Reconstruction of Ancestral Gene Orders

Algorithms in Genome Research

January 19, 2018

Fitch's algorithm for a consistent most parsimonious reconstruction

For each adjacency α do:

1. **(Bottom-up phase)** Collect putative states for an adjacency α of each node v, stored in a candidate set $B(\alpha, v)$.

   1a. (Leaves) For each leaf l, set

   $$B(\alpha, l) := \begin{cases} \{1\} & \text{if } \alpha \in G_l \\ \{0\} & \text{otherwise.} \end{cases}$$

   1b. (Internal nodes) Assume an internal node u with children v and w.

   $$B(\alpha, u) := \begin{cases} B(\alpha, v) \cap B(\alpha, w) & \text{if } B(\alpha, v) \cap B(\alpha, w) \neq \emptyset, \\ B(\alpha, v) \cup B(\alpha, w) & \text{otherwise.} \end{cases}$$

2. **(Top-down refinement)** Reconstruct most parsimonious labeling $F(\alpha, v)$:

   2a. (Root)

   $$F(\alpha, \text{root}) := \begin{cases} s & \text{if } B(\alpha, \text{root}) = \{s\}, \\ 0 & \text{otherwise, i.e., } B(\alpha, \text{root}) = \{0, 1\}. \end{cases}$$

   Second case to avoid conflicts. (In the general Fitch version, any state from B can be chosen. But we require consistency.)

   2b. (Other nodes) Consider v with parent node p, and let $s = F(\alpha, p)$.

   $$F(\alpha, v) := \begin{cases} s & \text{if } s \in B(\alpha, v), \\ 1 - s & \text{otherwise, i.e., set it to the other state.} \end{cases}$$

Definition (Ci1P). A matrix (or sub-matrix) has the *Circular Ones Property (Ci1P)* if its columns can be ordered such that in each row, either the 1 entries are consecutive (there is no 0 entry between two 1 entries), or the 0 entries are consecutive (there is no 1 entry between two 0 entries); in other words the 1 entries are consecutive when the order of columns is viewed as a circle.

Definition (component-Ci1P). A matrix is *component-Ci1P* if its columns can be partitioned such that a row has 1s only in one part and each part is Ci1P.

MAX-ROW-component-Ci1P [3]. Given a matrix with weighted rows, find a subset of rows of maximum cumulative weight such that the obtained sub-matrix is component-Ci1P. If each row contains exactly two ones, this is possible in polynomial time.

Note. In [3], each column is assigned a maximum cardinality m which we here set to one and thus omit.

References

[1] G. Bourque and P. A. Pevzner. Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome research*, 12(1):26–36, 2002.

[2] 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.

[3] J. Maňuch, M. Patterson, R. Wittler, C. Chauve, and E. Tannier. Linearization of ancestral multichromosomal genomes. *BMC bioinformatics*, 13(19):S11, 2012.

[4] D. Sankoff, G. Sundaram, and J. Kececioglu. Steiner points in the space of genome rearrangements. *International Journal of Foundations of Computer Science*, 7(01):1–9, 1996.