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 6, discussion: 06.06.2025

Exercise 1 (Study the DCJ-indel model by hand)
Consider the linear genomes $\mathbb{A} = [\ 1\ \ \mathtt{X}\ \ \bar{2}\ \ \mathtt{Y}\ \ 3\ ]$ and $\mathbb{B} = [\ 1\ \ \mathtt{W}\ \ 2\ \ \mathtt{Z}\ \ 3\ ]$.

1. Draw the relational graph $AG(\mathbb{A}, \mathbb{B})$. What is the DCJ-indel distance between $\mathbb{A}$ and $\mathbb{B}$?
2. Find an optimal scenario with 3 indels, applying only optimal DCJs. Redraw the relational graph after the first DCJ.
3. Find an optimal scenario with 2 indels. Redraw the relational graph after the first DCJ.
   (Hint: Find a neutral DCJ that reduces the number of runs.)

Exercise 2 (DCJ-indel potential)
Prove the formula for the DCJ-indel potential λ of a cycle C of the adjacency graph $AG(A, B)$,

$$\lambda(C) = \begin{cases} 0 & \text{if } \Lambda(C) = 0 \\ 1 & \text{if } \Lambda(C) = 1 \\ \frac{\Lambda(C)}{2} + 1 & \text{otherwise } (\Lambda(C) \text{ even}) \end{cases},$$

where $\Lambda(C)$ is the number of runs of indels in C.
How does the corresponding formula for a path P with $\Lambda(P)$ runs look like?

Exercise 3 (Triangle inequality)
Consider the two genomes $\mathbb{A} = [\ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ ]$ and $\mathbb{B} = [\ \bar{3}\ \ 1\ \ 2\ \ \bar{6}\ \ 5\ ]$.

1. Find a third genome $\mathbb{C}$ such that the triangle inequality for the DCJ-indel distance does not hold, i.e.
$$\mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A}, \mathbb{B}) > \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A}, \mathbb{C}) + \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{C}, \mathbb{B}).$$
2. Remember the function $\mathrm{m}^{\mathrm{ID}}_{\mathrm{DCJ}}$ from class such that the triangle inequality holds for any set of three singular genomes. Which distances do you get for the genomes $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$ from above?