A New Data Structure for Genome Rearrangement Problems

Cleber V. G. Mira, 1 Joao Meidanis, 2

Keywords: Genome Representation, Computational Biology

1 Statement of the Problem.

A DNA chromosome has two strands with complementary orientation. A set of genes $E$ and a permutation $\Gamma$ over $E$ is called a gene system, denoted by $(E, \Gamma)$, when $\text{Supp}(\Gamma) = E$ and $\Gamma \Gamma = \iota$. The permutation $\Gamma$ associates each gene to its complementary. For instance, given the set of genes $E = \{-4, -3, -2, -1, 1, 2, 3, 4\}$ and the permutation $\Gamma = (1 - 1)(2 - 2)(3 - 3)(4 - 4)$, the pair $(E, \Gamma)$ is a gene system since the support of $\Gamma$ is $E$ and $\Gamma \Gamma x = x$ for any $x \in E$.

Given a gene system $(E, \Gamma)$, a cycle $\alpha$ is called a strand when $x \in \text{Supp}(\alpha)$ implies $\Gamma x \notin \text{Supp}(\alpha)$ for each $x \in E$. The conjugation of a permutation $\alpha$ by $\beta$, denoted by $\beta \cdot \alpha$, is $\beta \alpha \beta^{-1}$ [2]. A chromosome is a product of two strands $\alpha$ and $\Gamma \cdot \alpha^{-1}$. Two chromosomes are disjoint when their supports are disjoint. A genome is a product of disjoint chromosomes. A fundamental property of genomes is: if $\pi$ is a genome, then $\Gamma \pi \Gamma = \pi^{-1}$.

A genome rearrangement problem consists of finding a minimum sequence of mutational events called rearrangement events that transform a genome into another.

2 Data Structure

The design of new data structures [3] allowed the improvement in performance of algorithms for genome rearrangement problems involving signed reversals [4] and transpositions [1]. We present a new data structure that is based on the algebraic formalism for genome rearrangements [2]. The Figure 1 illustrates the data structure.

The data structure can be used, with a few modifications, in algorithms for different genome rearrangement problems (e.g. involving generalized transpositions or fusions, fissions, and signed reversals).

Given the genomes $\pi$ and $\sigma$ in $(E, \Gamma)$, we can obtain a minimum sequence of fusions, fissions, and signed reversals that transforms the genome $\pi$ into $\sigma$ in $O(n)$, where $n = |E|$, using the data structure in Figure1.

References


1Research supported by grants from FAPESP(03/00731-3).
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
Figure 1: Genomes $\pi = (3 - 2 7)(-7 2 - 3)(-5 - 6)(6 5)(1 - 4 8)(-8 4 - 1)$ and $\sigma = (1 2 3 4 5 6 7 8)(-8 - 7 - 6 - 5 - 4 - 3 - 2 - 1)$ in the gene system $(E, \Gamma)$ where $E = \{x \mid |x| \in [8]\}$ and $\Gamma x = -x$ are represented by arrays $P$ and $S$. The permutation $\sigma \pi^{-1}$ is represented by the list $Q$. The array $T$ makes an entry indexed by a gene $x$ to point to the node containing $x$ and to the pair of circular lists in $Q$. In order to make the figure clear, we show the pointers of the first four genes of $T$ only.