

# Chromosomal Mutational Algebra I

---

**Dipankar Mazumdar**

Department of Mathematics  
Visva-Bharati (A Central University)  
Santiniketan 731235, India  
dmazumda@science.uva.nl, itsmedip@rediffmail.com

---

**Abstract:** This is a mathematical study of the chromosomal mutations. Chromosomal mutations are variations on chromosome number or chromosome structure. Chromosomal mutation is formulated as algebra. The mutational behaviour is well represented in this algebra. There is a brief description of how to introduce such an algebraic model.

**Keywords:** Chromosomal mutations, Chromosomal mutational algebra, Genome, Bioinformatics, Computational biology, Mathematical modeling.

---

## 1 Introduction

Chromosomal abnormalities lead to the diseases such as cancer. Thus the study of the chromosomal mutations has become very much important. Genetic comparison in molecular biology has undergone a major paradigm shift—a shift from gene comparison based on local mutations (i.e. insertions, deletions and substitutions of nucleotides) to chromosome comparison based on global rearrangements (i.e. inversions and transpositions of chromosomal segments). The genomes are studied by reversals recently [1, 2]. Ferretti et. al. [5] proposed a distance measure in the case where the order of the genes are unknown.

In this paper an algebra have been defined on the mechanisms of chromosomal mutations. This algebra gives a basic mathematical background to the burning issue of chromosomal mutation.

## 2 Chromosomal Mutation

Chromosomal mutations are variations from the wild-type condition in either chromosome number or chromosome structure. Chromosomal mutations can occur spontaneously or they can be induced by chemical or radiation treatment. Chromosomal mutations affect both prokaryotes and eukaryotes, as well as viruses. The association of genetic defects with changes in chromosome structure or chromosome number indicates that not all genetic defects result from simple mutations of single genes. Chromosomal mutations arise spontaneously or can be induced experimentally by chemicals or radiation mutagens, they can be detected by changes in the linkage arrangements of genes. Chromosomal mutations are significant causes of developmental disorders. Chromosomal mutations may involve parts of chromosomes rather than whole chromosomes. The four major types of such mutations are **deletions** (loss of a DNA segment), **duplications** (duplication of a DNA segment), **inversions** (change in orientation of a DNA segment within a chromosome without loss of DNA material) and **translocations** (movement of a DNA segment to another location in the genome without loss of DNA material) [8].

A *deletion* is a chromosomal mutation involving the loss of a segment of a

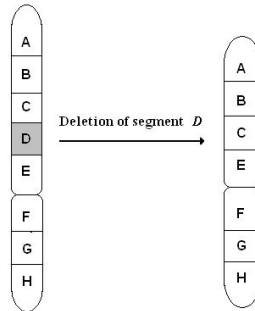


Figure 1: Deletion. An example: A number of human disorders are caused by deletions of chromosome segments. One human disorder caused by deletion is cri-du-chat syndrome, which results from an observable deletion of part of the short arm of chromosome 5. Children with cri-du-chat syndrome are severely mentally retarded, have a number of physical abnormalities and cry with a sound like mew of a cat (hence the name, which is French for ‘cry of the cat’) [Russell, 1998].

chromosome. The deleted segment may be located anywhere along the chromosome. A deletion starts where breaks occur in chromosomes. These breaks may be induced, for example, by agents such as heat, radiation (especially ionizing radiation), viruses, chemicals, transposable elements or by errors in recombination. Because a segment of chromosome is missing, deletion mutations cannot revert to the wild-type state [8].

A *duplication* is a chromosomal mutation that results in the doubling of a segment of a chromosome. Duplication have played an important role in the evolution of gene families. The size of the duplicated segment may vary considerably and duplicated segments may occur at different locations in the genome or in a *tandem* configuration (that is, adjacent to each other). When the order of genes in the duplicated segment is the opposite of the order of the original, it is *reverse tandem duplication*; when the duplicated segments are tandemly arranged at the end of a chromosome, it is a *terminal tandem duplication*.

An *inversion* is a chromosomal mutation that results when a segment of a chromosome is excised and then reintegrated in an orientation of  $180^\circ$  (degrees) from the original orientation. When the inverted segment includes

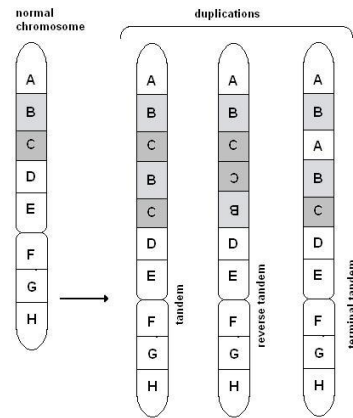


Figure 2: Forms of chromosome duplications.

the centromere, the inversion is called a *pericentric inversion*. When the inverted segment occurs on one chromosome arm and does not include the centromere, the inversion is called a *paracentric inversion*. Genetic material is not lost when an inversion takes place, although there can be phenotypic consequences.

Breaking off a portion of the chromosome and inserting it elsewhere in the chromosome is referred to as a *transposition*.

A *translocation* is a chromosomal mutation in which there is a change in position of chromosome segments and the gene sequences they contain. There is no gain or loss of genetic material involved in a translocation. Two simple kinds of translocations occur. The *intrachromosomal translocation* (within a chromosome) involves a change in position of a chromosome segment within the same chromosome. The other kind *interchromosomal* (between chromosomes) *translocation* involves the transfer of a chromosome segment from one chromosome into a nonhomologous chromosome. If the latter translocation involves the transfer of a segment in one direction from one chromosome to another, it is a *nonreciprocal translocation*. If it involves the exchange of segments between the two chromosomes it is a reciprocal translocation [8].

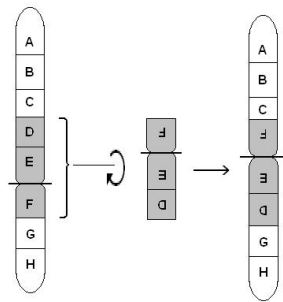


Figure 3: Pericentric inversion. When the inverted segment includes the centromere, the inversion is called a *pericentric inversion*.

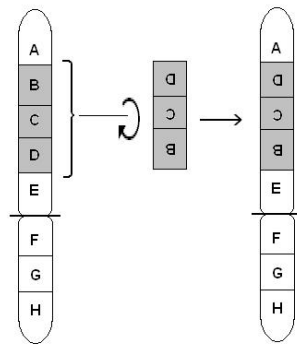


Figure 4: Paracentric inversion. When the inverted segment occurs on one chromosome arm and does not include the centromere, the inversion is called a *paracentric inversion*.

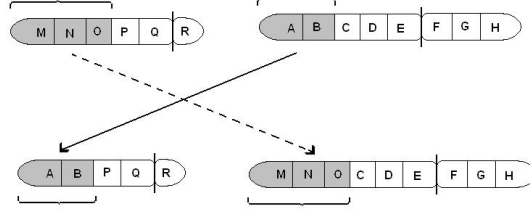


Figure 5: Translocation.

### 3 The construction of Chromosomal Mutational Algebra

#### 3.1 Uni-chromosomal Mutational Algebra

|                             |                                                 |
|-----------------------------|-------------------------------------------------|
| <b>Inversion</b>            | A reversal in the order of a chromosome portion |
| <b>Gene duplication</b>     | The copying of sections of chromosomes          |
| <b>Gene loss / deletion</b> | The deletion of sections of chromosomes         |
| <b>Transposition</b>        | The relocation of sections of chromosome        |

Table 1: The chromosomal mutation includes these in the case of single chromosome genomes.

**Definition 1.**  $S = \langle S^*, \Delta, d, \iota \rangle$  will be called an uni-chromosomal mutational algebra, if  $S = \{x_1, x_2, \dots, x_n\}$  ; is a set of chromosomal segments or portions  $x_1, x_2, \dots, x_n$  and  $S^*$  is the monoid generated by the elements of  $S$  and  $\Delta, d, \iota$  are three unitary operators defined below.

$\Delta$  is a unitary operator and is called the *deletion operator* defined by  $\Delta(x_i) = \Lambda$ ,  $\Lambda$  is the null segment of chromosome.

$d$  is a unitary operator and is called the *duplication operator* defined by  $d(x_i) = x_i \ x_i$ .

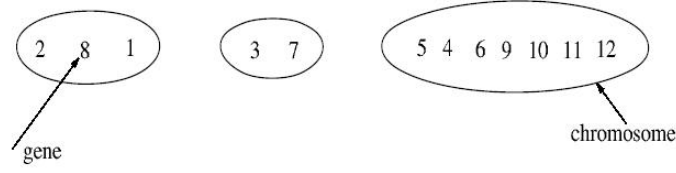


Figure 6: A genome with 12 genes and 3 chromosomes [DasGupta and Wang, 1998].

$\prime$  is a unitary operator and is called the *inversion operator* defined by  $(x_i \ x_j)' = x_j' \ x_i'$ .

All the operators  $\Delta, d, \prime$  can be extended to general cases like,

$$\begin{aligned} \Delta(x_i \ x_{i+1} \ \cdots \ x_k) &= \Lambda \\ d(x_i \ x_{i+1} \ \cdots \ x_k) &= \underbrace{x_i \ x_{i+1} \ \cdots \ x_k}_{\text{prefix}} \underbrace{x_i \ x_{i+1} \ \cdots \ x_k}_{\text{suffix}} \\ (x_i \ x_{i+1} \ \cdots \ x_k)' &= x_k' \ x_{k-1}' \ \cdots \ x_{i+1}' \ x_i' \end{aligned}$$

In this formulation all the uni-chromosomal mutations are defined as operators. *Transposition* may be formulated by the other operators. So this uni-chromosomal mutational algebra is a representation of the biological uni-chromosomal mutation.

### 3.2 Multiple-chromosomal Mutational Algebra

|                      |                                                         |
|----------------------|---------------------------------------------------------|
| <b>Fusion</b>        | Two chromosomes combine and become one chromosome       |
| <b>Fission</b>       | A chromosome splits and becomes two chromosomes         |
| <b>Translocation</b> | The changes of prefixes and suffixes of two chromosomes |

Table 2: The chromosomal mutation includes these in the case of multiple chromosome genomes.

If A, B, C, D are chromosomal segments (any or all of them may be the full chromosome) and  $\Lambda$  is the null segment, then we can represent *translocation*,

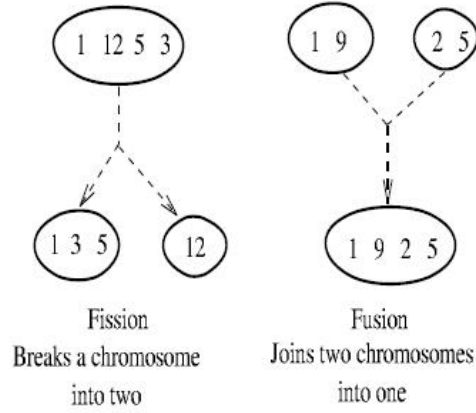


Figure 7: Different mutation operations [DasGupta and Wang, 1998].

*fusion* and *fission* as,

Translocation:  $(A,B)$  and  $(C,D) \rightarrow (A,D)$  and  $(C,B)$   
 Fusion:  $(A,B)$  and  $(C,D) \rightarrow (A,B,C,D)$  and  $(\Lambda, \Lambda)$   
 Fission:  $(A,B)$  and  $(\Lambda, \Lambda) \rightarrow (A, \Lambda)$  and  $(\Lambda, B)$

It is sufficient to model *translocation*, as *fission* and *fusion* are already represented by it. To model the *chromosomal translocation*, two operators are defined. *Translocation operator*  $\mathbf{T}$  and *rearrangement operator*  $\mathbf{R}$ . The chromosome may be represented as a set. The order of the elements of this set is known.

$\mathbf{T}$  is defined as,

$$\mathbf{T}^i(\mathbf{x}_1 \cdots \mathbf{x}_i \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_n) \\ = (\mathbf{x}_1 \cdots \mathbf{x}_{i-1} \mathbf{x}_{i+1} \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_n \mathbf{x}_i)$$

and

$$\mathbf{T}_j(\mathbf{x}_1 \cdots \mathbf{x}_i \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_n) \\ = (\mathbf{x}_1 \cdots \mathbf{x}_i \cdots \mathbf{x}_m \mathbf{y}_j, \mathbf{y}_1 \cdots \mathbf{y}_{j-1} \mathbf{y}_{j+1} \cdots \mathbf{y}_n)$$

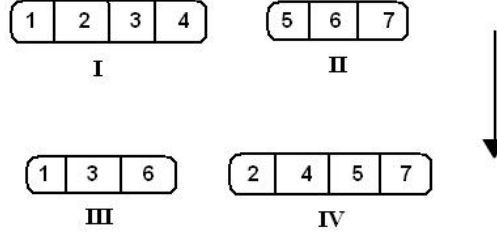


Figure 8: Two chromosomal segments are given by  $(\mathbf{1\ 2\ 3\ 4})$  and  $(\mathbf{5\ 6\ 7})$ , represented by **I** and **II**. Applying the operators **T** and **R** we get the new segments  $(\mathbf{1\ 3\ 6})$  and  $(\mathbf{2\ 4\ 5\ 7})$  represented by **III** and **IV**.

**R** is defined as,

$$\begin{aligned} & \mathbf{R}_{kl}^{ij}(\mathbf{x}_1 \cdots \mathbf{x}_i \cdots \mathbf{x}_j \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_k \cdots \mathbf{y}_l \cdots \mathbf{y}_n) \\ &= (\mathbf{x}_1 \cdots \mathbf{x}_{i-1} \mathbf{x}_j \mathbf{x}_{i+1} \cdots \mathbf{x}_{j-1} \mathbf{x}_i \mathbf{x}_{j+1} \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_{k-1} \mathbf{y}_l \mathbf{y}_{k+1} \cdots \mathbf{y}_{l-1} \mathbf{y}_k \mathbf{y}_{l+1} \cdots \mathbf{y}_n) \end{aligned}$$

In particular,

$$\mathbf{R}^i(\mathbf{x}_1 \cdots \mathbf{x}_i \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_n) = (\mathbf{x}_1 \cdots \mathbf{x}_{i-1} \mathbf{x}_{i+1} \cdots \mathbf{x}_m \mathbf{x}_i, \mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_n)$$

and

$$\mathbf{R}_j(\mathbf{x}_1 \cdots \mathbf{x}_i \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_j \cdots \mathbf{y}_n) = (\mathbf{x}_1 \cdots \mathbf{x}_i \cdots \mathbf{x}_m, \mathbf{y}_1 \cdots \mathbf{y}_{j-1} \mathbf{y}_{j+1} \cdots \mathbf{y}_n \mathbf{y}_j)$$

All  $x_i$ 's and  $y_j$ 's are chromosomal segments. Each chromosome can be represented as an ordered set of  $x_i$ 's. If  $I$  is a chromosome then  $I = \{x_i \ x_{i+1} \ x_{i+2} \ \cdots \ x_{i+p}\}$  where  $i$  is some positive integer and  $p$  is some natural number.

**Example:** The chromosomal segments  $(\mathbf{1\ 2\ 3\ 4})$  and  $(\mathbf{5\ 6\ 7})$  transforms under *translocation* to the new segments  $(\mathbf{1\ 3\ 6})$  and  $(\mathbf{2\ 4\ 5\ 7})$ .

Applying the translocation operator **T** to the chromosomal segments  $(\mathbf{1\ 2\ 3\ 4})$  and  $(\mathbf{5\ 6\ 7})$  we get,

$$\begin{aligned}
\mathbf{T}^2(1\ 2\ 3\ 4, \ 5\ 6\ 7) &= (1\ 3\ 4, \ 5\ 6\ 7\ 2) \\
\mathbf{T}_2(1\ 3\ 4, \ 5\ 6\ 7\ 2) &= (1\ 3\ 4\ 6, \ 5\ 7\ 2) \\
\mathbf{T}^3(1\ 3\ 4\ 6, \ 5\ 7\ 2) &= (1\ 3\ 6, \ 5\ 7\ 2\ 4)
\end{aligned}$$

Then applying the rearrangement operator  $\mathbf{R}$ ,

$$\begin{aligned}
\mathbf{R}_1(1\ 3\ 6, \ 5\ 7\ 2\ 4) &= (1\ 3\ 6, \ 7\ 2\ 4\ 5) \\
\mathbf{R}_1(1\ 3\ 6, \ 7\ 2\ 4\ 5) &= (1\ 3\ 6, \ 2\ 4\ 5\ 7)
\end{aligned}$$

We have obtained the desired segments  $(1\ 3\ 6)$  and  $(2\ 4\ 5\ 7)$ .

Therefore,

$$\mathbf{R}_1\mathbf{R}_1\mathbf{T}^3\mathbf{T}_2\mathbf{T}^2(1\ 2\ 3\ 4, \ 5\ 6\ 7) = (1\ 3\ 6, \ 2\ 4\ 5\ 7)$$

i.e. the problem of the figure [8] is represented by chromosomal operators as

$$\mathbf{R}_1\mathbf{R}_1\mathbf{T}^3\mathbf{T}_2\mathbf{T}^2(\mathbf{I}, \mathbf{II}) = (\mathbf{III}, \mathbf{IV}).$$

## 4 Conclusion

In this work we have developed models of chromosomal mutations. We have introduced the concept of chromosomal mutational algebra. The uni-chromosomal mutations and multiple-chromosomal mutations have been studied. There is a huge scope to explore chromosomal mutations with various mathematical structures. This algebraic structure is capable of solving real life biological problems, though they are not studied extensively in this work. The work leads to a basic formulation of a biological phenomenon- chromosomal mutations, and can be extended further.

## References

- [1] Bafna, V., and Pevzner, P., Sorting by transpositions, Proc. of 6th Ann. ACM-SIAM Symp. on discrete Algorithms, (1995), 614-623.
- [2] Bafna, V., and Pevzner, P., Genome rearrangements and sorting by reversals. SIAM Journal on Computing, (1996), 25(2), 272-289.

- [3] Costello, E., and Wilson, D.C., A case-based approach to gene finding, In: Proceedings of the Fifth International Conference on Case-Based Reasoning Workshop on CBR in the Health Sciences, (1993), 19-28.
- [4] DasGupta, B., and Wang, L., (1998), Selected Topics in Computational Biology. <http://citeseer.nj.nec.com/dasgupta98selected.html>
- [5] Ferretti, V., Nadeau, J.H., and Sankoff, D., Original Synteny, Proc. of 7th Ann. Symp. on Combinatorial Pattern Matching, (1996), 159-167.
- [6] Hannenhalli, S. and Pevzner, P., Transforming Cabbage into Turnip: polynomial algorithm for sorting signed permutations by reversals, Journal of the ACM, (1999), 46(1), 1-27.
- [7] Pevzner, P.A., Computational Molecular Biology: An algorithmic approach, MIT Press, 2000.
- [8] Russell, P.J., Genetics, Benjamin/Cummings, 1998.