

Algorithms in Comparative Genomics

Bielefeld University, WS 2019/20

Dr. Roland Wittler

<https://gi.cebitec.uni-bielefeld.de/teaching/2019winter/cg>

Exercise sheet 4, 07.11.2019

Exercise 1 (Genome halving under DCJ)

Given an all-duplicates genome with three linear chromosomes

$$G = [3 \ 5 \ -4 \ 2 \ -5] \ [2 \ 1] \ [3 \ 4 \ 1],$$

solve the genome halving problem under the DCJ distance, i.e., find a perfectly duplicated genome H with smallest DCJ distance to G .

Exercise 2 (Guided genome halving under DCJ)

Research the *guided* genome halving problem under the DCJ distance and its complexity.

Exercise 3 (Genome halving under the breakpoint distance)

The following paragraph about the genome halving problem for the breakpoint distance d is from the paper “Multichromosomal median and halving problems under different genomic distances” by Eric Tannier, Chunfang Zheng and David Sankoff, BMC Bioinformatics 10:120, 2009. (Notation: For a gene g , $g1$ and $g2$ denote the two paralogous copies of g ; similarly, for a gene extremity x , the two copies are denoted $x1$ and $x2$.)

Let Δ be an all-duplicates genome on a gene set $\mathcal{G}$, and G be the graph on the vertex set containing (1) all the extremities of the genes in $\mathcal{G}$, and (2) one supplementary vertex t_x for every gene extremity x . For any pair of gene extremities x, y , draw an edge in G weighted by zero, one or two according to the number of adjacencies in Δ among $x1y1$, $x1y2$, $x2y1$, and $x2y2$. Now for any vertex x , draw an edge xt_x weighted by half the number of telomeres among $x1$ and $x2$ in Δ . Finally, put an edge of weight 0 between t_xt_y for all pairs of gene extremities x, y .

For a genome M on $\mathcal{G}$, define a perfect matching, also called M , by including edges xy and t_xt_y for each adjacency xy , and an edge xt_x for each telomere x . Let $w(M)$ be the weight of the matching M .

1. Prove the following:

Claim. For a genome M on $\mathcal{G}$, the perfect matching M thus constructed satisfies $w(M) = 2n - d(\Delta, M)$.

2. How can this construction be used to solve the genome halving problem under the breakpoint distance?

Hint: Have a closer look at the all-duplicates genome $\Delta = [1 \ 2 \ 3] \ [1 \ -3 \ -2]$.