A new Data Structure for Genome Rearrangement Problems

Cleber Mira, Joao Meidanis

Keywords: Genome Rearrangements, Computational Biology

1 Abstract.

A genome of a species is a set of macromolecules (chromosomes) that encode in certain molecular segments (genes) the information necessary to produce every protein in a live being. A genome may be transformed into other genome by the action of mutational events called rearrangement events. A rearrangement event usually involves the exchange of segments of genes, change in the orientation of genes, and joining or breaking chromosomes.

Given two genomes the genome rearrangement problem consists of finding a minimum sequence of rearrangement events of a certain kind that transforms one genome into another. The number of rearrangement events in the sequence, usually called genomic distance, is used as a parameter in phylogeny reconstruction.

The minimum number of rearrangement events that transforms one genome into another can be used as parameter for the comparative analysis of genomes. We propose a new data structure for genome rearrangement problems based on an algebraic model for genomes. The data structure allows one to perform certain rearrangement events like fissions, fusions, reversals, transpositions, and block-interchanges in constant running time. Using such data structure we devise an algorithm for finding a minimum sequence of Fissions, Fusions, and Signed Reversals that transforms a genome into another. It is possible, with a few modifications, to use the same data structure to devise an algorithm for finding a minimum sequence of Generalized Transpositions that transforms a genome into another. Both algorithms solve the genome rearrangement problem in linear time. Those algorithms can be used to obtain the genomic distance in linear time.

We hope that the study of new representations for genomes and data structures for those representations encourage the research of new algorithms for genome rearrangement problems.

1Research supported by grants from FAPESP.
2Institute of Computing, University of Campinas (UNICAMP), Sao Paulo, Brazil E-mail: cleber@ic.unicamp.br
3Scylla Bioinformatics, Sao Paulo, Brazil E-mail: meidanis@scylla.com.br