Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye

Algorithms in Comparative Genomics
Summer 2018

Exercises

Number 6, return 2018 June 22

1. Consider the linear genomes $A = [\ 1 \quad a \quad -2 \quad b \quad 3\ ]$ and $B = [\ 1 \quad c \quad 2 \quad d \quad 3\ ]$.

   (a) Draw the adjacency graph $AG(A, B)$. What is the DCJ-indel distance between A and B?

   (b) Find a scenario with 3 indels, applying only optimal DCJs. Redraw the adjacency graph after the first DCJ.

   (c) Find a scenario with 2 indels. Redraw the adjacency graph after the first DCJ.
   (Hint: Find a neutral DCJ that reduces the number of runs.)

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

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

   where $\Lambda(C) > 0$ is the number of runs of indels in C.

   How does the corresponding formula for a path P with $\Lambda(P)$ runs look like?

3. Consider the following two genomes $A = [\ 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6\ ]$ and $B = [\ -3 \quad 1 \quad 2 \quad -6 \quad 5\ ]$.

   (a) Find a third genome C such that the triangle inequality for the uncorrected DCJ-indel distance does not hold, i.e.

$$d_{\text{DCJ-indel}}(A, B) > d_{\text{DCJ-indel}}(A, C) + d_{\text{DCJ-indel}}(C, B).$$

   (b) Show how the distance can be corrected such that the triangle inequality holds again.