# The Algebraic Theory for Genome Rearrangements

Universität Bielefeld
Technische Fakultät

**Pedro Feijão[1,*] and João Meidanis[2]**

[1]Genome Informatics, Faculty of Technology, Bielefeld University, Germany [2]University of Campinas, Institute of Computing, Brazil and Scylla Bioinformatics, Brazil
[*]pfeijao@cebitec.uni-bielefeld.de

UNICAMP

## Introduction

**Genome rearrangements** are evolutionary events where large, continuous pieces DNA shuffle around, changing the order of genes in the genome of a species. Gene order data can be very useful in estimating the evolutionary distance between genomes, and also in reconstructing the gene order of ancestral genomes.
In this poster we will present the main ideas of the **Algebraic Theory** for genome rearrangements, proposed by Meidanis and Dias [1], and some recent results by Feijão and Meidanis [2].

## Algebraic Theory Definitions

- A **permutation** is a bijection in a set $\mathcal{G}$. Permutations are composed by **k-cycles**, where $k$ is the number of elements in the cycle. For instance, the permutation $(1\ 2\ 3)(4\ 5\ 6)$ has two 3-cycles, as shown in the figure to the right.

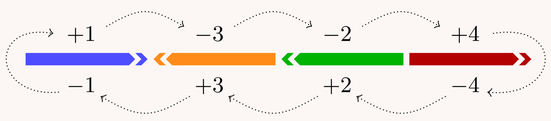

- A **1-cycle** represents a fixed element, and can be ommited in the notation of a permutation.
- The **product** $\pi\sigma$ of permutations $\pi$ and $\sigma$ is obtained by composing both permutations.
- The **identity permutation** $i$ is the permutation where every element is fixed.
- Every permutation $\pi$ has an **inverse** $\pi^{-1}$ such that $\pi\pi^{-1} = \pi^{-1}\pi = i$ .
- The **norm** $\|\pi\|$ of a permutation $\pi$ is a measure of its *rearrangement power*. More formally, it is the minimun number of 2-cycles needed to compose it.

**Example:** $\pi = (1\ 2\ 3\ 4) = (1\ 2)(2\ 3)(3\ 4) \Rightarrow \|\pi\| = 3$

## Modeling Genomes

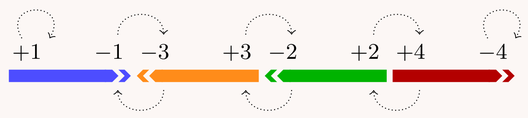
In the Algebraic Theory, **signed permutations** are used to model genomes in two ways:

- **The Chromosomal model:** permutations model the *order* of genes in the **chromosome**. The signs of each integer correspond to the direction of the genes.




$\pi_{\text{chr}} = (+1\ -3\ -2\ +4\ -4\ +2\ +3\ -1)$

- **The Adjacency Model:** permutations model the **adjacencies** between genes. The signs correspond to the **head** $(+)$ and **tail** $(-)$ of each gene.




$\pi_{\text{adj}} = (+1)(-1\ -3)(+3\ -2)(+2\ +4)(-4)$

**Algebraic Property:** The **chromosomal** and **adjacency** models are related by a product with the permutation $\Gamma = (-1\ +1)(-2\ +2)\cdots(-n\ +n)$.

**Example:** $(+1\ -3\ -2\ +4\ -4\ +2\ +3\ -1)\Gamma = (+1)(-1\ -3)(+3\ -2)(+2\ +4)(-4)$

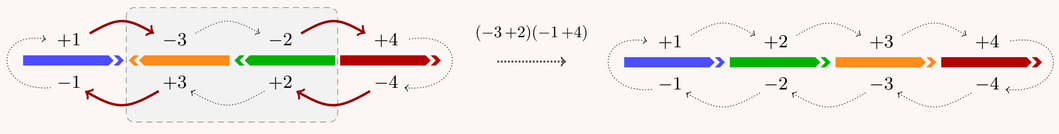
## Modeling Rearrangement Operations

Rearrangement operations are modeled by the **product** $\rho\pi$ of a permutation $\rho$ with a genome $\pi$. Common **2-break** operations (e.g., reversals, translocations) are modeled by the permutation

$$\rho = (u\ v)(\pi\Gamma v\ \pi\Gamma u)$$

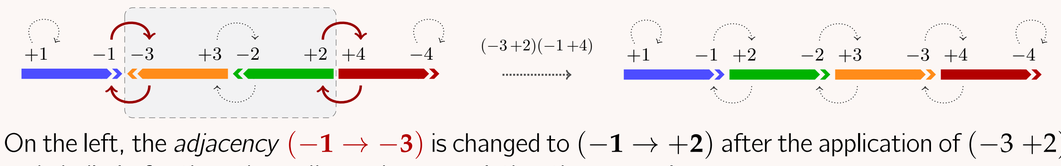
where the choice of elements $u$ and $v$ determine the kind of operation. **For instance:**

A **reversal** on the genome $\pi$ below can be applied with the permutation $\rho = (-3\ +2)(-1\ +4)$ :




On the left, the edge $(+\mathbf{1} \to -\mathbf{3})$ is changed to $(+\mathbf{1} \to +\mathbf{2})$ after the application of $(-3\ +2)$. Similarly, all the red edges are changed by $\rho$ resulting in the reversal.

**Algebraic property:** Any given **rearrangement operation** is modeled by the **same permutation** in both the **chromossomal** and **adjacency** models.




On the left, the *adjacency* $(-\mathbf{1} \to -\mathbf{3})$ is changed to $(-\mathbf{1} \to +\mathbf{2})$ after the application of $(-3\ +2)$, and similarly for the other adjacencies, completing the reversal.

The **weight** of a algebraic rearrangent operation $\rho$ is usually defined as $w(\rho) = \frac{\|\rho\|}{2}$.

This weight attribution agrees with other commonly used models where **2-breaks** have weight 1, and more complex operations such as **transpositions** and **block-interchanges** have weight 2.

## Solving the Algebraic Rearrangent Problem

The **Algebraic Rearrangement Problem** can be defined as:

Given genomes $\pi$ and $\sigma$, find permutations $\rho_1, \rho_2, \ldots, \rho_n$ that **minimally** transform $\pi$ into $\sigma$. Formally:

- $\rho_n \ldots \rho_2\rho_1\pi = \sigma$ and $\rho_i \ldots \rho_2\rho_1\pi$ is a valid genome for every $i = 1, \ldots, n$.
- The **algebraic distance**, defined as $d(\pi, \sigma) = \sum_{i=1}^{n} \frac{\|\rho_i\|}{2}$, is minimum.

Using the fact that $\rho_n \ldots \rho_2\rho_1\pi = \sigma$ we arrive at $\sigma\pi^{-1} = \rho_n \ldots \rho_2\rho_1$. The permutation $\sigma\pi^{-1}$ is very important for solving the rearrangement problem. In fact, we have that [2]:

The **algebraic distance** is given by $d(\pi, \sigma) = \frac{\|\sigma\pi^{-1}\|}{2}$.

This distance can be computed in **linear time**, and rearrangement scenarios between $\pi$ and $\sigma$ can be found by decomposing $\sigma\pi^{-1}$.

**Example:** For the genomes of the last section, $\pi = (+1\ -3\ -2\ +4\ -4\ +2\ +3\ -1)$ and $\sigma = (+1\ +2\ +3\ +4\ -4\ -3\ -2\ -1)$, we have that

$$\sigma\pi^{-1} = (-1\ +4)(-3\ +2) \quad \Rightarrow \quad \|\sigma\pi^{-1}\| = 2 \quad \Rightarrow \quad d(\pi, \sigma) = 1$$

and $\rho = \sigma\pi^{-1} = (-1\ +4)(-3\ +2)$ is exactly the operation that transforms $\pi$ into $\sigma$, as we saw in the previous section.

## Relationship with Double-Cut-and-Join

The **Double Cut and Join** (DCJ) was proposed as an *universal* rearrangement operation [3]. DCJ models several rearrangement operations by applying two cuts in a genome followed by two joins on the same extremities.

Focusing on the genome adjacencies, there are **three types of DCJ operations** [4]:

- Replacing adjacencies $\{p, q\}, \{r, s\}$ with $\{p, s\}, \{q, r\}$.
- Replacing adjacencies $\{p, q\}, \{r\}$ with $\{p, r\}, \{q\}$.
- Replacing adjacencies $\{p\}, \{q\}$ with $\{p, q\}$, and the opposite operation.

All these operations have **weight 1**. In the Algebraic Theory, these operations are modeled as shown in the table below, and we see that there is a **difference in the weight** of **single cut or join** operations.

**Table 1**: DCJ operations modelled with algebraic theory.

| Double-Cut-and-Join | | Algebraic Theory | | |
|---|---|---|---|---|
| **Operation** | **Weight** | **Related Operation** | **Norm** | **Weight** |
| $\{p, q\}, \{r, s\} \to \{p, r\}, \{s, q\}$ | 1 | $\rho = (p\ s)(q\ r)$ | $\|\rho\| = 2$ | 1 |
| $\{p, q\}, \{r\} \to \{p, r\}, \{q\}$ | 1 | $\rho = (p\ q\ r)$ | $\|\rho\| = 2$ | 1 |
| $\{p, q\} \longleftrightarrow \{p\}, \{q\}$ | **1** | $\rho = (p\ q)$ | $\|\rho\| = 1$ | $\frac{1}{2}$ |

This small difference is also visible in the distance equations derived from the **Adjacency Graph**, proposed by Bergeron, Mixtacki and Stoye [4]:

$$d_{DCJ}(\pi, \sigma) = N - (C + I/2) \quad \text{and} \quad d(\pi, \sigma) = N - (C + P/2)$$

where $N$ is the number of genes, $C$ is the number of cycles, , $I$ the number of **odd** paths and $P$ is the **total** number of paths in the adjacency graph.

## Conclusion

With the recent results of Feijão and Meidanis, allowing the modeling multichromosomal linear and circular genomes, the algebraic theory is a interesting alternative for solving rearrangement problems, with a different perspective that might complement the usual combinatorial, graph-theoretical approach. Some of the future directions for investigation include:

- Study other rearrangement problems, such as **median** and **genome halving** under the algebraic theory.
- Expand the algebraic theory to allow **inequal gene content** and **gene duplications** between the genomes.

## References

[1] Meidanis J, Dias Z. An alternative algebraic formalism for genome rearrangements. In: *Comparative Genomics: Empirical and Analyitical Approaches to Gene Order Dynamics, Map Alignment and Evolution of Gene Families*, edited by Sankoff D, Nadeau J, pp. 213–223. Kluwer Academic Publishers. 2000;.

[2] Feijao P, Meidanis J. Extending the Algebraic Formalism for Genome Rearrangements to Include Linear Chromosomes. In: *BSB 2012, LNBI 7409*, edited by de Souto M, Kann M. Heidelberg: Springer-Verlag Berlin. 2012; pp. 13–24.

[3] Yancopoulos S, Attie O, Friedberg R. Efficient sorting of genomic permutations by translocation, inversion and block interchange. *Bioinformatics*. 2005;21(16):3340–3346.

[4] Bergeron A, Mixtacki J, Stoye J. A unifying view of genome rearrangements. *Lecture Notes in Computer Science*. 2006; 4175:163–173.