \{\, [\,\bar{1}\ 4\,]\,,\ [\,\bar{5}\ \bar{3}\ 2\ \bar{4}\ 1\ \bar{2}\ 6\ \bar{6}\ 3\ 5\ 7\ \bar{7}\,]\,\}$

$\mathbb{D}_2 = \{\, (\,\bar{4}\ 2\ 3\ 4\,)\,,\ (\,\bar{2}\ 1\ \bar{1}\ \bar{3}\,)\,\}$

1. Calculate the SCJ halving distance.
2. Find two optimal SCJ scenarios sorting $\mathbb{D}_i$ into perfectly duplicated genomes $\mathbb{P}_i$ and $\mathbb{P}'_i$, such that $\mathbb{P}_i$ has the maximum and $\mathbb{P}'_i$ has the minimum number of telomeres.

Exercise 3 (SCJ and breakpoint median)
Consider the following canonical genomes:

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

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 $\mathbb{M}^k_{\text{SCJ}}$ of $\mathcal{S}^k$ and give its score.
2. Is there another SCJ median of $\mathcal{S}^k$ that is distinct from $\mathbb{M}^k_{\text{SCJ}}$?
   (Justify your answer by giving a distinct median or explaining why it does not exist.)
3. Is $\mathbb{M}^3_{\text{SCJ}}$ 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 $\mathbb{M}^3_{\text{SCJ}}$?
   (Justify your answer by giving a distinct median or explaining why it does not exist.)