Algorithms in Comparative Genomics

Universität Bielefeld, WS 2021/2022
Dr. Marília D. V. Braga · Leonard Bohnenkämper

https://gi.cebitec.uni-bielefeld.de/teaching/2021winter/cg

Exercise sheet 8, 16.12.2021

Exercise 1 (Singular DCJ-indel model) (10 pts)

Consider the following two singular circular genomes:

$$\mathbb{A} = (\, 1\ a_1\ \bar{2}\ a_2\ 3\ a_3\ \bar{4}\ a_4\,) \quad \text{and}$$

$$\mathbb{B} = (\, 1\ b_1\ 2\ b_2\ 3\ b_3\ 4\ b_4\,)$$

1. For each cycle C of the relational graph $RG(\mathbb{A}, \mathbb{B})$:
   (a) Give the number of runs $\Lambda(C)$;
   (b) Give the run-type of C (ε, $\mathcal{A}$, $\mathcal{B}$ or $\mathcal{AB}$);
   (c) Compute the minimum number of indels $\lambda(C)$ that are necessary for sorting C separately.

2. Compute the DCJ-indel distance $d_{DCJ}^{ID}(\mathbb{A}, \mathbb{B})$.

3. In the following three subtasks you need to find optimal sequences of DCJ and indel operations sorting $\mathbb{A}$ into $\mathbb{B}$. Please list all intermediate genomes of each sorting sequence and, at each step, indicate the affected adjacencies and the corresponding applied operation (specifying for each indel whether it is an insertion or a deletion and for each DCJ whether it is an inversion, circular excision or circular integration):

   (a) Give an optimal sequence of DCJ and indel operations sorting $\mathbb{A}$ into $\mathbb{B}$, using only internal gaining DCJs and indels. (The number of indels here is given by the sum of the indel-potentials of the cycles in $RG(\mathbb{A}, \mathbb{B})$.)

   (b) An internal neutral DCJ operation applied to a cycle C with at least 4 runs can decrease the number of runs of C by -2, consequently decreasing the indel-potential of C by -1. Give another optimal sequence of DCJ and indel operations sorting $\mathbb{A}$ into $\mathbb{B}$, now **maximizing** the number of neutral DCJs. (Such a sorting sequence has exactly two indels per cycle of $RG(\mathbb{A}, \mathbb{B})$ that has at least 2 runs.)

   (c) A losing DCJ operation applied to two cycles, merging two pairs of runs, can decrease the overall indel-potential by -2. Give another optimal sequence of DCJ and indel operations sorting $\mathbb{A}$ into $\mathbb{B}$, now **maximizing** the number of losing and neutral DCJs. (Such a sorting sequence has exactly two indels for the whole set of cycles of $RG(\mathbb{A}, \mathbb{B})$ that have at least 2 runs.)

Exercise 2 (Deducting path recombinations) (6 pts)

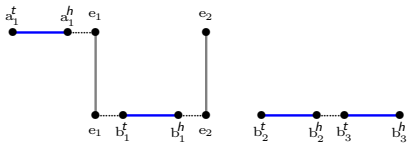
For each of the following deducting path recombinations, draw the cuts and the resultants of the given types with the joins, so that the given $\Delta_{\text{DCJ}}^{\lambda}$ is achieved. (Recall that both cuts must be done in the same genome and each cut either breaks an adjacency or is next to a telomere.)

1. Example: $\Delta_{\text{DCJ}}^{\lambda} = -1$

$\mathbb{AA}_{\mathcal{AB}} \quad + \quad \mathbb{BB}_{\mathcal{A}} \quad \Rightarrow \quad \mathbb{AB}_{\varepsilon} + \quad \mathbb{AB}_{\mathcal{BA}}$

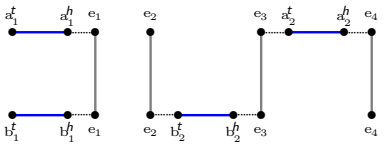
2. $\Delta_{\text{DCJ}}^{\lambda} = -1$

$\mathbb{AA}_{\mathcal{AB}} \quad + \quad \mathbb{BB}_{\mathcal{B}} \quad \Rightarrow \quad \mathbb{AB}_{\varepsilon} \quad + \quad \mathbb{AB}_{\mathcal{AB}}$



3. $\Delta_{\text{DCJ}}^{\lambda} = -1$

$\mathbb{AB}_{\mathcal{AB}} \quad + \quad \mathbb{AB}_{\mathcal{BA}} \quad \Rightarrow \quad \mathbb{AB}_{\mathcal{A}} \quad + \quad \mathbb{AB}_{\mathcal{B}}$



4. $\Delta_{\text{DCJ}}^{\lambda} = -2$

$\mathbb{AA}_{\mathcal{AB}} \quad + \quad \mathbb{BB}_{\mathcal{AB}} \quad \Rightarrow \quad \mathbb{AB}_{\mathcal{A}} \quad + \quad \mathbb{AB}_{\mathcal{B}}$