Algorithms in Comparative Genomics

Exercise sheet 5, 12.5.2023

Exercise 1 (DCJ model) (8 pts)

Given the following pair of canonical genomes:

$$\mathbb{A} = \{\, [\,2\ \bar{1}\ \bar{4}\ \bar{3}\,]\,,\ [\,5\ 7\ 6\ 9\ 8\ 10\,]\,\} \quad \text{and} \quad \mathbb{B} = \{\, [\,1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\,]\,\}$$

1. Draw the relational graph $RG(\mathbb{A}, \mathbb{B})$ and compute the DCJ distance between $\mathbb{A}$ and $\mathbb{B}$.

2. Give an optimal DCJ sorting scenario from $\mathbb{A}$ to $\mathbb{B}$. Name the operations in your sorting scenario (fusions, fissions, inversions, translocations, etc).

3. Generalize your procedure to a polynomial algorithm transforming any pair of canonical genomes into another with an optimal number of operations.

4. If your optimal sorting scenario contains circular chromosomes in two consecutive intermediate genomes, find an alternative optimal scenario in which each circular intermediate chromosome is immediately reintegrated (no need to worry about running time of your procedure).

5. Can you find yet another alternative optimal scenario without circular intermediates?

Exercise 2 (Solution space of sorting by DCJ) (6 pts)

1. Given genomes $\mathbb{A} = \{\,(1\ 2\ 3)\,\}$ and $\mathbb{B} = \{\,(1\ 3\ 2)\,\}$, how many different optimal DCJ scenarios sorting $\mathbb{A}$ into $\mathbb{B}$ can you find?

2. Given two canonical genomes $\mathbb{G}_1$ and $\mathbb{G}_2$, let c_k be a k-cycle in $RG(\mathbb{G}_1, \mathbb{G}_2)$ (with $k \geq 4$), and let $E_\alpha(\mathbb{G}_1, c_k)$ be the set of edges corresponding to adjacencies of genome $\mathbb{G}_1$ that are in c_k.

   What is the number of distinct DCJ operations that modify genome $\mathbb{G}_1$ and split c_k into two cycles?

   (*Hint: For each pair of edges in $E_\alpha(\mathbb{G}_1, c_k)$, there is exactly one DCJ splitting c_k into two cycles.*)