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 9, 7.1.2021

Exercise 1 (Canonical DCJ model with capping) (6 pts)

Consider the following canonical genomes:

$$\begin{aligned} \mathbb{A} &= [4\ 1]\ \ [6\ 3\ 2\ 5]\ \ [7\ 8\ \bar{9}] \quad \text{and} \\ \mathbb{B} &= [1\ 2\ 3\ 4]\ \ [5\ 6]\ \ [7\ 8\ 9] \end{aligned}$$

1. Construct the relational graph $RG(\mathbb{A}, \mathbb{B})$.
2. For each component C of the relational graph $RG(\mathbb{A}, \mathbb{B})$, give the type of C (cycle, $\mathbb{AB}$-path, $\mathbb{AA}$-path or $\mathbb{BB}$-path).
3. Compute the DCJ distance $\mathrm{d}_{\mathrm{DCJ}}(\mathbb{A}, \mathbb{B})$.
4. Modify the relational graph into an optimal capped relational graph with the minimum number of caps.
5. Derive a pair of canonical circular genomes from the capped relational graph, **maximizing** the number of circular chromosomes in each genome.

Exercise 2 (Singular DCJ-indel model with capping) (12 pts)

Consider the following singular genomes:

$$\begin{aligned} \mathbb{A} &= [\mathrm{a}_1\ 4\ 1]\ \ [\mathrm{a}_2\ 6\ \mathrm{a}_3\ 3\ \mathrm{a}_4\ 2\ \mathrm{a}_5\ 5]\ \ [\mathrm{a}_6\ \mathrm{a}_7\ 7\ 8\ \mathrm{a}_8\ \bar{9}]\ \ [\mathrm{a}_9] \quad \text{and} \\ \mathbb{B} &= [\mathrm{b}_1\ 1\ \mathrm{b}_2\ 2\ \mathrm{b}_3\ 3\ \mathrm{b}_4\ 4\ \mathrm{b}_5]\ \ [5\ \mathrm{b}_6\ 6]\ \ [\mathrm{b}_7\ 7\ 8\ \mathrm{b}_8\ 9]\ \ (\mathrm{b}_9\ \mathrm{b}_{10}) \end{aligned}$$

1. Give the sets of genes $\mathcal{G}_\star$, $\mathcal{A}$ and $\mathcal{B}$.
2. Construct the relational graph $RG(\mathbb{A}, \mathbb{B})$.
3. For each component C of the relational graph $RG(\mathbb{A}, \mathbb{B})$:
   (a) Give the type of C (cycle, singleton, $\mathbb{AB}$-path, $\mathbb{AA}$-path or $\mathbb{BB}$-path).
   (b) Give the number of runs $\Lambda(C)$;
   (c) Give the run-type of C (ε, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{AB}$ or $\mathcal{BA}$);
   (d) Compute the minimum number of indels $\lambda(C)$ that are necessary for sorting C separately.
4. Find all chains of deducting recombinations.
5. Compute the DCJ-indel distance $\mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A}, \mathbb{B})$.
6. Modify the relational graph into an optimal capped relational graph with the minimum number of caps.
7. Derive a pair of singular circular genomes from the capped relational graph, minimizing the number of circular chromosomes in each genome.