

# Stochastic Model for Genetic Recombination because of Crossing over and Chromatid Exchange

Jugal Gogoi

Department of Mathematics  
Dibrugarh University-786004

Tazid Ali

Department of Mathematics  
Dibrugarh University-786004

Rajeev Sarmah

Centre for Bioinformatics Studies  
Dibrugarh University-786004

## ABSTRACT

Genetic recombination by crossing over and chromatid exchange in eukaryotes is one of the major events leading to variance, among individuals of a population set. Therefore, quantification of the recombination would be central to the understanding of the genetic diversity, genealogical differences, disease variants and its maintenance within the population. Hence in this article an attempt is made to model genetic recombination stochastically and prove the Haldane's mapping function of genetic recombination by applying exponential series and Poisson process.

## Keywords

Gene, Crossing-over, Recombination, mutation, stochastic model.

## 1. INTRODUCTION

Genetic recombination by crossing over and chromatid exchange is fundamental in eukaryotes, leading to diversity within a population. While mutation generates new gene variants for natural selection to work on, recombination ensures that new combinations of genes are generated. In eukaryotic diploid organisms crossing over and chromatid exchange during meiosis cell division determines the combination gene arrangement in parental chromosomes, which plays an important role in the appearance of new types in a population [1]. It is thought that the component characteristics from two parents to be extracted and then reassembled in different combinations with expectation of producing an offspring that has the good characteristics of both parents. This can work only if it is possible to identify which parts of each parent should be extracted [2].

There are a number of approaches for detecting recombination in biological sequence, e.g, distance-based [3,4], substitution distribution-based [5], compatibility-based [6,7], and phylogenetic-based[8], but stochastic model is lacking. A stochastic model for recombination may lead to resourceful analysis and realm of the mechanism. Much of the biological researches on recombination are based on experimental data analyzed using certain statistical measures. However, presently mathematical models are widely been used for understanding of biological processes and even in linkage analysis. As the time period for regeneration in higher eukaryotic organisms is much larger, the recombination data by experiments always seems to be too little, thus models would help for understanding such processes. "It turns out to be impossible to carry out what are known as multilocus analysis without models".

A useful source of information regarding historical recombination events can be found in population genetic data, and recent advances in sequencing and genotyping technologies have greatly increased the availability of such

data. Given a sample of population data, patterns of similarities between different sequences can provide information regarding the genealogical history of the sample. However, interpretation of the observed patterns can be problematic without an understanding of the process that generated the patterns in the first place. Statistical modeling of the evolutionary process by which the data was generated can therefore provide a useful tool by which patterns of variation can be understood. Recombination is one of the key evolutionary processes shaping the building of genomes. Quantifying the effect of recombination is fundamental to our understanding of how genetic diversity is generated and maintained in a population, for the design and analysis of studies aimed at uncovering the genetic basis for disease. Ignoring the occurrence of recombination may influence the analysis of genetic data and the conclusions derived from it. In this paper, an attempt is made to model recombination, and derives Haldane's mapping function [9] with the help of Poisson distribution. Haldane (1919) considered a two-strand model where the occurrence of crossover events between two loci are independent of one another. With these assumptions Haldane found the relation:

$$d = \left(\frac{-1}{2}\right) \log(1 - 2r).$$

where,  $r$  is the recombination fraction and  $d$  is the map distance.

## 2. DEFINITION AND ASSUMPTIONS FOR THE MODEL

### 2.1. Counting Process

A counting process [10] is a stochastic process  $N(t), t \geq 0$  with values that are '+ve', integer and increasing;

- $N(t) \geq 0$ .
- $N(t)$  is an integer
- If  $s \leq t$  then  $N(s) \leq N(t)$

If  $s < t$ , then  $N(t) - N(s)$  is the number of events occurred during the interval.

### 2.2. The Poisson Process

A counting process  $N(t), t \geq 0$  is a Poisson process [11] with rate  $\lambda$  if

- $N(0) = 0$ ,
- $N(t)$  has independent increments,
- $N(t) - N(s) \sim \text{Poisson}(\lambda(t - s))$  for  $s < t$

In order to model recombination, we need to model the appropriate stage of meiosis [12] namely reciprocal meiotic exchange between non-sister chromatids. This takes place at the 4-strand stage, although there are models of the relevant features of meiotic products (Fig-1), which do not begin at the 4-strand. It is suitable to consider chromosome arms

separately, as the exchanges on separate arms generally seem to be independent.

A stochastic model for the process of meiosis therefore may consist of two parts:

1. A stochastic process determining the locations of the chiasmata in the four chromatids.
2. A stochastic process indicating for each chiasma which two of the four chromatids take part in the chiasma.

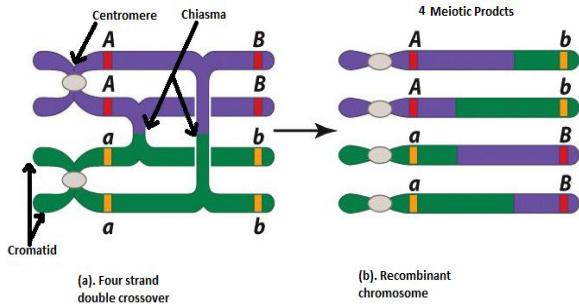


Fig-1

The standard model of recombination uses a Poisson point process, and it is assumed that all strand choice is random, that is, the pair of strands involved in any given exchange in anyone of the four possible non-sister pairs with probability  $\frac{1}{4}$  independently of the choices of strands for all other exchanges, and the position of exchanges. This assumption is known as No Chromatid Interference, abbreviated NCI [13].

### 3. THE MODEL

In a given meiosis cell division, crossover are points, because of which there may be chances of alteration in syteny (arrangement of genes) of a chromosome in the resulting gamete.

A recombination occurs between two loci (positions) on a chromosome if the genes which are copied to the gamete at those two loci derive from the two different parental chromosomes. Thus there is a recombination between two loci on a chromosome if there are an odd number of crossovers between them, because an even number of crossovers (usually 0, maybe 2) would return these genes to their original chromosomal arrangement.

It can be proved that crossovers occur Poisson process as follows:

Let  $k$  be the number of crossover, since recombination depends upon crossing over, therefore we can take recombination as a function of  $k$  (say  $r(k)$ ). Then it follows the following properties of Poisson process:

1.  $r(k = 0) = 0$ , since if there is no crossover then recombination is impossible.
2.  $r(k)$  has independent increments, since the number of crossover that occur in disjoint intervals are independent.
3.  $r(t) - r(k) \sim \text{Poisson}(1(t - k))$  for  $k < t$

The Haldane's model allows no interference at all, i.e., crossing over in disjoint intervals is independent, or the

presence of one crossover does not alter the chance of others occurring nearby. So, the coincidence function has the constant value 1. The unit of genetic distance is Morgan. It is defined as the expected number of crossovers between two loci. Since expectations are additive, so is this measure of distance along a chromosome. [This is always true, even without the assumption that crossovers follow a Poisson process.] If genetic distance is measured in Morgans, the rate of the Poisson process is 1.

Therefore, the probability of an odd number of events in a given interval length  $d$  in a Poisson process rate 1 [9], for that we proceed as follows:

The number of crossovers

$$X \text{ is Prob}(d): P(X = k) = \frac{e^{-d} d^k}{k!}, \dots (1)$$

(by Poisson distribution with parameter  $d$ )

We want to sum

$$P(X = k) \text{ over } k = 1, 3, 5, 7, 9, \dots$$

Now

$$e^d = 1 + d + \frac{d^2}{2!} + \frac{d^3}{3!} + \frac{d^4}{4!} + \dots$$

And

$$e^{-d} = 1 + (-d) + \frac{(-d)^2}{2!} + \frac{(-d)^3}{3!} + \frac{(-d)^4}{4!} + \dots$$

$$= 1 - d + \frac{d^2}{2!} - \frac{d^3}{3!} + \frac{d^4}{4!} + \dots$$

Now

$$e^d - e^{-d} = 2 \left( d + \frac{d^3}{3!} + \frac{d^5}{5!} + \dots \right)$$

$$= 2 \cdot e^d \cdot P(X \text{ odd}) \text{ [from (1)]}$$

therefore,

$$P(\text{recombination}) = P(X \text{ odd}) = \left(\frac{1}{2}\right) (1 - e^{-2d}).$$

or

$$r = \frac{1}{2} (1 - e^{-2d})$$

or

$$e^{-2d} = 1 - 2r$$

or

$$d = (-1/2) \log (1 - 2r)$$

Which is the Haldane's mapping function. This function defines a relation between recombination fraction and crossover.

When  $d=0$ , the probability of recombination is 0. i.e., when loci are very close, probabilities of recombination is 0. When loci are far apart, they segregate almost independently; then the recombination frequency is almost  $1/2$ . i.e., when they are on different chromosomes ( $d$  infinite), they do segregate independently; recombination frequency  $1/2$ . So under Haldane's model, the recombination frequency is always between 0 and  $1/2$ , as also it was seemed in Mather's formula [14].

#### 4. CONCLUSION

In this paper, exponential series as well as Poisson process and distribution are used to prove the Haldane's mapping function  $d = (-1/2)\log(1 - 2r)$  for genetic recombination, which is much easier than the earlier process [15] of proving the same with the help of probability distribution. Here it is also shown that in Mather's Formula and Haldane's map functions both have the same inferences.

#### 5. ACKNOWLEDGEMENT

The research is a part of research project "Mathematical modeling of evolutionary biology" funded by Department of Science & Technology, Govt. of India.

#### 6. REFERENCES

- [1] Hart, D.L. 2000. A Primer of Population Genetics, Sinauer Associates Inc. 3<sup>rd</sup> edition, 96-98.
- [2] Watson, R. A., Pollack, J. B. 2000. Combination and Recombination in a Genetic Algorithm. Dynamical and Evolutionary Machine Organization Group .
- [3] Weiller, G.F. 1998. Phylogenetic profile: a graphical method for detecting genetic recombinations in homologous sequences. *Molecular Biology and Evolution* (15), 326-335.
- [4] Etherington, G.J., Dicks, J. and Roberts, I.N. 2005. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics* (21), 278-281.
- [5] Sawyer. 1989. Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* (6), 526-538.
- [6] Jakobsen, I.B. and Easteal, S. 1996. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. *Computer applications in the Biosciences* (12), 291-295.
- [7] Bruen, T.C., Philippe, H. and Bryant, D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* (172), 2665-2681.
- [8] Hein (1990): Reconstructing evolution of sequences subject to recombination using parsimony. *Mathematical Biosciences* (98), 185-200.
- [9] Haldane, J.B.S .1919. *Journal of Genetics* (8), 299-309.
- [10] Ross, S. M. 1995. *Stochastic Processes*. Wiley. 2<sup>nd</sup> edition.
- [11] Kingman; Charles, John Frank .1992. *Poisson processes*. Vol. 3. Clarendon Press.
- [12] Bernstein, H and Bernstein, C. 2010. Evolutionary origin of recombination during meiosis. *BioScience* 60(7), 498-505.
- [13] Cobbs, G. 1978. Renewal process approach to the theory of Genetic linkage: case of no chromatid interference, *Journal of Theory of Genetic Linkage* . *Genetics* 139.
- [14] Van der Vaart, A.W. 2008. *Statistics in Genetics*, Chapters 1,2,11.
- [15] Liu, B.H . 1997. *Statistica Genomics: Linkage, Mapping, and QTL Analysis*. CRC Press. 18-19.