

Algorithms in Genome Research
Winter 2017/2018

Exercises

Number 11, Discussion: 2018 January 26

1. **Median of Three.**

Consider the problem of finding a median M of three genomes A , B and C for a distance d satisfying the *triangle inequality* property: $d(X, Y) \leq d(X, W) + d(W, Y)$. Prove the well-known lower bound

$$d(M, A) + d(M, B) + d(M, C) \geq \frac{1}{2}(d(A, B) + d(A, C) + d(B, C))$$

2. **Small Parsimony—hands-on.**

Consider the phylogenetic tree $((A, B), (C, D))$; with the following gene orders (signed permutations) at its leaves. Perform the bottom-up phase of the Fitch Algorithm for each adjacency (presence/absence). Only consider adjacencies that occur at any leaf.

$A = 1\ 2\ 3\ 4$
 $B = 1\ 3\ 2\ 4$
 $C = 1\ 2\ 3\ 4$
 $D = 2\ 1\ 3\ 4$

- Would initializing the root with the presence of an adjacency in case of ambiguity result in consistent sets of adjacencies?
- How many adjacencies would be reconstructed, if instead the absence of an adjacency is preferred? What is the total cost of the scenario?
- Let X be the set of all adjacencies with candidate set $\{1\}$ or $\{0, 1\}$ at the root node. Formulate a corresponding instance of the MAX-ROW-component-Ci1P, (manually) find a maximum weight consistent subset $X' \subseteq X$, and perform the top-down phase starting with the corresponding root labeling. Do you obtain the same total cost?

3. **Small Parsimony—theory.**

In its original formulation [1], the authors proceed as in Exercise 2b, and prove in Lemma 6.2 that no conflicting adjacencies would be assigned to any node in the tree during the top-down phase. Slightly modify this proof such that it also holds when proceeded as in Exercise 2c.

References

- [1] P. Feijao and J. Meidanis. SCJ: a breakpoint-like distance that simplifies several rearrangement problems. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 8(5):1318–1329, 2011.