# Analysis of Sorting by Transpositions based on Algebraic Formalism [1]

**Cleber V. G. Mira, [2] Joao Meidanis, [3]**

**Keywords:** Genome Rearrangements, Computational Biology

## 1 Statement of the Problem.

Genome rearrangements analysis focus on the relative positions of the same block of genes at two or more distinct genome sequences. Some mutational events, such as *transpositions*, affect the genome sequences solely in their ordering of blocks of genes. Given a permutation representing a genome $\pi = (\pi_1\ \pi_2\ \dots\ \pi_n)$, a *transposition* $\tau(i,j,k)$ for $1 \leq i < j \leq n$ and $1 \leq k \leq n$, but $k \notin [i,j]$, is the following operation on $\pi$.

$$\tau(i,j,k)\pi = (\pi_1\ \pi_2\ \dots\ \pi_{i-1}\ \pi_j\ \dots\ \pi_{k-1}\ \pi_i\ \dots\ \pi_{j-1}\ \pi_k\ \dots\ \pi_n)\,,$$

if $i < j < k$.

The *problem of transposition distance* consists of finding the minimum number of transpositions to transform one genome into another. That is:

$$\sigma = \tau_t\ \tau_{t-1}\ \dots\ \tau_1\ \pi$$

The number $t$ is the transposition distance $d_\tau(\pi,\sigma)$ between two genomes $\pi$ and $\sigma$. For example, consider the following sequence of transpositions which order the permutation (4 3 2 1 5):

$$\begin{aligned} \tau(1,4,5)\pi &= (1\ 4\ 3\ 2\ 5) \\ \tau(2,4,5)\tau(1,4,5)\pi &= (1\ 2\ 4\ 3\ 5) \\ \tau(3,4,5)\tau(2,4,5)\tau(1,4,5)\pi &= (1\ 2\ 3\ 4\ 5) \end{aligned}$$

## 2 Algebraic Formalism

The permutations can be analyzed through a graph representation called *cycle graph* [1]. However, we represent the permutations and transpositions by means of a new algebraic formalism developed by Meidanis and Dias [2]. In this approach a genome is described as a permutation on the symmetric group over $\{0,\ 1\ \dots\ n\}$. But we are interested in the cycle decomposition of the permutations in $S_n$. Since the transposition event does not change the orientation of a block, only one of the strands is considered in its cycle decomposition representation. A genome in the Algebraic Formalism is usually represented as:

$$\pi = (0\ \pi_1\ \pi_2\ \pi_3\ \dots\ \pi_n)(\overline{\pi_n}\ \dots\ \overline{\pi_3}\ \overline{\pi_2}\ \overline{\pi_1}\ -0)$$

Observe that the "dummy block" zero is used in this representation. The earlier permutation is a product of two disjoint cycles, each one representing a strand of the genome. As the

---

[1] Research supported by grants from FAPESP.

[2] Institute of Computing, University of Campinas (UNICAMP), Sao Paulo, Brazil E-mail: `cleber@ic.unicamp.br`

[3] Scylla Bioinformatics, Sao Paulo, Brazil E-mail: `meidanis@scylla.com.br`

transposition event does not change the orientation of the blocks of genes, we will not consider the strand which has the block $-0$.

This permutation is seen as a function which induces a circular order of its elements, such that $\pi_{i+1} = \pi(\pi_i)$. The *identity permutation*, 1, in the permutation group is $(1)(2)(3)\ \ldots\ (n)$. Each element in an 1-cycle in the cycle decomposition of a permutation is called a *fixed element*. Fixed elements are usually omitted in the cycle decomposition representation. The *support*, $Supp(\pi)$, of a permutation $\pi$ is the subset of elements not fixed in $\pi$.

The product of permutations, $\pi\sigma$, is performed in this way: for each element $x \in [n]$ is applied the composition $(\pi\sigma)(x) = \pi(\sigma(x))$. For instance, consider this example: $(3\ 2\ 5\ 1)(6\ 4\ 2) = (1\ 3\ 2\ 6\ 4\ 5)$. The *inverse permutation* of $\pi$ is the permutation $\pi^{-1}$, such that $\pi\pi^{-1} = 1$. To obtain the inverse permutation of a cycle $\pi$ is easy — the inverse of $\pi = (\pi_1\ \pi_2\ \ldots\ \pi_n)$ is $\pi^{-1} = (\pi_n\ \pi_{n-1}\ \ldots\ \pi_1)$. A permutation $\tau$ *divides* a permutation $\pi$, $\tau|\pi$, if and only if $|\pi\tau^{-1}| = |\pi| - |\tau|$, where $|\pi|$ is the *norm* of $\pi$; i.e. the minimum 2-cycle decomposition of $\pi$.

A transposition in this new approach is the permutation $\tau(\pi_u, \pi_v, \pi_w) = (\pi_u\ \pi_v\ \pi_w)$. To apply a transposition in the genome $\pi$ is to perform the product $\tau\pi$. For instance: $(4\ 2\ 5)(0\ 1\ 4\ 3\ 2\ 5) = (0\ 1\ 2\ 4\ 3\ 5)$.

A transposition $\tau$ is *applicable* to $\pi$ if $\tau\pi$ is a strand. Also, a transposition $\tau$ is applicable to $\pi$ if and only if $\tau|\pi$. There exists transpositions that are not applicable to a genome $\pi$. For example: $(4\ 5\ 2)(0\ 1\ 4\ 3\ 2\ 5) = (0\ 1\ 5)(2)(4\ 3)$. The length of a cycle $\alpha$ in the cycle decomposition of a permutation $\pi$ is $|Supp(\alpha)|$. A cycle is *odd*, if its length is odd.

## 3 Transposition Distance Bounds.

Let $|\pi|_3$ denotes the minimum number of 3-cycles $\tau_1, \tau_2, \ldots, \tau_k$, where $k = |\pi|_3$, such that $\pi = \tau_1\tau_2\ldots\tau_k$. The algebraic approach provides the following lower bound to the transposition distance. Notice that given a genome $\pi$ and $\tau_1\ \tau_2\ \ldots\ \tau_k\ \pi = \sigma$, such that $k$ is minimum, then $\tau_1\ \tau_2\ \ldots\ \tau_k = \sigma\pi^{-1}$. Therefore:

**Proposition 3.1 (Lower Bound)** $d_\tau(\pi, \sigma) \geq |\sigma\pi^{-1}|_3$.

The formula $\sigma\pi^{-1}$, which is called *Quotient*, is very important in the algebraic theory because it straightforwardly provides lower bounds for others rearrangement problems [2] and gives an algebraic relationship between the genomes $\pi$ and $\sigma$. Next we state that the previous lower bound is equivalent to the best known lower bound [1].

**Proposition 3.2** $|\pi|_3 = \frac{(n - c_{odd}(\pi))}{2}$

A *split* is a transposition not applicable to $\pi$. If we permit splits besides transpositions, then the split+transposition distance, $d_{st}(\pi, \sigma)$, is:

**Proposition 3.3 (Split+Transposition Distance)** $d_\tau(\pi, \sigma) = |\sigma\pi^{-1}|_3$.

# References

[1] V. Bafna and P. A. Pevzner, 1995. Sorting by Transpositions. In: *Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*,San Francisco, USA, pp. 614–623

[2] J. Meidanis and Z. Dias 2000.An Alternative Algebraic Formalism for Genome Rearrangements. In: *Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment and Evolution of Gene Families* "D. Sankoff and J. H. Nadeau",editors, Kluwer Academic Publishers. pp 213–223.