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 4, 19.11.2020

Exercise 1 (Circular breakpoint median) (6 + 3* pts)

1. Let $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$ be canonical circular genomes. For another canonical circular genome $\mathbb{M}$, let:

$$\mathrm{s}_{\mathrm{BP}}(\mathbb{M}) \quad = \quad \mathrm{d}_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_1) \quad + \quad \mathrm{d}_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_2) \quad + \quad \mathrm{d}_{\mathrm{BP}}(\mathbb{M}, \mathbb{C}_3)\,.$$

   Show that

$$\mathrm{s}_{\mathrm{BP}}(\mathbb{M}) \quad \geq \quad 2n \quad - \quad 2a_3 \quad - \quad a_2\,,$$

   where $n = |\mathcal{G}_\star|$ and $a_i = |\{xy : \phi(xy, \mathbb{C}_{1..3}) = i\}|$, i. e., a_3 is the number of adjacencies common to $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$ and a_2 is the number of adjacencies that occur in exactly two genomes among $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$.

2. Let $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$ be canonical circular **unichromosomal** genomes. Show how the problem of computing a circular unichromosomal breakpoint median of $\mathbb{C}_1$, $\mathbb{C}_2$ and $\mathbb{C}_3$ can be reduced to the Travelling Salesman Problem (TSP).

3. (3* extra pts) Explain how to extend the NP-hardness proof of breakpoint median of unichromosomal circular genomes to the breakpoint median of unichromosomal linear genomes.

Exercise 2 (DCJ halving) (6 pts)

For each of the following duplicated genomes, compute the DCJ halving distance $\mathrm{h}_i = \mathrm{h}_{\mathrm{DCJ}}(\mathbb{D}_{\mathtt{i}})$ and find a perfectly duplicated genome $2{\cdot}\mathbb{H}_{\mathtt{i}}$ with a matching between the genes of $2{\cdot}\mathbb{H}_{\mathtt{i}}$ and $\mathbb{D}_{\mathtt{i}}$, giving a halving scenario with h_i optimal DCJ operations that transform $\mathbb{D}_{\mathtt{i}}$ into $2{\cdot}\mathbb{H}_{\mathtt{i}}$.

1. $\mathbb{D}_1 = [3\ 5\ \bar{4}\ 2\ \bar{5}]\ \ [2\ 1]\ \ [3\ 4\ 1]$
2. $\mathbb{D}_2 = (3\ 5\ \bar{4}\ 2\ \bar{5})\ \ (2\ 1\ \bar{1}\ 3\ 4)$

Exercise 3 (DCJ halving) (4 pts)

Denote by $\rho\mathbb{G}$ the genome obtained after applying a DCJ operation ρ to a genome $\mathbb{G}$.

Now consider the duplicated genome:

$$\mathbb{D} = [\bar{4}\ 1\ \bar{4}\ \bar{3}\ 2]\ \ [\bar{2}\ 3\ 1]\ \ [5\ \bar{5}].$$

List all possible optimal (1st step) DCJ halving operations that could be applied to $\mathbb{D}$, that is, the set of DCJ operations $R = \{\rho : \mathrm{h}_{\mathrm{DCJ}}(\mathbb{D}) = \mathrm{h}_{\mathrm{DCJ}}(\rho\mathbb{D}) + 1\}$.