

Algorithms for Genome Rearrangement
Summer 2017

Exercises

Exercise 07, guest lecture 02.06.2017

1. Consider the following two genomes: (6 P)

$A = (1 \ a \ 3 \ b \ -8 \ 7 \ -6 \ c \ 4 \ 5 \ d \ 9 \ 11 \ e \ -13 \ 10 \ -12 \ f \ 14 \ -2 \ 15 \ 17 \ g \ 22 \ h \ 18 \ 20 \ 19 \ 21 \ i \ 16 \ j)$

$B = (1 \ 2 \ z \ 3 \ y \ 4 \ 5 \ x \ 6 \ 7 \ 8 \ w \ 9 \ 10 \ v \ 11 \ 12 \ 13 \ 14 \ u \ 15 \ t \ 16 \ 17 \ s \ 18 \ r \ 19 \ q \ 20 \ 21 \ p \ 22 \ n)$

- (a) Give the sets \mathcal{G} , \mathcal{A} and \mathcal{B} .
- (b) Construct the Mastergraph $MG(A, B)$.
(The top row contains all \mathcal{G} -adjacencies of genome A and the bottom row contains all \mathcal{G} -adjacencies of genome B .)
- (c) For each cycle C of the Mastergraph:
Give the label-type of C (unlabelled, \mathcal{A} -cycle, \mathcal{B} -cycle, \mathcal{AB} -cycle)
Give the number of “segments”: $\Lambda(C)$
Compute the minimum number of indels $\lambda(C)$ that are necessary for C (using the formula from the lecture):

$$\lambda(C) = \left\lceil \frac{\Lambda(C) + 1}{2} \right\rceil$$

- (d) Compute the overall DCJ-indel distance for the two genomes using:

$$d_{\text{DCJ}}^{\text{id}}(A, B) = d_{\text{DCJ}}(A, B) + \sum_{C \in MG(A, B)} \lambda(C)$$

2. Consider the following two genomes: (6 P)

$A = (1 \ x \ -2 \ y \ -3 \ 4)$

$B = (1 \ r \ 2 \ s \ 3 \ t \ 4)$

- (a) Give an *optimal* sorting scenario (it has minimum length: $d_{\text{DCJ}}^{\text{id}}$ steps).
- (b) How can $d_{\text{DCJ}}^{\text{id}}$ be achieved if the *merging* of labels (or “segments”) is done by DCJs instead of insertions or deletions?
- (c) With the ideas from b) find a different *optimal* scenario that uses more DCJs and fewer indels.

3. If we have linear genomes we might have paths. What are the different paths that can exist in combination with labels? What is the best λ we can achieve for each? (for fun)