

## Recombination and Genetic Diversity <sup>1</sup>

T.C. COUTINHO<sup>2</sup>, Campus Alto Paraopeba, UFSJ - Univ. Fed. de São João Del Rei, 36420-000 Ouro Branco, MG, Brazil.

T.T. DA SILVA<sup>3</sup>, Departamento de Física e Matemática, UFSJ - Univ. Fed. de São João Del Rei, 36420-000 Ouro Branco, MG, Brasil.

G.L. TOLEDO<sup>4</sup>, Departamento de Tecnologia em Engenharia Civil, Computação e Humanidades, UFSJ - Univ. Fed. de São João Del Rei, 36420-000 Ouro Branco, MG, Brazil.

**Abstract.** In this paper we present a spatial stochastic model for genetic recombination, that answers if diversity is preserved in an infinite population of recombining individuals distributed spatially. We show that, for finite times, recombination may maintain all the various potential different types, but when time grows infinitely, the diversity of individuals extinguishes off. So under the model premisses, recombination and spatial localization alone are not enough to explain diversity in a population. Further we discuss an application of the model to a controversy regarding the diversity of “Major Histocompatibility Complex” (MHC).

**Keywords.** Genetic recombination, spatial stochastic model, Major Histocompatibility Complex (MHC).

### 1. Introduction

Mendelian laws of inheritance, when applied to infinite populations under random mating, lead to Hardy-Weinberg laws, which state that gene and genotype proportions do not change after the first generation [2]. When considered over finite populations without mutation, random genetic drift leads the population to homozigosity, even in the presence of recombination. Our aim is, then, to investigate how the proportion of different genotypes varies in an infinite population that is distributed spatially, trying to verify the role of recombination in this setting, mainly its implication for population diversity.

In order to build the model, we consider some hypotheses which we explicit in the sequel:

---

<sup>1</sup>This work was partially supported by the Brazilian National Research Council-CNPq, under Grants 314729/2009-7 and 503851/2009-4 and by FAPEMIG, under Grant 253-10 and 04719-10(PRONEM). Part of the paper was presented at CMAC 2011 - Sudeste [5].

<sup>2</sup>tthamaraccoutinho@gmail.com

<sup>3</sup>timoteo@ufs.br

<sup>4</sup>lealtoledo@ufs.br

- i) The population consists of haploid individuals;
- ii) There are an infinite number of individuals, each occupying a position in  $Z$ ;
- iii) A newborn individual is always formed by the contribution of genes from two distinct individuals;
- iv) We do not consider any biochemical or metabolic influence on the genetic inheritance, *i.e.*, mutations do not occur, nor any kind of error during the process of genetic inheritance; besides there are no selective forces acting over the population.

The model recalls the “Voter Model”, a stochastic model originally developed to study the interaction of two distinct populations competing for a territory [3]. Stochastic models treat naturally random fluctuations that usually happen in the environment. In population genetics, *e.g.*, it is natural to assume that allele frequency variation is influenced by probabilistic factors. Then, through the knowledge of the population state in a generation, and given a reproduction scheme for individuals, we can determine the probability of reappearance of a sample of genes in the next generation [6].

The modelling procedure can be described briefly as follows. We dispose each individual in different positions for each time step. An individual’s genes one step ahead are inherited from the recombination of its neighbours’ genes in the current step with equal probability. This originates a stochastic process that will be analysed by a dual process. The building of this dual process allows us to look back on the evolution of the population and retrieve information about which individuals at time step 0 donated the genes that constitute some individual at time step  $n$ . That is, the dual process retrieves the genealogy of genes in the population. We will propose the modelling for 2- and 3-loci individuals, noting that the last gives opportunity for more recombination to occur.

In the next section we propose the models and obtain some conclusions from them. In Section 3. we discuss an application to a controversy regarding the diversity of “Major Histocompatibility Complex” (MHC). MHC molecules play a key role in many immune functions, consequently, these molecules arise special medical interest, since they are directly related to organ and tissue rejection, to pathogenic susceptibility, as well as to individual variability regarding the susceptibility to disorders of self-immune aetiology.

## 2. Mathematical modelling

### 2.1. The 2-loci model

Consider an infinite population consisting of haploid individuals, for which we analyse two distinct loci  $A$  and  $B$ . Each individual is at a point of  $Z$  and each locus admits only two alleles. In the first generation, the individuals at odd points will reproduce, their new genes will be a recombination of their neighbours’ genes, in such a way that, if the individual at position  $i - 1$  is, *e.g.*,  $A_1B_1$  and the individual at position  $i + 1$  is  $A_2B_2$ , then the individual at position  $i$  in the first generation will be either  $A_1B_2$  or  $A_2B_1$  with equal probabilities. In the second generation, the individuals at even points will reproduce by the recombination of their neighbours. And so on. The model building is adapted from the voter model [7].

See the Diagram 1 for an example of how the model evolves.

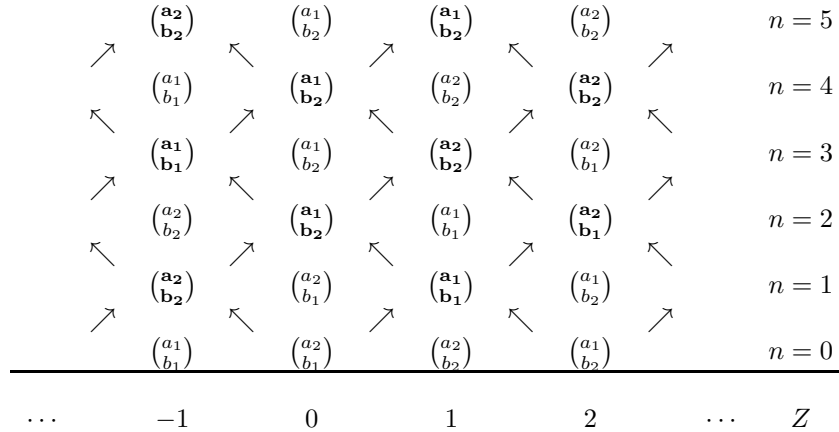


Diagram 1: Voter model modified, adapted to genetics.

Let us develop mathematically the model: Let  $\{V(i, n), i \in Z, n \in \mathbb{N}\}$  be a set of random variables uniformly distributed on the interval  $[0, 1]$ . Define the intervals  $I_1 = [0, 1/2[$  and  $I_2 = [1/2, 1]$ . For each  $n \in \mathbb{N}$  and  $i \in Z$  consider the random vector  $X(i, n) = [x_1(i, n) \ x_2(i, n)]$ , which for  $k = 1, 2$ ,  $x_k(i, n)$  has either the value 0 or 1 (*only two distinct alleles per locum*). So,  $X(n) : Z \rightarrow \{0, 1\}^2$ . We define, then, the dynamics of the model in the following way: if  $i + n$  is odd, then  $X(i, n) = X(i, n-1)$ ; if  $i + n$  is even, then

$$X(i, n) = \sum_{\alpha=1}^2 [x_1(i + (-1)^\alpha, n-1) \ x_2(i - (-1)^\alpha, n-1)] \mathbf{I}_{[V(i, n) \in I_\alpha]}.$$

The initial distribution is given by  $P(X(i, 0) = [a \ b]) = \pi_{ab}$ , for all  $i \in Z$ , with  $\sum_{a, b=0, 1} \pi_{ab} = 1$  and  $\pi_{ab} > 0$ , for  $a, b \in \{0, 1\}$ . The function  $\mathbf{I}_A$  is the characteristic function of the set  $A$ . The initial distribution of  $a$ 's in the first coordinate and  $b$ 's in the second are, respectively,

$$P(x_1(\cdot, 0) = a) = \sum_{j=0}^1 \pi_{aj}, \quad \text{and} \quad P(x_2(\cdot, 0) = b) = \sum_{i=0}^1 \pi_{ib},$$

with  $a, b = 0, 1$ .

### 2.1.1. Dual process and genealogy

Consider, for each  $(i, n) \in Z \times \mathbb{N}$ , the process  $Y^{i, n}(k) = (y_1^{i, n}(k), y_2^{i, n}(k))$ , such that  $Y^{i, n}(k) : Z \rightarrow Z^2$  is given by  $Y^{i, n}(0) = (i, i)$ ;  $Y^{i, n}(1) = (i, i)$ , if  $i + n$  is odd; and, if  $k = 1$  and  $i + n$  is even, or if  $k > 1$

$$Y^{i, n}(k) = Y^{i, n}(k-1) + \sum_{\alpha, \beta=1, 2} (-1)^\alpha (\delta_{\beta 1}, -\delta_{\beta 2}) \mathbf{I}_{[V(y_\beta^{i, n}(k-1), n-k+1) \in I_\alpha]},$$

where  $\delta_{\beta k}$  is equal to 1 if  $\beta = k$ , and equal to zero otherwise. This process represents the genealogy for the individual at position  $i$ , in generation  $n$ .

The following Diagram 2 represents a possible genealogy for an individual at generation 5.

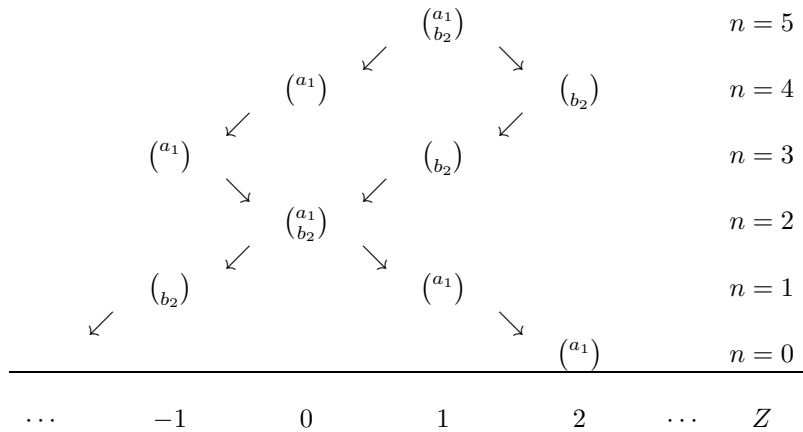


Diagram 2: Genealogy for the individual at position 1, generation 5.

By construction of the process  $Y^{i,n}$  we have the following

**Lemma 2.1** (Duality relation, gene phylogeny). *The following duality identity is valid:*

$$X(i, n) = \left[ x_1(y_1^{i,n}(n), 0) \quad x_2(y_2^{i,n}(n), 0) \right]. \tag{2.1}$$

The genotype of the individual at position  $i$ , generation  $n$ , consists of the gene at the first locus of the individual at the random position  $y_1^{i,n}(n)$  and of the gene at the second locus of the individual at the random position  $y_2^{i,n}(n)$ , both pertaining to the initial generation.

**Theorem 2.1.** *The proportion of genotypes, from the first generation on, keeps constant.*

The proof can be found in Appendix A.

**2.1.2. Diversity loss**

Consider the following equality, whose validity is shown in Appendix B:

$$P(X(i, n) \neq X(j, n)) = \left( 1 - \sum_{a,b=0,1} \pi_{ab}^2 \right) P(Y^{i,n}(n) \neq Y^{j,n}(n)). \tag{2.2}$$

This equality translates mathematically an ancestrality relation between two individuals chosen at random from generation  $n$ . That is, we can infer that the probability of two individuals having distinct genotypes is associated with the probability of their genes having come from distinct ancestors in the initial generation. Applying expression (2.2) and letting  $n$  grows to infinity, we arrive at the following

**Theorem 2.2** (Genetic diversity loss). *The probability of  $X(i, n)$  being different from  $X(j, n)$  goes to zero for large  $n$ . It follows that the genetic diversity does not keep itself on the population.*

*Proof.* We may consider  $y_1^{i,n}$  a symmetric random walk in  $Z$ , without any loss of generality. On the other hand,  $y_2^{i,n}$  walks in  $Z$  independently of  $y_1^{i,n}$ , except when  $y_1^{i,n} = y_2^{i,n}$ , because when they meet each other, if  $y_1^{i,n}(k+1) = y_1^{i,n}(k) + 1$ , then we must have  $y_2^{i,n}(k+1) = y_2^{i,n}(k) - 1$ , but if  $y_1^{i,n}(k+1) = y_1^{i,n}(k) - 1$ , then  $y_2^{i,n}(k+1) = y_2^{i,n}(k) + 1$ . The behaviour of  $y_1^{j,n}$  and  $y_2^{j,n}$  is analogous.

So, with probability one,  $y_1^{i,n}$  will couple with  $y_1^{j,n}$  when  $n$  grows to infinity, since they are unidimensional symmetric recurrent random walks [8]. In the case that  $y_2^{i,n}$  is already equal to  $y_2^{j,n}$ , then there is nothing else to prove. In the case that  $y_2^{i,n}$  is different from  $y_2^{j,n}$ , we can change the point of view and consider that  $y_2^{i,n}$  and  $y_2^{j,n}$  are independent processes from  $y_1 = y_1^{i,n} = y_1^{j,n}$ . Thus  $y_2^{i,n}$  and  $y_2^{j,n}$ , unidimensional symmetric random walks, will couple with each other with probability one when  $n$  increases.  $\square$

## 2.2. The 3-loci model

The model for three loci constitutes an extension of the model for two loci. For 3 distinct loci  $A, B$  and  $C$ , we will have the following recombination possibilities. If the individual at position  $i - 1$  is, for example,  $A_1B_1C_1$  and the individual at position  $i + 1$  is  $A_2B_2C_2$ , then the individual at position  $i$  will be either  $A_1B_2C_2$  or  $A_1B_1C_2$  or  $A_1B_2C_1$  or  $A_2B_1C_2$  or  $A_2B_1C_1$  or  $A_2B_2C_1$  with equal probabilities.

We consider  $\{V(i, n)\}$  and  $\{U(i, n)\}$  two sets of  $[0, 1]$ -uniformly distributed random variables. We define  $J_\beta = \left[\frac{\beta-1}{3}, \frac{\beta}{3}\right]$ ,  $\beta = 1, 2$ , and  $J_3 = \left[\frac{2}{3}, 1\right]$ . For each  $n \in \mathbb{N}$  and  $i \in Z$  let  $X(i, n) = [x_1(i, n) \ x_2(i, n) \ x_3(i, n)]$  be the random vector where, for  $k = 1, 2, 3$ ,  $x_k(i, n)$  takes the values 0 or 1. That is,  $X(n) : Z \rightarrow \{0, 1\}^3$ . The initial distribution is given by  $P(X(i, 0) = [a \ b \ c]) = \pi_{abc}$ . The model dynamics is: if  $i + n$  is odd, then  $X(i, n) = X(i, n - 1)$ ; if  $i + n$  is even, then

$$X(i, n) = \sum_{\alpha=1}^2 \sum_{\beta=1}^3 \begin{bmatrix} x_1(i - (-1)^{\alpha+\delta_{\beta 1}}, n-1) \\ x_2(i - (-1)^{\alpha+\delta_{\beta 2}}, n-1) \\ x_3(i - (-1)^{\alpha+\delta_{\beta 3}}, n-1) \end{bmatrix}^T \mathbf{I}_{[V(i,n) \in I_\alpha]} \mathbf{I}_{[U(i,n) \in J_\beta]}.$$

### 2.2.1. Dual process and genealogy

We also build the dual process for this model. Consider, for each  $(i, n) \in Z \times \mathbb{N}$ , the process  $Y^{i,n}(k) = (y_1^{i,n}(k), y_2^{i,n}(k), y_3^{i,n}(k))$ , such that  $Y^{i,n}(k) : Z \rightarrow Z^3$  is given by  $Y^{i,n}(0) = (i, i, i)$ ;  $Y^{i,n}(1) = (i, i, i)$ , if  $i + n$  is odd; and if  $k = 1$  and  $i + n$  is even, or if  $k > 1$ , then

$$\begin{aligned} Y^{i,n}(k) &= Y^{i,n}(k-1) + \\ &\sum_{\alpha=1}^2 \sum_{\beta=1}^3 \sum_{\gamma=1}^3 (-1)^{\alpha+1} ((-1)^{\delta_{\beta 1}} \delta_{\gamma 1}, (-1)^{\delta_{\beta 2}} \delta_{\gamma 2}, (-1)^{\delta_{\beta 3}} \delta_{\gamma 3}) \\ &\quad \times \mathbf{I}_{[V(Y_\gamma^{i,n}(k-1), n-k+1) \in I_\alpha]} \mathbf{I}_{[U(Y_\gamma^{i,n}(k-1), n-k+1) \in J_\beta]}. \end{aligned}$$

By construction, it follows

**Lemma 2.2** (Duality relation). *The following duality identity is true:*

$$X(i, n) = \left[ x_1(y_1^{i,n}(n), 0) \quad x_2(y_2^{i,n}(n), 0) \quad x_3(y_3^{i,n}(n), 0) \right] \quad (2.3)$$

**Theorem 2.3.** *The proportion of genotypes keeps constant from the first generation on.*

We will skip the proof since it is, *mutatis mutandis*, analogous to the proof of Theorem 2.1.

### 2.2.2. Loss of diversity

The following identity is valid:

$$P(X(i, n) \neq X(j, n)) = \left( 1 - \sum_{a,b,c=0,1} \pi_{abc}^2 \right) P(Y^{i,n}(n) \neq Y^{j,n}(n)). \quad (2.4)$$

The demonstration is analogous to the demonstration of equality (2.2).

**Theorem 2.4.** *The probability of being  $X_n(i)$  different from  $X_n(j)$  goes to zero when  $n$  increases. Therefore it follows that the genetic diversity does not keep itself in the population.*

*Proof.* We may consider, without loss of generality, that  $y_1^{i,n}$  and  $y_2^{i,n}$  are symmetric random walks in  $Z$ , independent of each other. On the other hand,  $y_3^{i,n}$  moves in  $Z$  independently from  $(y_1^{i,n}, y_2^{i,n})$ , except when  $y_1^{i,n} = y_2^{i,n} = y_3^{i,n}$ , because when they meet, we have the following possible implications:

- if  $y_1^{i,n}(k+1) = y_1^{i,n}(k) + 1$  and  $y_2^{i,n}(k+1) = y_2^{i,n}(k) + 1$ , then  $y_3^{i,n}(k+1) = y_3^{i,n}(k) - 1$ ,
- if  $y_1^{i,n}(k+1) = y_1^{i,n}(k) - 1$  and  $y_2^{i,n}(k+1) = y_2^{i,n}(k) - 1$ , then  $y_3^{i,n}(k+1) = y_3^{i,n}(k) + 1$ ,
- otherwise, the movement of  $y_3^{i,n}$  to the left or to the right happens with equal probabilities.

We can analyse analogously the movement of  $y_1^{j,n}$ ,  $y_2^{j,n}$  and  $y_3^{j,n}$ .

So, the process  $y_1^{i,n}$  will almost surely couple with  $y_1^{j,n}$  when  $n$  increases, and in the same way,  $y_2^{i,n}$  will a.s. couple with  $y_2^{j,n}$  when  $n$  increases, since they are unidimensional recurrent symmetric random walks [8].

In the case that  $y_3^{i,n}$  is already equal to  $y_3^{j,n}$ , then the proof ends.

In the case that  $y_3^{i,n}$  is different from  $y_3^{j,n}$ , we can change the point of view and take, for example,  $y_3^{i,n}$ ,  $y_3^{j,n}$  and  $y_1$  as independent from each other, where  $y_1 = y_1^{i,n} = y_1^{j,n}$ ; besides the movement of  $y_2 = y_2^{i,n} = y_2^{j,n}$  will depend on that of  $y_3^{i,n}$  and  $y_3^{j,n}$ , if  $y_3^{i,n} = y_3^{j,n}$ . We conclude, therefore, that when  $n$  goes to infinity, the probability of  $y_3^{i,n}$  and  $y_3^{j,n}$  coupling with themselves goes to one. □

### 2.3. Discussion

Firstly, in each model, we establish a duality relation between the stochastic processes  $X$  and  $Y$ . It follows, by the stochastic process coupling technique, that, in both models, the probability of two individuals being genetically distinct,  $P(X(i, n) \neq X(j, n))$ , goes to zero when  $n$  goes to infinity. That is, the genetic diversity disappears from the population as time goes by.

#### 2.3.1. A comparison between the models

When we augment the number of loci from two to three, the diversity is maintained longer when recombination is present. To illustrate this behaviour see Figure 1 that shows the simulated mean time for  $Y^{0,n}$  and  $Y^{j,n}$  to couple, for various values of  $j$  ( $j = 2, 12, 22, \dots, 102$ ). The more  $j$  is distant from 0, it takes longer, in mean, for  $Y^{0,n}$  and  $Y^{j,n}$  to assume the same value in  $Z^2$  or  $Z^3$ . But the coalescence times in  $Z^3$  are longer than in  $Z^2$ .

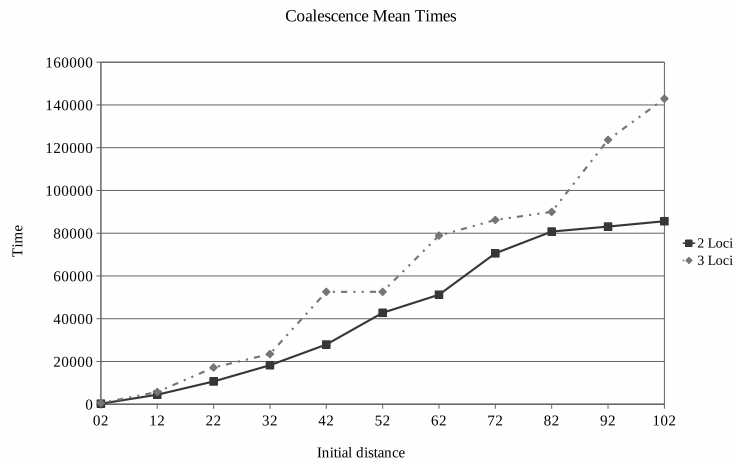


Figure 1: Variation of the coalescence mean time with the initial distance between individuals.

## 3. Application

### 3.1. Recombination and diversity of the MHC

The immune system of an organism is composed of cells and molecules responsible for the defense against infections. Even strange non-infectious substances may

generate immune responses [10]. This is the case about rejection to grafting and to transplantations performed between two people immunologically incompatible.

The role played by the immune system is to exhibit antigens against microorganisms that invade the body to the lymphocytes that eliminate these pathogens. Specialized proteins, the Human Leukocyte Antigens (**HLA**), execute this function; they are codified by a highly polygenic, polymorphic system, called “Major Histocompatibility Complex” (**MHC**).

The term “major histocompatibility complex” derived from researches in which tissues were transplanted between members of the same species. Rejection occurring in many transplantations was thought of being determined by one gene solely, that was called the major histocompatibility gene. Later, it was discovered that this gene was in fact a complex, an ensemble of genes inherited as one that since has been known as the major histocompatibility complex (MHC). Today, it is known that each species has an MHC containing multiple genes.

MHC genes appear in all vertebrates, in humans they are designated **Human Leukocyte Antigens** (HLA), since they were initially detected in leukocytes. The human MHC is codified mainly by a region of the 6th chromosome that contains more than 200 genes [9]. At least six polymorphic genic loci, separated and organized in clusters, were defined in a unique area of the 6th chromosome [4]. They are the most polymorphic of the human genome, having hundreds of stable forms (alleles) for each gene in the population already described. For example, a gene of the human MHC that is polymorphic is HLA-B. Nowadays it has more than 150 alleles described. Nevertheless, this polymorphism is not valid for all MHC genes, some of them have little polymorphism or are monomorphic. Approximately 224 genetic loci were identified englobing 3.5 megabases (Mgb) of DNA in MHC regions. Possibly 180 genes are expressed and around 40% of them have some function in the immune system. This region was one of the first “multimegabase” of the human genome which was completely sequenced [9].

MHC polymorphism is a consequence of vertebrates’ evolutionary response against invasion by microorganisms; thus it reassures the continuity of the species, even in the presence of pandemics. Some individuals may survive a pandemic due to the protective effect of MHC genetic polymorphism. The polymorphisms at the binding region with the antigen determine the specificity of peptide binding. Therefore the MHC molecule binds only with some few peptides among the many at disposal around the cellular micro-environment [9]. Because of polymorphism it is improbable the existence of two individuals that express identical MHC molecules. This huge **diversity** is the main obstacle for organ and tissue transplantation success.

MHC molecules have another essential characteristic: they are polygenic. Being **polygenic** means that these molecules are codified by multiple independent genes. They are inherited in clusters called haplotypes and **expressed co-dominantly** in each individual [10].

### 3.1.1. Controversy

Another important MHC characteristic is **recombination**. Nevertheless the hypothesis that recombination contributes to the diversity of MHC throughout popu-



lations is still disputed, since few comparative researches have computed estimates of this complex recombination rates [11].<sup>5</sup> The spatial stochastic model for recombination presented above shows that *recombination is able to maintain MHC diversity in a population through long time periods*, but when time goes to infinity, diversity goes to zero almost surely.

The model puts in relevance the importance of polymorphism, polygeny and recombination to the diversity of MHC molecules. Other issues such as the MHC codominant pattern or the existence of more than 3 alleles for one locus for most of MHC genes are not considered.<sup>6</sup>

On the other hand, the variability of MHC system evokes a series of questions of scientific interest on its own, related to MHC uncommon polymorphism, natural evolution, biological function of its diverse genes and their actions on the immune system. Due to MHC genetic polymorphism it is improbable to find two individuals that express identical MHC molecules. This such great diversity is the main obstacle to successful organ and tissue transplantations [10]. Nonetheless, according to the conclusions of the mathematical model developed above, this diversity will extinguish off in the long run. Therefore, the observed diversity of MHC molecules is not likely to depend on their high polymorphism, high polygeny, or on the great number of loci involved in recombination; if it is not a transient effect, this diversity may be due to other factors such as mutation.

### 3.2. Other practical issues

Recombination is recognized as an important factor potentially leading to evolution advantage in populations [2], due to its role on the maintenance of population diversity. But recombination solely, in spatially distributed infinite populations, is not able to maintain diversity for longer times, in the context proposed by the models described in this paper, for a finite number of loci. However, further research should be developed in order to put in relevance other characteristics not considered so far, for example, reproduction of diploid individuals, selective pressure, dominance relation between genes, or number of alleles per locum. It is likely that, *e.g.*, increasing the number of possible alleles for each locus, diversity will take longer to disappear from the population.

Another important aspect is the rate of recombination which may not be the same or constant through the population. This is a relevant issue, *e.g.*, for phylogenetic tree estimation. If high rates of recombination are common in MHC genes, re-evaluation of many inference-based phylogenetic analyses of MHC loci, such as estimates of the divergence time of alleles and trans-specific polymorphism, may be required [11].

---

<sup>5</sup>In the absence of recombination, the genes of HLA complex are inherited as an isolated unity of the 6th chromosome, the **haplotype**; the probability of two brothers being HLA-identical is 25%, according to Mendel laws: the child inherits a haplotype from the father and another from the mother.

<sup>6</sup>It is worth noting that the MHC molecule is codified by genes pertaining to 6 loci, each proper subset of them having potential probability of recombination.

## 4. Conclusion

We proposed a mathematical model capable to verify the interference of recombination in the diversity of a spatially distributed infinite population. From the model, we conclude that, as time increases, the probability of taking two distinct individuals with the same genetic load, goes to one. Besides, the greater the number of recombining loci considered, the longer the population diversity is maintained.

When the model was applied to the recombination of MHC molecules, we found that recombination was not a sufficient cause to the maintenance of MHC diversity.

## Appendix

### A Proof of Theorem 2.1

By the duality relation (2.1), for  $n \geq 1$ , we have

$$\begin{aligned}
 P(X(i, n) = [a \ b]) &= P\left([x_1(y_1^{i,n}(n), 0) \ x_2(y_2^{i,n}(n), 0)] = [a \ b]\right) \\
 &= \sum_{r,s \in Z} P([x_1(r, 0) \ x_2(s, 0)] = [a \ b]) P(Y^{i,n}(n) = (r, s)) \\
 &= \sum_{r,s \in Z} \left(\sum_{j=0}^1 \pi_{aj}\right) \left(\sum_{i=0}^1 \pi_{ib}\right) P(Y^{i,n}(n) = (r, s)) \\
 &= \left(\sum_{j=0}^1 \pi_{aj}\right) \left(\sum_{i=0}^1 \pi_{ib}\right) \sum_{r,s \in Z} P(Y^{i,n}(n) = (r, s)) \\
 &= \left(\sum_{j=0}^1 \pi_{aj}\right) \left(\sum_{i=0}^1 \pi_{ib}\right)
 \end{aligned}$$

### B Proof of Equality (2.2)

$$\begin{aligned}
 P(X(i, n) \neq X(j, n)) &= P\left(\left[x_1(y_1^{i,n}(n), 0) \ x_2(y_2^{i,n}(n), 0)\right] \neq \left[x_1(y_1^{j,n}(n), 0) \ x_2(y_2^{j,n}(n), 0)\right]\right) \\
 &= \sum_{r_i \neq r_j \text{ OR } s_i \neq s_j} P\left(\left[x_1(r_i, 0) \ x_2(s_i, 0)\right] \neq \left[x_1(r_j, 0) \ x_2(s_j, 0)\right]\right) \\
 &\quad \times P(Y^{i,n}(n) = (r_i, s_i), Y^{j,n}(n) = (r_j, s_j)) \\
 &= \sum_{r_i \neq r_j \text{ OR } s_i \neq s_j} [1 - P([x_1(r_i, 0) \ x_2(s_i, 0)] = [x_1(r_j, 0) \ x_2(s_j, 0)])] \\
 &\quad \times P(Y^{i,n}(n) = (r_i, s_i), Y^{j,n}(n) = (r_j, s_j)) \\
 &= \left(1 - \sum_{a,b=0}^1 \pi_{ab}^2\right) \sum_{r_i \neq r_j \text{ OR } s_i \neq s_j} P(Y^{i,n}(n) = (r_i, s_i), Y^{j,n}(n) = (r_j, s_j)) \\
 &= \left(1 - \sum_{a,b=0,1} \pi_{ab}^2\right) P(Y^{i,n}(n) \neq Y^{j,n}(n))
 \end{aligned}$$

## References

- [1] A.K. Abbas, A.H. Lichtman, "Imunologia Celular e Molecular", 3a. ed., Elsevier, Rio de Janeiro, 2009.

- [2] R. Bürger, “The mathematical theory of selection, recombination, and mutation”, John Wiley & Sons, Chichester, 2000.
- [3] P. Clifford, A. Sudbury, A model for spatial conflict, *Biometrika Trust*, **60** (1973), 581–588.
- [4] R. Coico, G. Sunshine, “Immunology: a short course”, 6th ed., John Wiley & Sons-Blackwell, New Jersey, 2009.
- [5] T.C. Coutinho, T.T. Da Silva, MHC, Recombinação e Diversidade Genética, *Congresso de Matemática Aplicada e Computacional - Sudeste*, Uberlândia, 2011.
- [6] W. Ewens, G.R. Grant, “Statistical Methods in Bioinformatics: an introduction”, 2nd ed., Springer, New York, 2005.
- [7] P. Ferrari, A. Galves, “Acoplamento e processos estocásticos”. IMPA, Rio de Janeiro, 1997.
- [8] O. Kallenberg, “Foundations of Modern Probability”, 2nd ed., Springer, Berlin, 2002.
- [9] P.S.C. Magalhães, M. Böhlke, F. Neubarth, Complexo Principal de Histocompatibilidade (MHC): codificação genética, bases estruturais e implicações clínicas, *Rev. Med. UCPEL (Pelotas)*, **2** (2004), 54–59.
- [10] A.M. Miranda Vilela, “A diversidade genética do complexo principal de histocompatibilidade (MHC) e sua relação com a susceptibilidade para doenças auto-imunes e câncer.” SBG, Ribeirão Preto, 2007.
- [11] H. Schaschl, P. Wandeler, F. Suchentrunk, G. Obexer-Ruff, S.J. Goodman, Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates, *Heredity*, **97** (2006), 427–437.