Bielefeld University
Faculty of Technology

Algorithms in Comparative Genomics, Winter 2018/19

Dr. Daniel Dörr

Exercises

Exercise 05, 22.11.2018

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 adjacency graph $AG(A, B)$.
   (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 adjacency graph:
   Give the label-type of c (ε-, $\mathcal{A}$-cycle, $\mathcal{B}$-cycle, $\mathcal{AB}$-cycle)
   Give the number of runs: $\Lambda(c)$
   Compute the minimum number of indels $\lambda(c)$ that are necessary for c (using the formula from the lecture notes):

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

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

   $$d_{\mathrm{DCJ}}^{id}(A, B) = d_{\mathrm{DCJ}}(A, B) + \sum_{\substack{c \in AG(A,B),\\ \Lambda(c) \geq 1}} \lambda(c)$$

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_{DCJ}^{id} steps).

   (b) How can d_{DCJ}^{id} be achieved if the *merging* of labels (or runs) 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.

Discussion of solutions in tutorial on 29.11.2018