Bridging the gap between the evolutionary dynamics and the molecular mechanisms of meiosis : a model based exploration of the *PRDM9* intra-genomic Red Queen

Alice GENESTIER[1], Laurent DURET[1], Nicolas LARTILLOT[1*]

[1] Universite Lyon 1, CNRS, UMR 5558, Laboratoire de Biometrie et Biologie Evolutive, Villeurbanne, France

Corresponding author :

[*] nicolas.lartillot@univ-lyon1.fr

Abstract

Molecular dissection of meiotic recombination in mammals, combined with population-genetic and comparative studies, have revealed a complex evolutionary dynamics characterized by short-lived recombination hot-spots, highly suggestive of an intra-genomic Red Queen. At the center of this exceptional dynamics is a key gene, *Prdm9*. Initially discovered for its role in hybrid sterility, *Prdm9* was subsequently recognized to have a double molecular function during meiosis, first recruiting the double strand break machinery and thus determining the location of hot spots ; second, facilitating the pairing of homologous chromosomes. This latter function requires symmetric binding of *Prdm9* (on both homologues). However, in many respects, the evolutionary mechanisms underlying this complex dynamics are still poorly understood. Here, we present a theoretical model of the evolutionary dynamics of meiotic recombination integrating current knowledge about the underlying molecular mechanisms. Our modeling work gives important insights into the evolutionary dynamics of recombination. Most importantly, a Red Queen spontaneously emerges as an outcome of the interplay between the dual molecular functions of *PRDM9*. Thus, on one side, hot spots are eroded by biased gene conversion, itself a result of the recruitment of the double-strand break machinery by *Prdm9*. On the other side, the reduced symmetrical binding of *Prdm9* caused by the loss of high affinity binding sites induces a net positive selection eliciting new *Prdm9* alleles recognizing new targets. The model also offers new insights about the influence of the genetic dosage of *PRDM9* on the dynamics of the Red Queen, which can paradoxically result in negative selection on new *PRDM9* alleles

entering the population. Finally, calibrating the model based on current empirical knowledge shows there is no need for much more realism in our model to correctly fit the empirical data.

Introduction

In eukaryotes, meiosis is a fundamental step in the reproduction process, allowing the formation of haploid cells from a diploid cell. This process requires, beforehand, the success of a key step, the pairing of homologous chromosomes. This pairing phase allows on the one hand, proper segregation of homologous chromosomes in daughter cells and, on the other hand, the implementation of recombination and genetic shuffling, generating new combinations of alleles. Finally, meiosis and recombination are at the heart of questions of hybrid sterility and speciation [1],[2],[3]. However, the evolutionary dynamics of meiotic recombination still remains poorly understood. A correct understanding of this dynamics requires to articulate the population genetics processes, on one side, and the molecular mechanisms of meiosis, on the other side. In the present work, we present an attempt in this direction, using theoretical and simulation models.

In mammals and many other eukaryotes, recombination points are not uniformly distributed along the chromosomes [4]. Instead, they cluster into regions of the genome, called recombination hotspots [5], of 1 to 2 kb long, often located outside of genes, where the frequency of crossing-over occurrence is more than 50 times that observed in the rest of the genome [6]. More than 30,000 hotspots were found in humans [7],[4], and around 40,000 in mice [8]. These hotspots are characterized in humans by the presence of a sequence motif (13-mer) determining up to 40% of crossing-overs [9]. This motif has itself helped to identify the gene coding for a zinc finger protein responsible for hotspots location, *PRDM9* [9],[10]. Interestingly, *Prdm9* was originally discovered in 2009 [2] for its role in hybrid sterility, suggesting a connection between the process of recombination and speciation in mammals. Once bound to its target site, the *PRDM9* protein induces the recruitment of the double strand break (DSB machinery) near the target site [11]. This produces a single-stranded end that searches for its homologous sequence on the other chromosome and is then repaired, using the sequence at the same locus on the homologous chromosome as the template. This repair leads to the conversion of sequences located near the break (and, in particular, often overlaps the site targeted by *PRDM9* [12]). Some of these gene conversion events lead to the formation of crossing-overs (CO), while others are repaired without exchange of flanking regions (non crossing-overs, NCO).

Crucially, when allelic variation exists at a given target motif, DSBs will form more frequently on the allele for which *PRDM9* has the highest affinity (the ‘hot’ allele). Given that DSBs are repaired by using

the intact homologue as a template, this process leads to the preferential replacement of 'hot' alleles by allelic motifs for which *PRDM9* has a lower affinity. Thus, this mechanism, called DSB-driven biased gene conversion (dBGC), favors the transmission of mutations that inactivate hotspots. At the population scale, dBGC leads to an increase in the frequency of the mutant inactive versions of binding sites, as well as their fixation in the population. This self-destruction phenomenon, commonly called the "hotspot conversion paradox" [13], leads to the progressive erosion of *PRDM9* binding sites at the genome scale, over short evolutionary times (in the order of 10,000 to 100,000 generations [14]), which therefore raises the question of the maintenance of recombination in the long term. As a solution to the paradox, a model of intra-genomic Red Queen has been proposed [15], which works as follows : the disappearance of recombination hotspots compromises the proper functioning of meiosis and greatly reduces the fertility of individuals. In this context, new *PRDM9* alleles recognizing different target sites in the genome and thus restoring recombination, would be positively selected. They would eventually replace the alleles currently segregating in the population. This model predicts a rapid evolutionary dynamics of recombination landscapes, as well as a strong positive selection on the zinc finger domain of *PRMD9*.

Several empirical observations support this model. First, fine scale recombination landscapes differ between closely related species, like between humans and chimpanzees [10],[16], or between modern humans and Denisovan [14], suggesting a fast turnover of recombination hotspots. Second, the zinc finger domain, encoded by a mini-satellite, mutates rapidly [17] thus providing the necessary mutational input for the Red Queen to run. Finally, non-synonymous substitutions in the zinc finger domain of *PRDM9* are more frequent than synonymous substitutions, suggesting the presence of a strong positive selection acting on *Prdm9* DNA binding domain [18],[3].

Analyses combining computer simulations and theoretical analyses under Wright-Fisher models, have also given many arguments in favor of the Red Queen model [15],[19]. However, current theoretical models remain silent on two essential points : first of all, what exact mechanism explains the positive selection acting on *PRDM9* to restore recombination ? Current models invoke a decline in fertility caused by the erosion of recombination, but without providing a precise explanation of this point. Second, these models do not provide any explanation, at this stage, of the link between the intra-genomic Red Queen and the phenotype of hybrid sterility induced by *PRDM9* observed in the mouse [1], [20].

In this respect, a series of molecular and comparative analyses conducted on the mouse provide some explanations. First, the differential erosion of *PRDM9* binding sites between mouse subspecies (due to different *PRDM9* alleles eroding their respective target sites within each subspecies) leads to asymmetrical

binding configurations in hybrid individuals [21]. In a F1 hybrid, each of the two *PRDM9* alleles has eroded its targets in the population from which the transmitting parent originated, but not in the other population. Each allele will therefore tend to bind preferentially to the still active sites present on the opposite chromosome, in *trans* (not eroded), but not in *cis* (eroded). Thus, there will be globally asymmetrical binding patterns of *PRDM9* across the whole genome. In turn, this asymmetry is suspected to be involved in the sterility phenotype. Indeed, Chip-Seq measurement of *PRDM9* binding asymmetry rates in a variety of mouse hybrids for different pairs of *PRDM9* alleles have uncovered a correlation between asymmetry rate and sterility [22],[23]. In addition, cytogenetic observations have shown that more asymmetrically bound chromosomes are less often correctly paired during metaphase I [23]. Finally, the introduction by transgenesis in mice of a *PRDM9* allele possessing the human zinc finger domain, which has never been present in the mouse species complex and has then not eroded its targets asymmetrically in either of the two populations, restores both a good binding symmetry and high levels of fertility.

In the light of these empirical results, a model has been proposed, according to which *PRDM9* would in fact have a dual role during meiosis : in addition to being responsible for the recruitment of the DSB machinery, it would also facilitate the pairing between the homologous chromosomes, thanks to its symmetrical binding. As indicated above, each DSB leads to the formation of a single stranded end, which then has to be repaired either on the homologous chromosome (the most common case and essential for the success of meiosis), or on the sister chromatid. This step of meiosis is critical, since it is what will allow the two homologues to associate, realize a CO and form their synapsis [24]. This step is also rate limiting, since the search for the complementary sequence is carried out over very long DNA sequences. In this context, *PRDM9* would act to bring the sites on which it is bound to the chromosomal axis. This is where symmetry plays a key role : when *PRDM9* is bound symmetrically, the two homologous loci are each brought closer to the chromosomal axis, thus facilitating the search for the complementary sequence by the single stranded end [25]. In contrast, in the case where *PRDM9* binds only on one of the two homologous loci, a possible DSB would be left as it is and repaired by the homologue during the synapse if performed

This mechanistic model based on the symmetrical binding of *PRDM9* provides a globally coherent explanation of the hybrid sterility phenotype. Indeed, in the hybrid, failure of meiosis is a consequence of the sequence asymmetry due to differential erosion in the two parental lineages. But such macroscopic asymmetric sequence patterns are unlikely in a panmictic context. On the other hand, statistical binding asymmetries might nevertheless occur and play a role. Thus a last question remains open however : could this dual role of *PRDM9* also explain the Red Queen evolutionary dynamics of recombination observed

within populations, *i.e.* in a panmictic context ?

To investigate this question, in this paper, we revisit the theoretical modeling of the intra-genomic Red Queen of *PRDM9*-dependent recombination. In contrast to previous work [19], [15] (but see recent work of Baker *et al.* [26]) , we explicitly model the molecular mechanism of meiosis and, more specifically, the function currently hypothesized for *PRDM9*. Our specific aim was to test whether the combined effects of dBGC and symmetry provide sufficient ingredients for running an intra-genomic Red Queen, and this, under empirically reasonable parameters values. These analyses were then summarized and compared with analytical developments to determine the features of the stationary regime. Then, we tested the impact of *PRDM9* genetic dosage on the Red Queen dynamics, observing that genetic dosage tends to act against diversity. Finally, a empirical calibration was attempted in order to explore the predictions of our model and to test the sensitivity of the dynamics of different values of parameters not yet precisely empirically determined.

Material and methods

The model

In this section, the Wright-Fisher model that was implemented for studying the evolutionary dynamics of *PRDM9* is described. This model is summarized in Figure 1, and its key parameters are listed in Table 1.

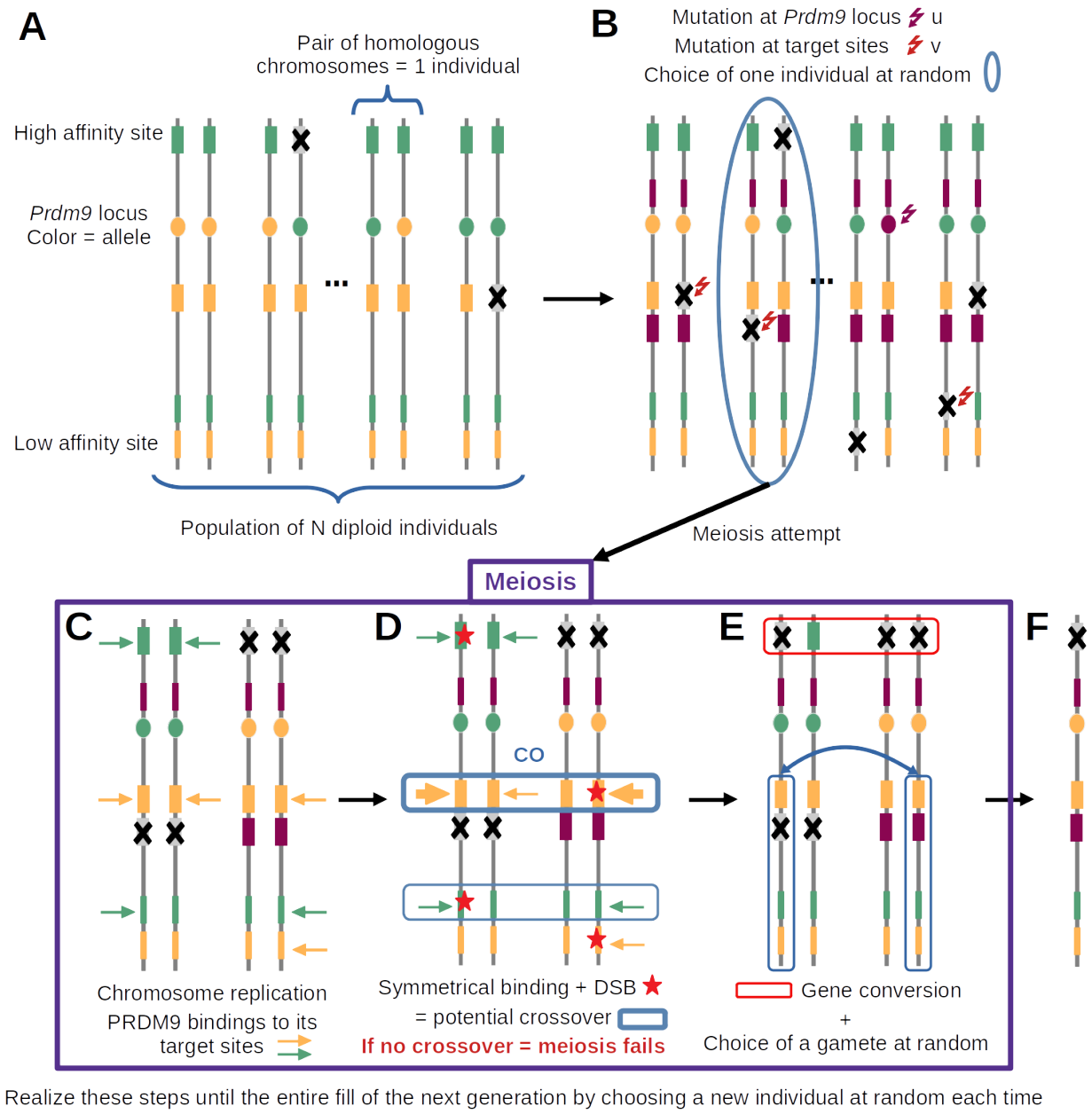


Figure 1. Diagram summarizing the different stages of the model. (A) Population of N individuals (= pairs of homologous chromosomes). Each chromosome has a gene locus *PRDM9* and sets of target sites of variable binding affinity. (B) Mutation at the *PRDM9* locus and target sites. (C) Chromosome replication and binding of *PRDM9* protein to their target sites. (D) Making DSBs and looking for breaks at symmetrically bound sites ; if none is found, meiosis fails and we start again at the stage of choosing the individual in the population (in B). (E) Completion of the crossing-over and (F) random choice of one of the 4 gametes of the tetrad, which will associate with another gamete produced in the same way, to form an individual in the new generation.

This model assumes of a population of N diploid individuals, whose genetic map is composed of a single chromosome. A *Prdm9* locus is located on this chromosome. Each *Prdm9* allele has a set of h binding sites of varying affinity : each site has an intrinsic binding affinity to the *PRDM9* protein associated to it. The positions of these binding sites along the chromosome are drawn uniformly at random and the affinities are drawn according to an exponential law of parameter $\bar{y}$. Each target site is associated to a unique allele and different sites cannot overlap. In practice, this is implemented by encoding the chromosome as an array of

L slots, such that, upon creation of a new *Prdm*9 allele by mutation, each binding site of this allele chooses one of the available slots uniformly at random.

Given the composition of the population of the current generation, the next generation, of the same population size, is generated as follows.

Mutations

The *PRDM9* locus mutates with a probability u per gene copy. Given a mutation, a new functional *PRDM9* allele is created to which are associated h new sites along the genome. These new sites possess different binding affinities, drawn according to an exponential law of mean $\bar{y}$ [20].

Next, each target site recognised by each allele currently present in the population mutates with a probability v. This type of mutation results in a complete inactivation of the target site, which can thus not be recognised anymore. Of note, all target sites are assumed to be present and monomorphic for the active variant at the time of the birth of the corresponding *Prdm*9 allele (*i.e.* we neglect new target sites that might arise by mutation).

Meiosis and reproduction

A meiosis is attempted on a randomly selected individual, according to the following steps. Of note, the meiosis can fail (in which case we assume that the meiocyte undergoes apoptosis) for multiple reasons, all of which will be described below.

In the simulations presented here, unless stated otherwise, we assumed that the reproductive success of an individual is directly proportional to the number of gametes that he produces, *i.e.* that its fitness is determined by the chance of success of each meiosis. Thus, whenever meiosis fails, then a new individual is chosen at random from the population and a new meiosis is attempted. However, in mammals, males generally produce an excess of gametes. It is thus likely that a fraction of meioses can fail without affecting the fitness of individuals. To model this phenomenon, we also performed some simulations where we allow several attempts of meiosis per individual to produce a gamete. The number of attempts (n_{mei}) was varied from $n_{mei} = 1$ (i.e. fitness proportional to meiosis success) to $n_{mei} = 5$ (i.e. a reduction in the number of gametes has no fitness coast up to -80%).

First, the two homologous chromosomes are replicated, thus creating a set of 4 chromatids. Then, *PRDM9* binds to its target site according to an occupation probability determined by the chemical equi-

librium. A binding site i has an affinity

$$K_i = \frac{[PS_i]}{[P]_{free}[S_i]} = \frac{x_i}{[P]_{free}(1-x_i)} \tag{1}$$

where $[PS_i]$ is the proportion of occupied target sites, $[P]_{free}$ is the concentration of free molecules of *PRDM9*, $[S_i]$ is the proportion of free target sites and x_i is the occupancy probability at the site i at equilibrium. Thus

$$x_i = \frac{[P]_{free}K_i}{1+[P]_{free}K_i} \tag{2}$$

We assume that *PRDM9* is not limiting, meaning that most *PRDM9* molecules are free $[P]_{free} \approx [P]_{tot}$; there is therefore a total absence of competition between binding sites. In this model, if we define the rescaled affinity as $y_i = [P]_{tot}K_i$, then the occupancy probability can be re-written :

$$x_i = \frac{y_i}{1+y_i} \tag{3}$$

The total number of sites bound by *PRDM9* is given by $k = \sum_{i=1}^{h} x_i$. If no site is bound in the meiocyte, *i.e.* if the sum of all the $x_i = k$ is lower than 1, meiosis fails. Once the occupation status of all binding sites of the meiocyte has been determined, an average number d of DSBs (DSB) are induced across the chromosome (in the experiments shown here, $d = 6$). More precisely, the total number k of sites bound by *PRDM9* over the four chromatids, is calculated (different from 0), then each site bound by *PRDM*9 undergoes a DSB with a probability equal to $p = \frac{d}{k}$ (in the case where $d > k$, $p = 1$). This procedure aims to model the regulation of the total number of DSBs through the genome, which in mammals seems to be independent from *PRDM9* binding [11], [1]. If, at the end of this step, no DSB has been induced in the genome, meiosis fails. Otherwise, the four chromatids are scanned for symmetrically bound sites, which are essential for crossing-over (CO) achievement. According to the current model of mammalian meiosis, when a site has been broken by the DSB machinery, a single strand is generated. This single stranded DNA will then search for a complementary sequence on which to hybridize, which, if it succeeds, will trigger the pairing of the homologues. In this context, symmetrical binding of *PRDM9* (i.e. binding at the same site on at least one chromatid of each homologue) has been shown to represent a key facilitating step achieving chromosome pairing (see introduction). In the absence of symmetrical binding, chromosome pairing is likely to fail leading to apoptosis of the meiocyte. In the simulator, this is implemented by scanning across the genome for sites that have been symmetrically bound by *PRDM9* and where a DSB has occured. Any

such site becomes a potential position for CO. If no symmetrically bound site has been detected, meiosis fails. Otherwise, if at least one symmetrical binding with a DSB has been detected, one of them is uniformly picked at random and becomes the initiation site for a CO. Only one CO is performed per meiosis (and this, in order to model CO interference). The chosen CO is performed (i.e. the chromosomal segments on either side of the chosen site are exchanged between the two chromatids involved in the CO) and all other sites that have been broken are repaired with the homologous site (NCO), during the synapsis between the two chromosomes. This succesful meiosis therefore produces 4 chromatids, with exactly one CO between two of them, and some events of gene conversion at all sites that have undergone a DSB. Of note, in the presence of inactive versions of the binding sites (created by mutations, see above), these gene conversion events are effectively implementing the hotspot conversion paradox [13]. Finally one chromatid is randomly picked among the tetrad. The whole procedure is repeated, starting from the choice of a new individual until another gamete is obtained. Then these two gametes merge and create a new diploid individual of the next generation.

All these steps, from the choice of an individual until the choice of a chromatid among the tetrad, are performed as many time as needed until the complete fill of the next generation. Thus, at least $2N$ meiosis are needed (two chromosomes for each N individual) if none of them fails. In practice, meiosis can fail, and this, at a frequency that will depend on the genotype of the individual (*PRDM9* alleles, distribution of active or inactive sites, distribution of binding affinity of each site). This model therefore brings out differences in fertility between individuals according to their *PRDM9* genotype and their haplotype (without invoking an arbitrary fitness function) and enables concrete implementation of biased gene conversion. These two points are a fundamental add-on compared to the models implemented in previous studies, which did not explicitly model the molecular mechanisms at the origin of biased gene conversion and selective differences among *Prdm9* alleles [15],[19].

Genetic dosage

In the case where the genetic dosage of *PRDM9* is taken into account, it is assumed that, for a given *PRDM9* allele, a homozygote will produce twice the concentration of the corresponding protein compared to a heterozygote. By assuming that $[P]_{tot}^{homo} = 2[P]_{tot}^{het}$, the probability that a *PRDM9* protein binds to a site i of rescaled affinity y_i is

$$x_i = \frac{cy_i}{1 + cy_i} \tag{4}$$

where $c = 1$ for a heterozygote and $c = 2$ for a homozygote. This implies that the binding probability to a target of affinity y is $x_i^{het} = \frac{y_i}{1+y_i}$ for a heterozygote and $x_i^{hom} = \frac{2y_i}{1+2y_i}$ for a homozygote.

Summary statistics

Several summary statistics were monitored every 100 generations during the run of the simulation program. They were used first, to evaluate the time until convergence to the equilibrium regime (burnin). The burnin was taken in excess, so as to be adapted to all simulation settings. In practice, the burnin is set at 10000 generations over a total of $50,000$ generations. Second, the summary statistics were averaged over the entire run (burnin excluded), thus giving a quantitative characterization of the equilibrium regime, as a function of parameters. These numerical measurements are a key to achieving a better understanding of the dependence of the stationary regime of the Red Queen dynamics on the model parameters.

The main statistics are the following : first, the *PRDM9* diversity in the population (D). Second, for each allele, the mean erosion of its target sites (θ), the mean symmetrical binding probability (q) and the mean fertility (w). Concerning the two statisitcs q and w, the equations presented here correspond to their analytical approximations used to override sampling limitations.

All of these statistics are now described in detail.

Diversity

The diversity is defined as the effective number of *PRDM9* alleles, or in other words as the inverse of the homozygosity [19] :

$$D = \frac{1}{\sum_i f_i^2} \tag{5}$$

where $f_{i,t}$ is the frequency of each allele i.

With this definition, when K alleles segregate each at frequency $\frac{1}{K}$ in the population, the diversity D is equal to K.

Mean activity

The mean activity corresponds to the proportion of sites that are still active for each allele. This fraction is noted $\theta_i \in [0,1]$ for allele i. The mean activity is the mean of the activity of all the current alleles weighted

by their frequency.

$$\theta = \sum_i \theta_i f_i \tag{6}$$

Equivalently, $z_i \approx 1 - \theta_i$ gives a measure of the level of erosion of the target sites of an allele. By extension, when averaged over all alleles in the population, it corresponds to the mean equilibrium erosion level.

Probability of symmetrical binding

The probability of symmetrical binding q corresponds to the mean probability of having a *PRDM9* protein bound on at least one of the two chromatids of the homologous chromosome at a certain position, given that a DSB has occured at this very position. In the case of a heterozygous individual at the *Prdm9* locus, with two alleles of mean erosion level z_1 and z_2, this quantity can be obtained analytically for a given complete diploid genotype and is given by :

$$q^{het}(z_1, z_2) = \frac{2 < x^2 >_{z_1} - < x^3 >_{z_1} + 2 < x^2 >_{z_2} - < x^3 >_{z_2}}{< x >_{z_1} + < x >_{z_2}} \tag{7}$$

and for a homozygous individual with an allele of mean erosion level z :

$$q^{hom}(z) = \frac{2 < x^2 >_z - < x^3 >_z}{< x >_z} \tag{8}$$

In these equations $x = \frac{cy}{1+cy}$ is the mean occupancy of a binding site of affinity y, and $< x >$ is the mean over all sites, taking into account their affinity distribution. This statistic was then averaged over all diploid complete genotypes present in the population at a given generation (and then averaged over all generations of the run).

Fertility

For a given genotype, the fertility can be computed analytically. Here we assumed that fertility in proportional to the rate of success of meiosis. Thus, it is equivalent to 1-(the mean probability of failure of meiosis) characterized by the absence of a DSB in a symmetric site. In turn, the number of DSBs in a symmetric site follows a Poisson law with parameter dq where d is the average number of DSBs per meiocyte and q is the mean probability of symmetrical binding for this allele. Thus, the mean fertility of an allele can

be expressed as follows :

$$w = 1 - e^{-dq} \tag{9}$$

For a given allele, and at a given time, this statistic was then averaged over all diploid complete genotypes carrying this allele present in the population at a given generation (and then averaged over all generations of the run).

Scaling experiments

In order to visualize how the equilibrium regime varies according to the model parameters, first, a central parameter configuration was chosen, which will represent a fixed reference across all scaling experiments.

The central parameters were chosen as follows. First, the parameters of population size N was set to $5,000$, corresponding to the maximum population size that can be afforded computationally. Then the number of bound sites per allele was set to $h = 400$ in order to get closer to the average number of target sites found on the smallest chromosome of the mouse (around 800 sites) while limiting the memory requirements. The parameter d representing the number average of DSBs in a meiocyte is set at 6 which corresponds to the approximate number of DSBs found on the smallest chromosome of the mouse. For uni-dimensional scaling, the parameter v, the mutation rate at target sites, was set to 2×10^{-6} and the parameter u was set to 5×10^{-5}. For the bi-dimensional scaling of d and $\bar{y}$, $u = 2 \times 10^{-4}$ and $v = 5 \times 10^{-5}$. Then only one parameter at a time (or two for bi-dimensional scaling) is systematically varied over a range of regularly spaced values on both sides of the reference configuration. This parameter is fixed along the entire simulation and is variable only between simulations. For each parameter value over this range, a complete simulation is run.

Once the equilibrium is reached for a given simulation setting, the summary statistics described above are averaged over the entire trajectory, excluding the burnin. These mean equilibrium values, which characterize and quantify the stationary regime of the model, were finally plotted as a function of the varying parameter(s).

Parameters	Description	Value
u	Mutation rate at the *Prdm9* locus	2×10^{-6} to 5×10^{-3}
v	Mutation rate at the targets	2×10^{-6} to 5×10^{-3}
N	Population size	$5,000$
h	Number of targets recognised by a new allele	400
d	Mean number of DSB per chromosome pair per meiocyte	6
y_i	Affinity of target i	variable
g	Gene conversion rate	variable
n_{mei}	Number of meiosis allowed per individual before reproduction failure	1 to 5

Variables	Description
f_i	Frequency of allele i [1]
θ_i	Proportion of target sites still active for allele i [1]
z_i	Level of erosion of allele i [1]
w_i	Fertility of allele i [1]
q_i	Probability of symmetrical binding of allele i [1]
ρ	Net rate of erosion per generation
x_i	Occupancy probability at site i at equilibrium
α	Linear response of the log-fitness as a function of erosion
τ	Turnover rate of *PRDM9* alleles
μ	Population-scaled mutation rate of *PRDM9*
D	Diversity D of *PRDM9* in the population at equilibrium
σ_0	Relative difference in fertility between homozygotes and heterozygotes for young alleles
$\sigma_{\bar{z}}$	Relative difference in fertility between homozygotes and hemizygotes for alleles of level of erosion $\bar{z}$

TABLE 1. Descritption of input parameters and output variables. These variables also change over time

Results

Intragenomic Red Queen

We first ran the model with a low mutation rate at the *PRDM9* locus ($u = 5 \times 10^{-6}$). A typical result is shown in figure 2,

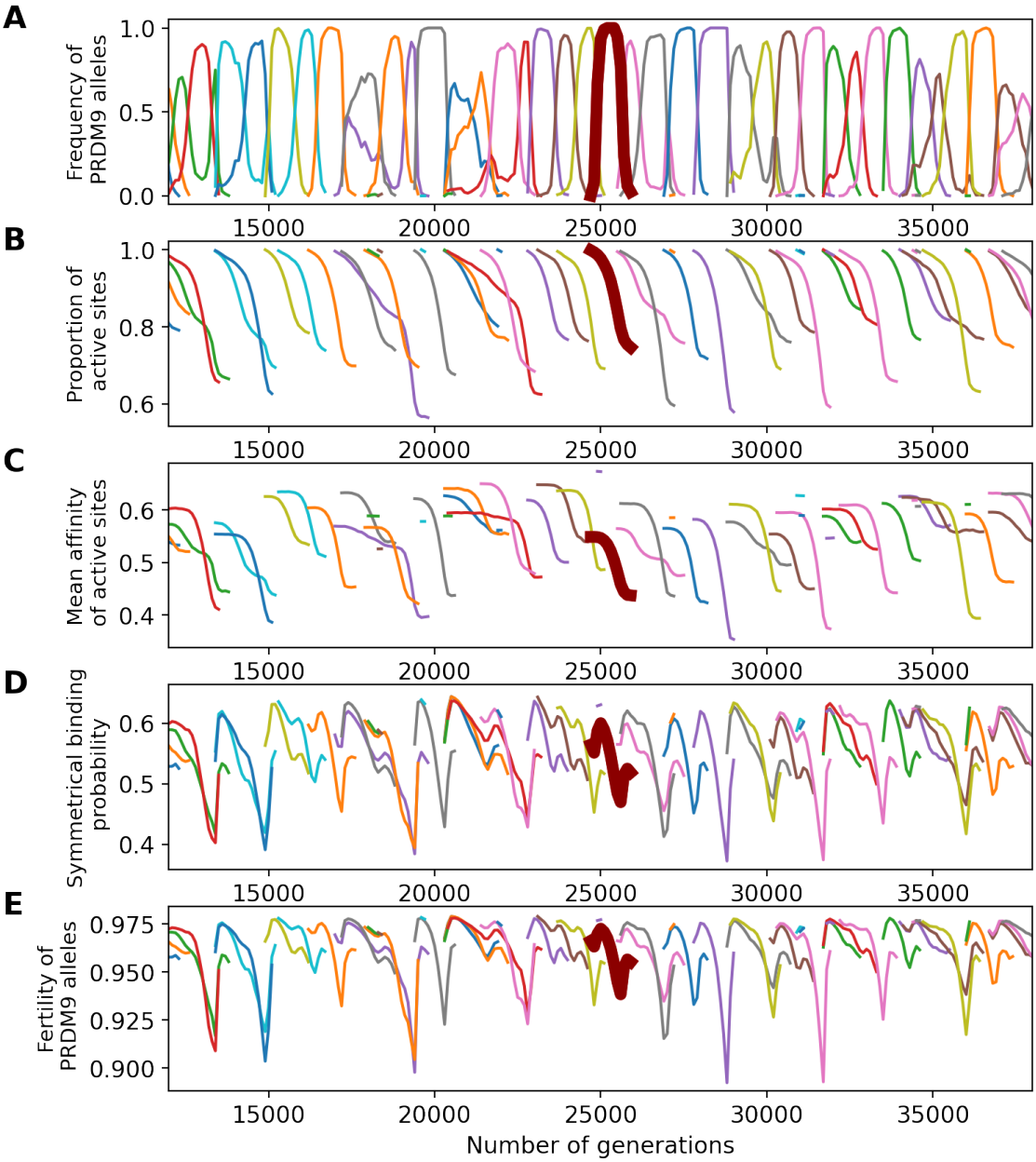


Figure 2. *PRDM9* dynamic : example of a monomorphic regime ($u = 5 \times 10^{-6}$ and $v = 5 \times 10^{-5}$). In all the graphs, each color corresponds to a different allele. Note that a given color can be reasigned to a new allele later in the simulation. This figure represents the evolution along time of (A) the frequency of each *PRDM9* allele and their corresponding (B) proportion of active sites, (C) mean affinity, (D) probability of symmetrical binding and (E) fertility. The bold trajectory represent a typical allele trajectory.

Figure 2A, representing the evolution along time of the frequency of each *PRDM9* allele in the population ($f_{i,t}$), shows a regime with few alleles segregating at any given time in the population, due to a low *PRDM9* mutation rate ($u = 5 \times 10^{-6}$). We call this a *monomorphic* regime, after Latrille *et al* [19]. The

dynamic is as follows. First, an allele appears in the population and invades, until reaching a frequency close to 1. Meanwhile, the activity of this allele θ decreases, due to the erosion of its target sites caused by inactivating mutations and biased gene conversion. Once the allele has eroded a fraction of around 20% of its sites, it is quickly replaced by one of the low-frequency alleles that keep being created by mutations at the *PRDM9* locus. This rapid replacement clearly suggests the presence of strong positive selection. This positive selection is entirely a consequence of the molecular mechanisms of meiosis such as implemented by the simulation model (the process implicated will be described below). The newly invading allele then erodes its target until being replaced by a new allele which in turn follows a similar trajectory, and so on. This cyclic process clearly shows that a Red-Queen process is operating, under the combined action of erosion of the *PRDM9* target sites by inactivating mutations and biased gene conversion, and positive selection on new *PRDM9* alleles.

A key aspect of the model is that it does not explicitly invoke a fitness function. Instead, the positive selection that seems to be operating on new alleles (Figure 2) is an emerging property of the mechanism of meiosis. This positive selection acting on new alleles can be more precisely understood by examining how the rate of *PRDM9* symmetrical binding (7),(8) and the mean fertility of a typical allele (9) evolve over the lifespan of this allele (Figure 2D and 2E), and how this relates to the level of erosion (6) (Figure 2B) and the mean affinity of the remaining non-eroded sites (Figure 2C). First, we observe a clear correlation between allele frequency and activity. When the frequency of an allele increases in the population, its general activity decreases, meaning that the proportion of still active sites for this allele decreases. This erosion (Figure 2B) seems to occur at a rate which is directly proportional to the frequency of the allele in the population (Figure 2A). This is expected, the more frequent an allele, the more often it will be implicated in a meiosis and the more opportunities its bounding sites will have to undergo gene conversion events. Second, erosion results in a decrease of the proportion of active sites for a given allele. But, more importantly, it results in a decrease of the mean affinity of the sites that are still active (Figure 2C). This reflects the fact that sites of high affinity are eroded more rapidly. This last point is also expected, since the sites of high affinity are more often bound by PRDM9 (the binding probability $= \frac{cy}{1+cy}$ is higher when the affinity y is higher), and thus are more often converted by inactive mutant versions. Third, sites of lower affinity are also less often symmetrically bound (i.e. bound simultaneously on both homologues). Davies *et al.* discovered in 2016 that the symmetrical binding of PRDM9 greatly facilitates the repair of DSBs generating recombination events that can result in crossovers essential for chromosome pairing and therefore for the success of meiosis [1]. One of the key features that captures the success of meiosis is the conditional rate of symmetrical binding (q). Since DSBs

are chosen uniformly at random among all bound sites, q corresponds to the probability that a DSB will occur in a symmetrically bound site and will thus contribute to a successful pairing of the homologues. This probability is a monotonous function of affinity (Figure 2D). The mean value of q over all sites of a given allele is lower for older alleles (Figure 2D). Finally, since, in our model, meiosis requires at least one DSB in a symmetrically bound site, the mean fertility of an older allele is lower (Figure 2E). Hence, new alleles (young alleles) will be positively selected at the expense of old alleles and will ultimately replace them in the population.

Thus, a key result obtained here is that the requirement of symmetrical binding for chromosome pairing, combined with preferential erosion of high affinity sites, is sufficient for creating differences in fitness (fertility) between old and young alleles, which in turn will spontaneously induce a Red Queen dynamics at the level of the population.

It should be noted that we observe a slight increase in the rate of symmetrical binding and in the mean fertility at the beginning and at the end of the life of the alleles. The reason of this is that these two summary statistics are defined, for each allele, as a mean over all diploid genotypes segregating in the population at any given time. As a result, when old alleles have declined to a low frequency, they often find themselves in a heterozygous state with new alleles, which restores the rate of symmetrical binding and thus the fertility of the corresponding diploid individual. Likewise, when a new allele appears in the population it is in a heterozygous state with an older allele, which gives a lower rate of symmetrical binding and fertility than being in the homozygous state.

The simulation shown above (Figure 2) was run under a low mutation rate, hence resulting in a *monomorphic* regime. Running the simulation under higher mutation rates for PRDM9 (higher u) results in a *polymorphic* regime, where many alleles segregate together at any given time. In this regime, of which an example of simulation is shown in figure 3, the Red Queen process is also operating, except that many alleles are simultaneously segregating (3A), at a lower frequency (around $\frac{1}{D}$, where D is the diversity of effective number of allele). As in the previously shown monomorphic regime, each allele undergoes erosion (3B), primarily of its higher affinity sites (3C) ; again causing a decrease in symmetrical binding rate (3D) and fertility (3E). But, due to the high mutation rate, there is a much lower erosion leading to a lower decrease in fertility than in a monomorphic regime. Note that in figure 3, the scale of the axes of the ordinates are not the same as in figure 2.

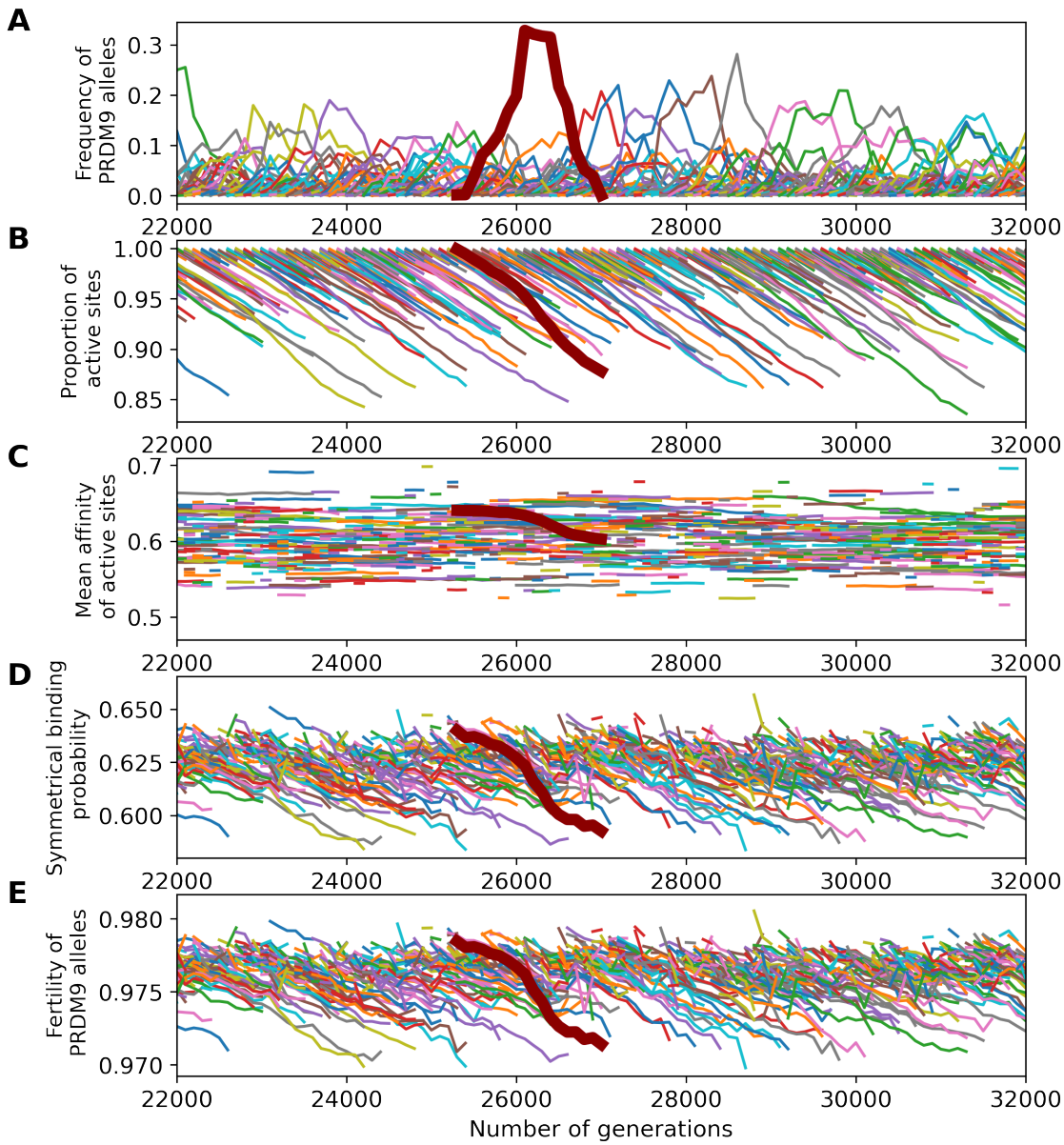


Figure 3. *PRDM9* dynamic : example of a polymorphic regime ($u = 5 \times 10^{-4}$ and $v = 5 \times 10^{-5}$). In all the graphs, each color corresponds to a different allele. Note that a given color can be reassigned to a new allele later in the simulation. This figure represents the evolution along time of (A) the frequency of each *PRDM9* allele and their corresponding (B) proportion of active sites, (C) mean affinity, (D) probability of symmetrical binding and (E) fertility. The bold trajectory represent a typical allele trajectory that escaped from genetic drift. Note that the scale of the axes of the ordinates are not the same as in figure 2

PRDM9 diversity and activity

One-dimensional scaling

The simulations shown in the previous section, have helped to establish that the mechanistic model induces a Red Queen dynamics, and to explain the positive selection on new alleles. However, it remains to be understood how the equilibrium regime quantitatively depends on the parameters of the model. To

address this issue, scaling experiments were performed (see methods). These experiments are helpful for determining how the different characteristics of the model at equilibrium (the allelic diversity at the *PRDM9* locus, the average level of erosion of the target sites or the fertility of the individuals) respond to variation in the values of the model parameters.

A first scaling experiment was carried out on the mutation rate u at the *Prdm9* locus. It immediately appears in Figure 4 A that u has a direct impact on the standing diversity, which is roughly proportional to $24Nu$, the population scaled mutation rate. Increasing u also reduces the equilibrium erosion level (Figure 4 B). This can be explained as follows. The equilibrium set point of the Red Queen is fundamentally the result of a balance between erosion and invasion. The rate of invasion by new alleles can be expressed as $2NuP_{inv}$ where P_{inv} corresponds to the probability of invasion. In this expression, if u increases, this also increases the rate of replacement of old alleles thus shifting the equilibrium towards lower equilibrium erosion level. These lower erosion levels imply a higher rate of symmetrical binding and a higher mean fertility (4C).

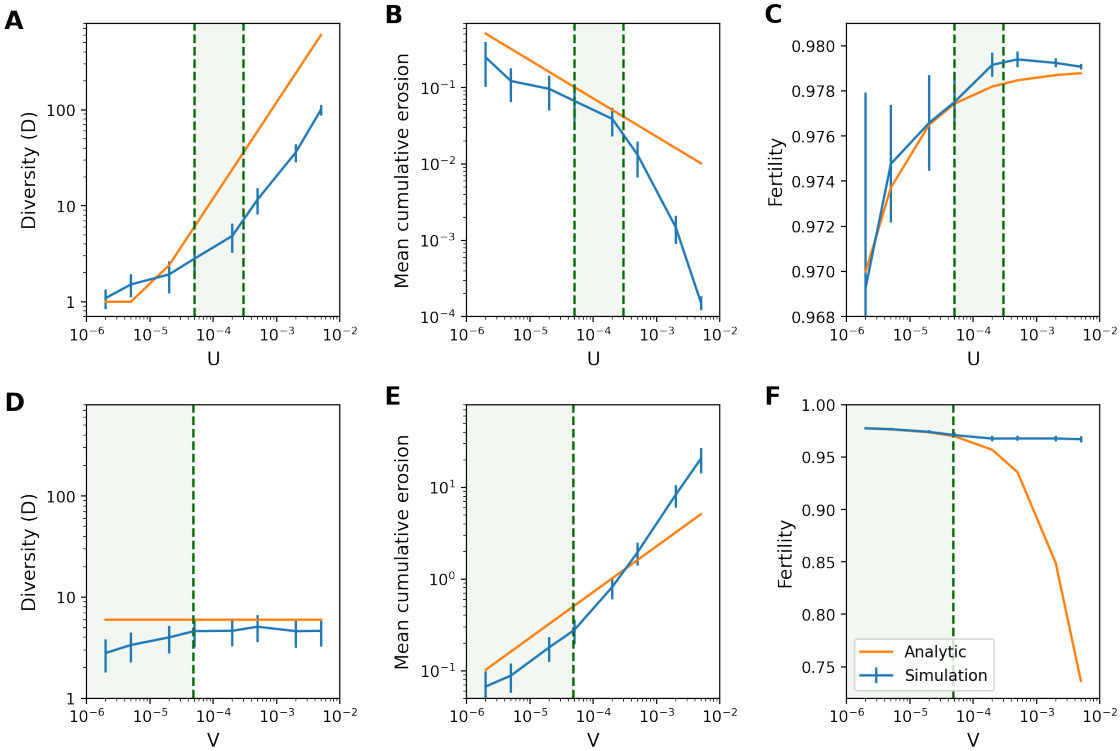


Figure 4. Summary statistics as a function of the *Prdm9* mutation rate u ($v = 2 \times 10^{-6}$) and the target mutation rate v ($u = 5 \times 10^{-5}$). Study of the Diversity (D) as a function of u (A) and v (D), the mean erosion at equilibrium as a function of u (B) and v (E) and the log fitness as a function of u (C) and v (F). Each graph represent the computer simulations with the standard variation for each run (blue) and the analytical approximations (orange). The area colored in green correspond to the area where the analytical model verifies the assumptions of high diversity, low erosion and strong selection.

A second scaling experiment was performed, now varying the rate v of target inactivation. From this experiment, it appears that v has a very weak impact on standing diversity (Figure 4 D). On the other hand, it has an important impact on equilibrium erosion levels (Figure 4 E) and on fertility 4 F), with an opposite

effect, compared to what was observed with u. In this case, the explanation comes from the fact that the rate of erosion is equal to $2Nvg$ where g is the proportion of sites that will be converted by the homologous sites among all inactivated sites at a population-wide rate of $2Nv$ (see supplementary materials equations (34)). Increasing v will therefore result in an increase of the rate of erosion at equilibrium, and correlatively, a decrease on the probability of symmetrical binding and of the mean fertility, because fewer high affinity sites still recognizable by *PRDM9* implies less chance of binding to the same site on the homologous chromosomes.

To summarize, the equilibrium set point of the Red Queen dynamic is directly impacted by the two parameters u and v, whose effects are opposed. In the case of diversity, u (or, more precisely, Nu) is the key determining factor, while v seems to have minor impact.

Comparison with analytical developments

In parallel with the simulator, the model was also explored analytically. The complete analytical developments can be found in Supplementary materials parts 1 to 3. The mathematical developments are based on the population genetics principles and using a self-consistent mean-field approaches described in Latrille *et al.* [19], although now it also takes into account the dual role of *PRDM9* during meiosis. Such analytical developments are useful in several respects. First, they give a deeper insight into the relationship between input and output parameters of the model. Second, they make it possible to overcome the limitations of simulation experiments in terms of computing and storage capacity, allowing, among other things, to extrapolate results to larger population size or larger genomes. Here, a simplified presentation is given, introducing the main formulae.

For these analytical developments to be tractable, three preliminary conditions were introduced. First, a polymorphic regime is assumed for *PRDM9*. This assumption makes it possible to neglect the presence and the impact on the dynamics of homozygous individuals, which facilitate the approximations. Second, we assume a strong selection regime. This allows us to rely on deterministic approximations for the typical trajectory of *Prdm9* alleles. Finally, we assume a weak mean erosion of the target sites at equilibrium, which allows us to linearize the equation.

First, to a good approximate, the rate of erosion of the targets of an allele depends on its frequency in the population :

$$\frac{d\theta}{dt} \approx -\rho f \theta \tag{10}$$

where $\rho = \frac{Nvd}{2h} \approx \frac{Nvg}{2h}$, with g as the gene conversion rate, is the net rate of erosion per generation. As a

result, the cumulated erosion of an allele of age t is :

$$z(t) \approx \rho \int_0^t f(X)\, dX \tag{11}$$

where $f(X)$ is the frequency of the allele at time X. z can thus be seen as the intrinsic age of the allele (i.e. an allele ages more rapidly if more frequent in the population). Because of selection, the frequency of an allele changes as a function of its erosion, or intrinsic age z.

Assuming weak erosion ($z << 1$) allows one to linearize the differential equation describing the evolution of the frequency, which gives :

$$\frac{dln(f)}{dt} \approx -\frac{\alpha}{2}(z - \bar{z}) \tag{12}$$

Equation 12 says that the variation in frequency of an allele is proportional to the difference of erosion between this allele (z) and the mean erosion of all other alleles at equilibrium ($\bar{z}$. The proportionality coefficient, α, corresponds to the linear response of the log-fitness as a function of erosion. It can be expressed as a function of the mechanistic parameters of the model, see supplementary materials equation (56). Thus, if an allele is younger, its erosion is lower than the mean erosion at equilibrium, the allele have eroded less target than the average and will increase in frequency. But when the allele reaches the mean erosion level and overtake it, this implies that the allele has a level of erosion higher than the average and will decrease in frequency in the population.

The previous equation shows the evolution of the frequency of a typical allele in the population, but it depends on the mean level of erosion $\bar{z}$ which is unknown. To determine its value, we can apply a self-consistent argument, essentially saying that all alleles entering the population have this typical dynamic. Using this self-consistent argument leads to the following explicit expression for $\bar{z}$ [19] :

$$\bar{z} \approx \sqrt{\frac{\rho}{\mu\alpha}} \approx \sqrt{\frac{vg}{u\alpha}} \tag{13}$$

Here, we find 4 constants of the model. There is first of all ρ, the net erosion constant, which by increasing will contribute to increasing the mean equilibrium erosion level $\bar{z}$. Then, in the denominator, there is $\mu = 4Nu$, the population-scaled mutation rate of *PRDM9*, which by increasing will allow a more frequent emergence of new alleles. This restricts the lifetime of alleles in the population and thus decreases the average cumulative erosion rate. Finally, we find again the constant α. If the latter increases, this means that the log-fitness will decrease more quickly as a function of erosion. As a result, the allele will be more

quickly eliminated (i.e. with less erosion), which lowers the equilibrium erosion level $\bar{z}$.

Note that this equation is the same as that presented on Latrille *et al*. However, in their paper, α and g were phenomenological parameters, set arbitrarily, while in our mechanistic model, they emerge directly from the molecular mechanisms of meiosis. As such, they can be expressed as functions of the model parameters (d, h, N or v, see supplementary materials equations (56) and (34)).This gives a better understanding of the mechanisms and the dependence between the outputs of the model and the input parameters.

From there, we can express all the mean equilibrium quantities over the population and across the whole simulation as a function of $\bar{z}$. Namely $\bar{\theta}$, the mean activity of the target sites, $\bar{w}$, the mean fertility of the alleles, and the diversity D of *PRDM9* in the population at the equilibrium :

$$\begin{cases} \bar{\theta} & \approx & 1-\bar{z} \\ \bar{w} & \approx & 1-e^{-d\bar{q}} \\ D & \approx & 24Nu \end{cases} \tag{14}$$

In the expression of $\bar{w}$ we can find the parameter $\bar{q}$ corresponding to the mean probability of symmetrical binding at the equilibrium. This parameter depends itself on $\bar{z}$.

These analytical developments have therefore made it possible to obtain direct approximate relationships between important parameters in the Red Queen mechanism such as the link between the erosion ρ and the restoration $\mu\alpha$ or the expression of the influence of the mechanism on the selective response to the erosion α.

The analytical approximations presented here were confronted with the results obtained directly using the simulation program (blue curve on Figures 4 A to F).

The model and the analytical approximation give qualitatively similar results in the range of parameters validating all the conditions, in particular for fertility (Figure4 C and Figure 4 F). Concerning the *PRDM9* diversity, we can obtain fairly substantial differences between the simulation results and the analytical approximations, up to a factor of 10, in the scaling of u (Figure 4 A). But the nature of the regime, polymorphic (many alleles segregating at the same time in the population) or monomorphic (only one allele present in the population at a time), is correctly predicted. In particular, we can say that the nature of the regime is directly and mostly determined by 24Nu and the level of erosion has almost no influence (Figure 4 D). Finally, the analytical approximations are less accurate for low and high u or high v. These correspond to strong erosion

regimes (low u and high v) or to regimes with weak selection (high u), for which the assumption of the analytical developments are not met.

Taking into account the *PRDM9* genetic dosage

The previous results were obtained with a model assuming the same concentration of the *PRDM9* product of a given allele in individuals that are either homozygous or heterozygous for this allele. Yet, in reality, gene dosage seems to be an important aspect of the regulation of PRDM9 expression [27]. To account for this fact, genetic dosage was introduced in the simulation model. This was done by assuming that a homozygote produces twice as many protein products as a heterozygote for a given allele. Analytically this can be expressed as follows. Without dosage the *PRDM9* binding probability x of a target site of affinity y is the same whatever the genotype (*i.e.* the number of *PRDM9* copies) and is equal to $x = \frac{y}{1+y}$. However, when the genetic dosage is taken into account, this probability becomes $x = \frac{2y}{1+2y}$ for a homozygote. As a result, in homozygotes, the occupancy of a site for a given affinity is increased, leading to a higher probability of symmetrical binding and fertility which could have an important impact on the Red Queen dynamics.

To illustrate this point, Figure 5 contrasts the results obtained with and without genetic dosage in an otherwise identical parameter configuration. We observe a drastic change of regime between the two settings. When the simulation without genetic dosage gives a polymorphic regime, the simulation with genetic dosage gives a strict monomorphic regime with an extremely dominant allele staying at a frequency close to 1 during several thousands of generations, a time during which new alleles apparently cannot invade the population. What happens here is that, due to gene dosage, homozygotes have a fitness advantage over heterozygotes. Given that, alleles at high frequency are more often present in a homozygous diploid genotype, they thus have an advantage over low frequency alleles. And so, paradoxically, old (but not too old) alleles at high frequency can have a higher mean fitness than new alleles that just appeared in the population, and this, in spite of their higher levels of erosion. This last point is confirmed by measuring the selection coefficient associated with new alleles (Figure 5C) : during a phase of domination of one allele, new alleles are strongly counter selected. When the dominant allele becomes too old, its homozygote advantage is no longer strong enough to compensate for its erosion. At this point, new alleles become positively selected (e.g. at around $15,000$ generations on figure 5C). The old allele then quickly disappears and all other alleles competes for invasion. This transient polymorphic regime is unstable however : as soon as one of the young alleles reaches a higher frequency than its competitors, it benefits from a homozogous advantage, ultimately evicting all competitors and thus becoming the new dominant allele.

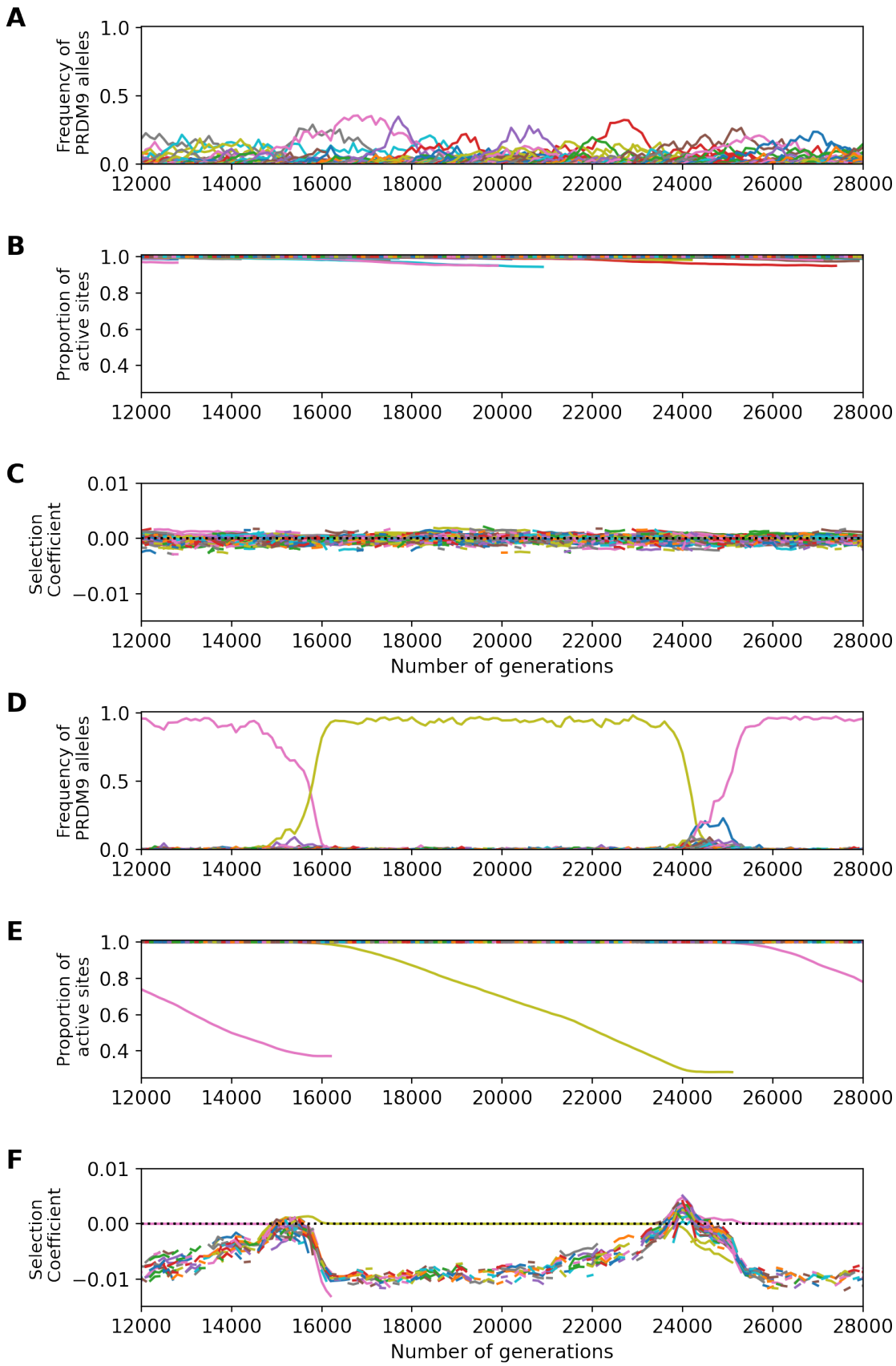


Figure 5. *PRDM9* dynamic : example of a *PRDM9* dynamic without *PRDM9* genetic dosage (A) to (C) and with genetic dosage (D) to (F) ($u = 5 \times 10^{-4}$ and $v = 2 \times 10^{-6}$). In all the graphs, each color corresponds to a different allele. Note that a given color can be reasigned to a new allele later in the simulation. This figure represents the evolution along time of (A),(D) the frequency of each *PRDM9* allele and their corresponding (B),(E) proportion of active sites and (C),(F) selection coefficient. Here, the genetic dosage changes the regime of the dynamic passing it from polymorphic to monomorphic.

The example on figure 5 corresponds to only one particular parameter settings showing that gene dosage can act against diversity. But is it a general feature or is it just restrained to specific conditions ? Is there a qualitative genetic dosage regime ? To address these questions, a perturbative development was conducted, starting from the analytical results shown above (for the complete analytical developments, see Supplementary materials part 4). Here, perturbative means that it is valid only when the impact of dosage is weak, however it will also give the conditions on the model parameters for this to be the case. Thus, with dosage, the evolution of the allele frequency through time is now given by :

$$\frac{dln(f)}{dt} = -\frac{\alpha^{het}}{2}(z-\bar{z}) + \sigma_0(f-\bar{f}) \tag{15}$$

where α^{het} corresponds to the α coefficient for a heterozygote and :

$$\sigma_0 = \frac{w^{hom}(0) - w^{het}(0,0)}{w^{het}(0,0)} \tag{16}$$

is the relative difference in fertility between homozygotes and heterozygotes for young alleles (or homozygous advantage).

We therefore note that the evolution of the frequency of an allele no longer depends solely on its intrinsic age (i.e. erosion level z) but also on its frequency at time t. Contrary to what was observed without dosage, the fate of an allele is now determined by a competition between two different effects : an age effect and a frequency effect. The age effect (everything else being equal, older alleles tend to have a lower fitness than younger alleles) was already present without dosage (equation 12). The frequency effect (everything being equal, more frequent alleles benefit from a homozygous advantage and will therefore tend to become even more frequent) is the specific contribution of dosage. Importantly, the frequency effect (which is proportional to σ_0) systematically acts against diversity.

Based on the equation 15, we can determine when the dosage will play an important role. Indeed these effects will be all the more pronounced as σ_0 is large. Also, there should be enough time between successive invations for eviction to take place. An analytical approximation for the time between invasions is given by :

$$\tau = \frac{2}{\mu\alpha\bar{z}} \tag{17}$$

Altogether we predict a qualitative change induced by gene dosage on *PRDM9* standing diversity when *Nu* is large (i.e. the regime would have been polymorphic without dosage) and $\sigma_0\tau$ is also large (i.e. eviction

due to homozygous advantage is sufficiently strong and has enough time to occurs, and thus the regime turns out to be monomorphic with dosage). Conversely, we predict a polymorphic regime when Nu is large and when the dosage effects are negligible, i.e. when $\sigma_0\tau$ is small.

These predictions were verified by conducting bi-dimensional scaling experiments. First, a scaling of the parameters u and v was performed to determine which parameter combination would lead to a change of regime (Figure 6). Once the scaling performed, we note that a certain number of simulations underwent a change of regime between the experiments without and with genetic dosage. In particular, the most pronounced differences that are observed between 6A and B relate to parameters allowing for a polymorphic regime without dosage (i.e. u high, or $4Nu >> 1$), a small erosion parameter v giving a fairly substantial latency time between two allele invasions (τ) and a high selection coefficient associated to genetic dosage σ_0. In all cases, the criterion that $\sigma_0\tau$ should be large is met (*i.e.* always > 3).

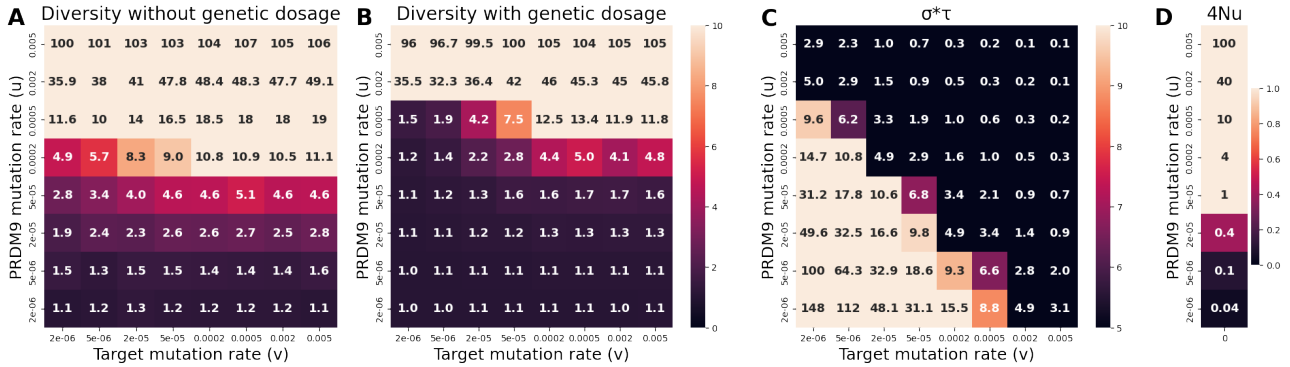


Figure 6. Change of regime between without and with genetic dosage according to mutation rates of *PRDM9* u and targets v. Diversity without (A) and with (B) genetic dosage. (C) Selection coefficient associated to dosage (σ) times the time between two allele invasion (τ). (D) $4Nu$ for each u

In order to get additional insights, another bi-dimensional scaling experiment was performed, this time allowing for the variation of the parameters d, which is the mean number of DSBs per meiosis per chromosome pairs and $\bar{y}$ corresponding to the mean value of the binding affinity distribution of all the *PRDM9* target sites (Figure 7). Here, a change of regime is essentially found when these two parameters are low. It is also well predicted by the criterion that $\sigma_0\tau$ should not be large (Figure 7C). Of note, both parameters, d and $\bar{y}$, play on σ_0, rather than on τ. Thus, if DSBs are limiting or sites tend to be of low affinity, a change in dosage can make a big difference in the chance of having at least one DSB in a symmetrically bound site.

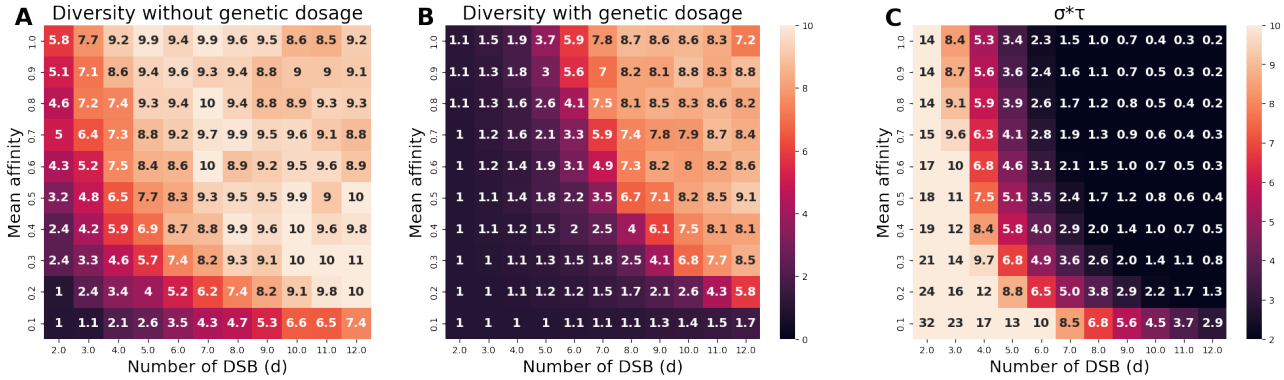


Figure 7. Change of regime between without and with genetic dosage according to mean DSBs per chromosome pairs per meiosis (d) and the mean affinity of targets ($\bar{y}$) (with $u = 2 \times 10^{-4}$ and $v = 5 \times 10^{-5}$). Diversity without (A) and with (B) genetic dosage. (C) Selection coefficient associated to dosage (σ) times the time between two allele invasion (τ)

These results show that there are new constraints applying to the Red Queen dynamics regime. In particular, we note that the conditions for obtaining a polymorphic regime are restricted to a relatively short time between two invasions of alleles as well as a fairly weak heterozygous disadvantage (haplo-insufficiency) in young alleles.

Empirical calibrations of the model

Empirical input parameters

Finally an empirical calibration of the model was attempted, using parameter values roughly estimated based on current knowledge in mammals and, more specifically, in the mouse. The effective population size (N_e) of different *Mus musculus* subspecies such as *Mus m. musculus* or *Mus m. domesticus* is around 10^5 [28]. The number of hostspots recognized per *PRDM9* alleles across the genome (h) is estimated between $15,000$ and $20,000$ [29]. Knowing that the mouse possesses 20 pairs of chromosomes this corresponds approximately to $h \approx 1000$ sites per chromosome. The mean number of DSBs (d) performed per individual is estimated between 200 and 250 per meiosis [30] which correspond to approximately 10 DSBs per chromosome pair per meiosis. Concerning the target mutation rate, considering that the mutation rate per nucleotide per generation across the mouse genome is 5.4×10^{-9} [31] and that the motifs characterising most of the hotspots in the mouse are several tens of nucleotides long [21], the inactivating mutation rate per generation (v) at the targets can be estimated at 10^{-7} [19]. The other mutation rate to determine is the one at the *PRDM9* locus (u). It has been estimated at (10^{-5}) in humans [17], but this is the raw mutation rate (including mutation that lead to either non-functional or functionally equivalent alleles). In contrast, the mutation rate u of the model is the net functional mutation rate (probability of creating a new functional allele recognizing entirely different targets), and the net rate is likely to be smaller than the raw experimental

estimate. Accordingly we tried different values around 10^{-5}. An exponential distribution for the affinities has a good fit to observed distribution obtained from Chip-seq experiments [20]. The mean, however, is unknown but can be determined by assuming an average of $5,000$ targets bound per meiocyte [32] out of $20,000$ available targets. We obtain a mean of approximately 0.2. We tried different values of $\bar{y}$ between 0.2 and 2.

In all simulations shown thus far, the fitness of an individual is assumed to be directly proportional to the number of gametes that this individual produces. In reality, there is an excess of gamete, in particular in males , such that the fitness might respond non-linearly to a decrease in the rate of success of meiosis. In the context of the present empirical calibration, we modeled this excess of gametes by leaving the possibility of carrying several meioses per individual before choosing a new individual from the population. The number of attempts (n_{mei}) ranged from $n_{mei} = 1$ (i.e. fitness proportional to meiosis success) to $n_{mei} = 5$ (thus mimicking a situation where an up to 80% reduction in the number of gametes would not affect the reproductive success of individuals).

Finally, since a population size of $N = 10^5$ is too large for running the simulator, a standard scaling argument was used, by setting $N = 5 \times 10^3$ and multiplying u and v by a factor 20 (i.e. using between $u = 6 \times 10^{-5}$ to $u = 6 \times 10^{-4}$ and $v = 2 \times 10^{-6}$). The other parameters are set for the smallest chromosome in mouse, so with lower h and d than the mean in the entire genome ($h = 800$, $d = 8$, $\bar{y} = 0.2$).

Model predictions

The simulator was calibrated based on these rough empirical estimates for its parameters, then run and monitored for several key summary statistics. There, we focus on the genetic diversity of PRDM9 (D), the proportion of eroded sites ($\bar{z}$), the mean haplo-insuficiency at the birth of the allele (σ_0) and the mean haplo-insuficiency over alleles sampled at the equilibrium regime ($\sigma_{\bar{z}}$). The predictions are shown in the table 2 and examples of Red Queen dynamics are shown in figures 8 and 9

u	$\bar{y}$	n_{mei}	D	$\bar{z}$	σ_0	$\sigma_{\bar{z}}$
6×10^{-5}	0.2	1	1	0.26	2.9×10^{-2}	5×10^{-2} [1]
6×10^{-4}	0.2	1	1.2	0.24	2.9×10^{-2}	4.5×10^{-2} [1]
6×10^{-5}	2	1	1.7	0.26	5.1×10^{-4}	9×10^{-4} [1]
6×10^{-5}	0.2	5	2.5	0.22	2.9×10^{-2}	7.5×10^{-2} [2]

TABLE 2. Results of the simulations after empirical calibrations. these tab shows the values of output statistics at the stationary equilibrium (*PRDM9* diversity D, mean activity θ, mean haplo-insuficiency at the birth of the allele (σ_0) and mean haplo-insuficiency at the equilibrium ($\sigma_{\bar{z}}$) depending on different values of parameter (mutation rate at *PRDM9* locus (u), mean of the affinity distribution ($\bar{y}$) and maximum number of meiosis allowed for each individual) with fixed parameters ($d = 8$, $h = 800$ and $v = 2 \times 10^{-6}$)

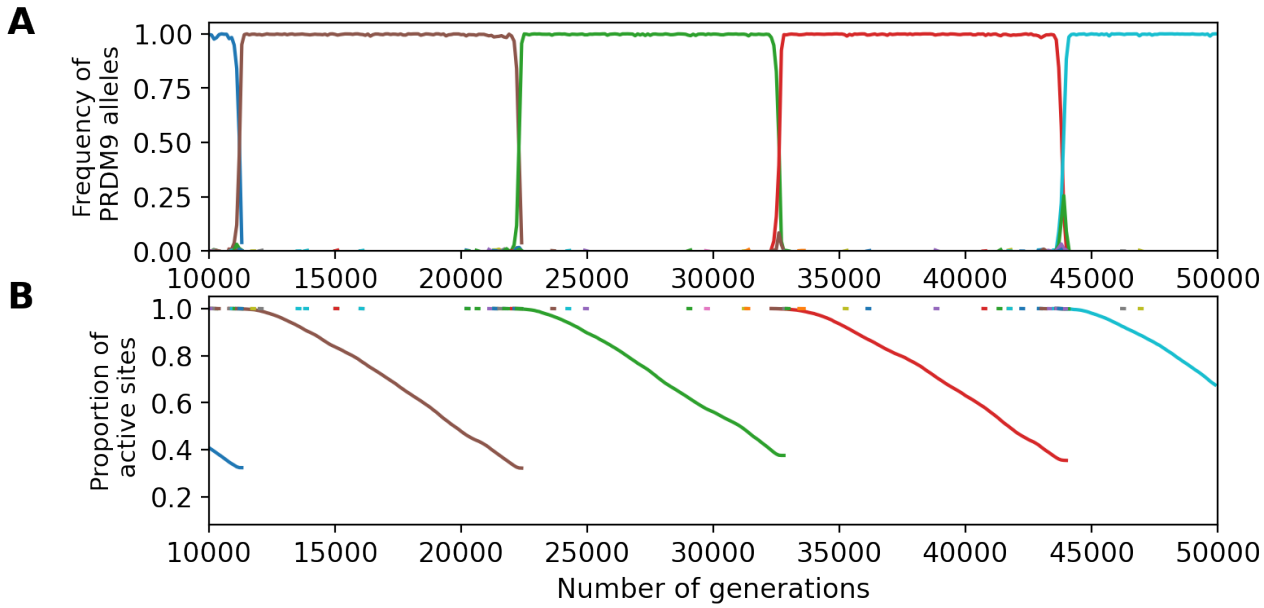


Figure 8. An example of a *PRDM9* dynamic with only one meiosis allowed per individual ($n_{mei} = 1$, $u = 6 \times 10^{-5}$ and $\bar{y} = 0.2$). In all the graphs, each color correspond to a different allele. If a color appears several generation after the death of an allele of the same color, it is considered as a complete different allele.This graph represents the evolution along time of the frequency of each PRDM9 allele (A) and their proportion of active sites (B). This corresponds to an eviction regime.

The first line of Table 2 reports the results obtained after simulations run with the lower range values for u, y and n_{mei}. The level of erosion (z) predicted by the model is consistent with what is known in the mouse (between 20% and 50% of erosion [20],[21],[1]). However, the predicted *PRDM9* diversity appears to be too low : these simulations result in a monomorphic regime (*i.e.* $D \sim 1$), whereas the *PRDM9* diversity observed in natural populations of mice is of the order of $D = 6$ to $D = 18$ [20] (See Supplementary materials section 5). Of note, the diversity predicted by the model is the diversity of functionally different alleles (owing to the assumption made by our model of non-overlapping sets of targets for different alleles). In reality, closely related alleles can share a substantial fraction of their targets [20]. However, even accounting for this redundancy, the empirical functional diversity is still typically greater than 1, of the order of 2 to 6 in

mouse subspecies [20] (See Supplementary materials section 5). Thus, the monomorphic regime obtained here seems to be in contradiction with empirical observations. This monomorphic regime can be explained by the fact that we are in a range of parameters involving an eviction regime where one allele dominates in the population while all the other alleles are counter-selected (Figure 8A). In turn, this eviction is caused by a non-negligible haplo-insufficiency, especially in old alleles (Table 2 column $\sigma_{\bar{z}}$), with a reduction of a few percents in the success of meiosis in hemizygotes (or, equivalently, heterozygotes with two alleles of same age), compared to homozygotes. Interestingly, such levels of haplo-insufficiency are comparable to those observed in old alleles in the mouse (B6, C3H [27]). The predicted haplo-insufficiency of young alleles is weaker (Table2 column σ_0). Such an age-dependency for the impact of gene dosage was previously suggested (Davies *et al.*).

Increasing the *PRDM9* functional mutation rate (u) by a factor 10 does not even allow to get out of this monomorphic regime with predicted *PRDM9* diversity still very close to one (second row of Table 2). Moreover, there is *a priori* no reason to consider that the *PRDM9* mutation rate should be higher in mice than in humans. Alternatively, increasing the mean affinity (or, equivalently, the concentration of *PRDM9* in the cell) allows for higher levels of *PRDM9* equilibrium diversity while maintaining empirically reasonable levels of erosion (third row of Table 2). However, the mean number of sites bound by *PRDM9* in a meiocyte would then predicted to be too large ($\sim 20,000$), compared to current empirical knowleldge ($\sim 5,000$ [32]). In addition, in this parameter regime, the model predicts very low levels of haplo-insuficiency ($\sigma_{\bar{z}} < 1 \times 10^{-3}$ for old alleles), substantially lower than empirically observed levels in the mouse ($\sigma_{\bar{z}} > 1 \times 10^{-2}$ [27]).

All these results were obtained assuming that the reproductive fitness is directly proportional to the success rate of meiosis, which, as mentioned above, is not realistic in the case of mammals. Assuming instead that gametes are not limiting by allowing for up to five meioses per individual (fourth row of Table 2), we observe a higher functional diversity with a still reasonable level of erosion (Figure 9). Indeed, we are no more in the eviction regime which tends to acts against diversity (Figure 8B). The predicted reduction in meiosis success in hemizygotes is also consistent with the ones observed empirically in mice (~ 2 to 5% [27]). Of note, by making meioses non-limiting, we decoupled reproductive fitness and the meiosis success rate, thus reducing the decline in fertility due to dosage, which makes it possible to get out of the eviction regime. Altogether, we deduce that there is no need for much more realism in our model to correctly fit the empirical data.

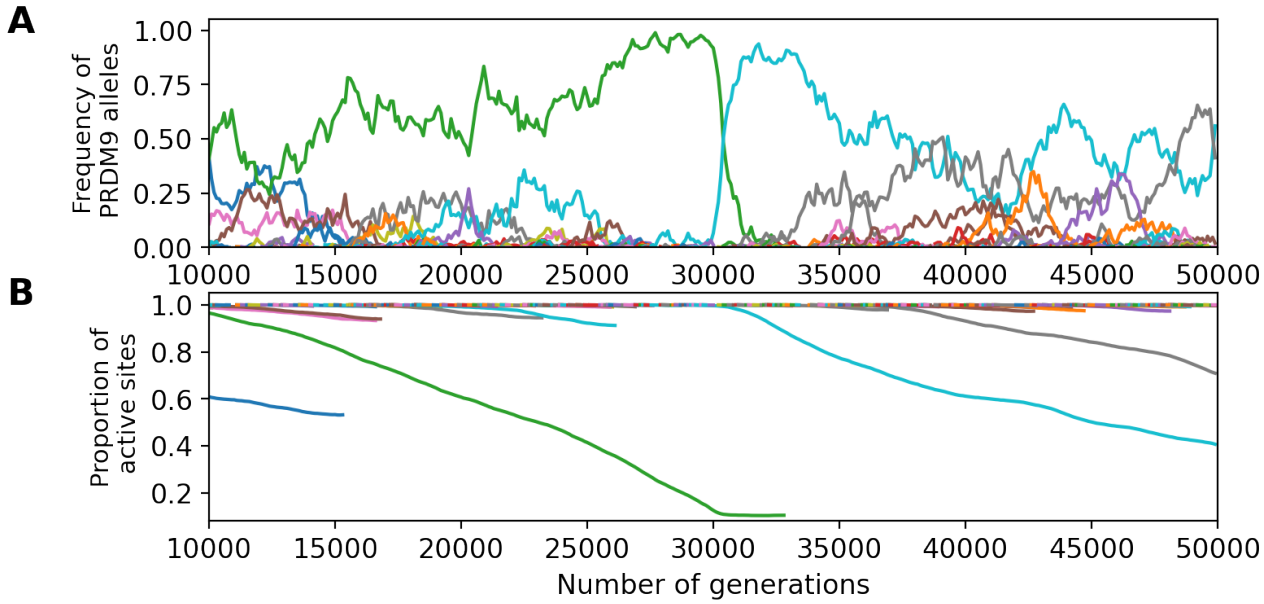


Figure 9. An example of a *PRDM9* dynamic with five meiosis per individual allowing up to 80% decrease in number of gametes ($n_{mei} = 5$, $u = 6 \times 10^{-5}$ and $\bar{y} = 0.2$). In all the graphs, each color correspond to a different allele. If a color appears several generation after the death of an allele of the same color, it is considered as a complete different allele.This graph represents the evolution along time of the frequency of each PRDM9 allele (A) and their proportion of active sites (B). This better corresponds to the polymorphic regime observed empirically.

Discussion

Fundamental role of the symmetrical binding of *PRDM9*

The first studies on the function of *Prdm9* uncovered its role in targeting DSBs at specific sites, by its capacity to set histone marks to recruit the DSB machinery [33],[34],[35],[36]. More recently, Davies *et al.* (2016) discovered that this gene plays another important role during meiosis, namely, by facilitating chromosome pairing [1]. In this process, the symmetrical binding of *PRDM9* to its target sites plays a key role. Indeed, by bringing back to the chromosomal axis the sites on the homologue, it presents it to the single strand of DNA produced by the DSB, thereby facilitating the search for the complementary sequence. This second function of PRDM9 combined with the differential erosion of the target sites between sub-species (which is itself a direct consequence of the first function of PRDM9, the recruitment of the DSB machinery), was found to be sufficient for explaining most aspects of the hybrid sterility pattern [1].

Here, we show that, similarly, these two aspects of the function of PRDM9 during meiosis, provide the necessary ingredients for explaining the intragenomic Red Queen occurring in a panmictic context - thus giving a globally coherent picture bridging the gap between the molecular mechanisms and the evolutionary

dynamics of recombination.

More fondamentally, the Red Queen results from a balance between erosion and invasion by new *PRDM9* alleles that are positively selected. On the one hand, the erosion is the direct consequence of the first role of *PRDM9* (recruitment of the DSB machinery), which triggers dBGC. On the other hand, the positive selection comes from differences in fertility between young alleles with brand new active target sites and ensuring a good fertility, and old alleles having eroded there high affinity targets leading to asymmetrical binding at low affinity sites generating bad fertility. Symmetry therefore plays an important role both in hybrid context where the differential erosion on both chromosomes generates sequence asymmetry, and in panmictic context where the erosion of high affinity targets generate a statistical asymmetry of binding.

In addition, scaling experiments have helped to better understand what determines the steady state regime and we recover the main results of Latrille *et al.* although now with a mechanistic foundation. In particular, the *PRDM9* diversity is mostly determined by the mutational input at the *PRDM9* locus u (and in a very weaker way by the target mutation rate v) while the mean erosion and fertility at equilibrium are the result of a balance between site erosion (v, g) and selection of new alleles (u, α).

Impact of genetic dosage on the Red Queen dynamics

The *Prdm9* genetic dosage implies that a homozygote has a *PRDM9* concentration for its allele which is twice that of each of the two alleles of a heterozygote. Everything else being equal, this increases the occupancy of each target and therefore the probability of symmetrical binding and fertility. This clearly impact the Red Queen dynamics for certain combinations of parameters by acting against diversity (polymorphic regime becomes monomorphic). In these cases, on one side, a dominant allele is present and by its high frequency in a homozygous state it is positively selected even when it has eroded many of its target. On the other side, all the other alleles necessarily appearing in the heterozygous state are counter selected. The higher the selection coefficient associated to the homozygous advantage (captured here by σ_0), the stronger the effect against diversity. This effect is however mitigated when *PRDM9* is in high concentration in the cell. Indeed, when the PRDM9 concentration increases, the difference in binding probability for a given activity target decreases between a homozygote and a heterozygote (Supplemental Figure 1). By extension, this reduces the fertility gap between a homozygote and a heterozygote. Thus, the homozygous advantage would become almost insignificant when PRDM9 is present in high concentration in the cell (Supplemental Figure 2).

To avoid the counter effect on PRDM9 diversity of the genetic dosage, PRDM9 should be present in

high concentration in the cell. Empirically, genetic dosage is present [27] but the regime is still polymorphic, at least in the mouse ($D \approx 4$) and in humans ($D \approx 2$). Based on our model and empirical data, we therefore hypothesize that the PRDM9 expression level has been selected at a sufficiently high level so as to limit the impact of gene dosage. This is an interesting hypothesis that could be studied in the future.

Comparison with Baker *et al.*'s model

Recently, a new article by Baker *et al.* [26] and dealing with the same subject as ours was published. Several conclusions agree with ours. Most notably, Baker *et al.* also recognize that the symmetric binding of *PRDM9* provides the key ingredient for running the evolutionary Red Queen of recombination. However, the two studies differ on some important points.

First, in our model, the amount of *PRDM9* protein is not limiting, whereas Baker *et al.* assume strong competition between targets for PRDM9 proteins. This raises the empirical question of whether or not *PRDM9* is limiting, a question that still remains open. In any case, our model shows that competition between targets is not a necessary feature to drive the evolutionary dynamics of *PRDM9*.

The second difference concerns the affinity distribution. Baker *et al.* use a two-heat distribution. Here we use an exponential distribution, which is motivated by results from Chip-seq experiments [20], although with some uncertainty regarding the shape of the distribution in the low affinity range, which is not captured by Chip-seq. These different affinity distributions have clear consequences on the behavior of models with respect to dosage. Indeed, increasing the genetic dosage can have two opposing effects. On the one hand, it increases the symmetrical binding at sites of intermediate affinity. On the other hand, it also results in more sites of low affinity sites being bound, and this, most often asymmetrically. Depending on the balance between these two effects, different outcomes are obtained. In our case, where the affinity distribution has a moderate variance, the increased symmetry in intermediate affinity site always wins. As a result, there is always an advantage to increasing dosage. That is, homozygotes always have an advantage over heterozygotes, which creates an eviction regime that tends to play against diversity. A contrario, in Baker's model, the two-heat affinity distribution results in a non-monotonous dependency, with a turning point, such that the dosage is in favor of homozygotes for young alleles, but in favor of heterozygotes for old alleles. This turning point is one potential solution to the problem of eviction. Here, we present an alternative solution, which is captured in our model by the statistics $\sigma_0\tau$. Intuitively, eviction does not take place if the homozygote advantage is sufficiently weak and erosion sufficiently rapid, such that ageing alleles are replaced before eviction has enough time to take place. In the end, we can find empirically reasonable

parameter configurations such that the Red Queen runs under a coherent and plausible regime. Of note, we also predict empirically reasonable values for the haplo-insufficiency of old alleles, close to what is observed in the mouse (for allele B6 and C3H [27]). This haplo-insufficiency of old alleles is not observed systematically in Baker's *et al.* model, but only occasionally and in small populations.

Altogether, it could be interesting to empirically investigate the haplo-insufficiency for young and old alleles over a broader range of alleles and species, in order to determine whether a lifelong homozygous advantage is systematic or occasional. More globally, it will be important to unravel the exact roles of affinity distribution, *PRDM9* concentration and competition between targets in the evolutionary dynamics and in the dosage effects. Finally, all this raises the question of a possible evolution of the *PRDM9* expression level, the affinity (whether in its distribution or in its mean) and the number of targets recognized per allele.

Conclusion

All the work carried out in this article made it possible to better identify the molecular roles of *PRDM9* in the evolutionary dynamics of the Red Queen and the impact of genetic dosage on the system. In addition to being able to study in greater depth the model, it could be relevant to study the influence of other parameters on the dynamics or to test the model in a bi-populational context to see if it predicts hybrid sterility that could lead to speciation.

Data accessibility

The model, and the codes to generate the figures can be found at `https://github.com/alicegenestier/Red_Queen_PRDM9_Panmictic.git`. The model was implemented in C++ and all figures were created using Python.

Competing interest

The authors have no competing interests.

Funding

Agence Nationale de la Recherche, Grant ANR-19-CE12-0019 / HotRec.

Acknowledgment

All simulations of this work were performed using the computing facilities of the CC LBBE/PRABI.

References

1. DAVIES B, HATTON E, ALTEMOSE N, HUSSIN JG, PRATTO F, ZHANG G, HINCH AG, MORALLI D, BIGGS D, DIAZ R, PREECE C, LI R, BITOUN E, BRICK K, GREEN CM, CAMERINI-OTERO RD, MYERS SR et DONNELLY P. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature. 2016 Feb; 530. Number : 7589 Publisher : Nature Publishing Group:171-6. DOI : 10.1038/nature16931. Available from: http://www.nature.com/articles/nature16931 [Accessed on: 2022 Mar 18]

2. MIHOLA O, TRACHTULEC Z, VLCEK C, SCHIMENTI JC et FOREJT J. A Mouse Speciation Gene Encodes a Meiotic Histone H3 Methyltransferase. Science. 2009 Jan 16; 323. Publisher : American Association for the Advancement of Science:373-5. DOI : 10.1126/science.1163601. Available from: https://www-science-org.docelec.univ-lyon1.fr/doi/10.1126/science.1163601 [Accessed on: 2022 Mar 18]

3. OLIVER PL, GOODSTADT L, BAYES JJ, BIRTLE Z, ROACH KC, PHADNIS N, BEATSON SA, LUNTER G, MALIK HS et PONTING CP. Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. PLOS Genetics. 2009 Dec 4; 5. Publisher : Public Library of Science:e1000753. DOI : 10.1371/journal.pgen.1000753. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000753 [Accessed on: 2022 Mar 18]

4. MYERS S, BOTTOLO L, FREEMAN C, MCVEAN G et DONNELLY P. A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome. Science. 2005 Oct 14; 310. Publisher : American Association for the Advancement of Science:321-4. DOI : 10.1126/science.1117196. Available from: https://www-science-org.docelec.univ-lyon1.fr/doi/10.1126/science.1117196 [Accessed on: 2022 Mar 18]

5. JEFFREYS AJ, KAUPPI L et NEUMANN R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nature Genetics. 2001 Oct; 29. Number : 2 Publisher : Nature Publishing Group:217-22. DOI : 10.1038/ng1001-217. Available from: http://www.nature.com/articles/ng1001-217 [Accessed on: 2022 Mar 18]

6. KAUPPI L, JEFFREYS AJ et KEENEY S. Where the crossovers are : recombination distributions in mammals. Nature Reviews Genetics. 2004 Jun ; 5. Number : 6 Publisher : Nature Publishing Group:413-24. DOI : 10.1038/nrg1346. Available from: http://www.nature.com/articles/nrg1346 [Accessed on: 2022 Mar 18]

7. MCVEAN GAT, MYERS SR, HUNT S, DELOUKAS P, BENTLEY DR et DONNELLY P. The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. Science. 2004 Apr 23 ; 304. Publisher : American Association for the Advancement of Science:581-4. DOI : 10.1126/science.1092500. Available from: https://www-science-org.docelec.univ-lyon1.fr/doi/10.1126/science.1092500 [Accessed on: 2022 Mar 18]

8. BRUNSCHWIG H, LEVI L, BEN-DAVID E, WILLIAMS RW, YAKIR B et SHIFMAN S. Fine-Scale Maps of Recombination Rates and Hotspots in the Mouse Genome. Genetics. 2012 Jul 1 ; 191:757-64. DOI : 10.1534/genetics.112.141036. Available from: https://doi.org/10.1534/genetics.112.141036 [Accessed on: 2022 Mar 18]

9. BAUDAT F, BUARD J, GREY C, FLEDEL-ALON A, OBER C, PRZEWORSKI M, COOP G et MASSY B de. PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. Science. 2010 Feb 12 ; 327. Publisher : American Association for the Advancement of Science:836-40. DOI : 10.1126/science.1183439. Available from: https://www-science-org.docelec.univ-lyon1.fr/doi/10.1126/science.1183439 [Accessed on: 2022 Mar 18]

10. MYERS S, BOWDEN R, TUMIAN A, BONTROP RE, FREEMAN C, MACFIE TS, MCVEAN G et DONNELLY P. Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. Science. 2010 Feb 12 ; 327. Publisher : American Association for the Advancement of Science:876-9. DOI : 10.1126/science.1182363. Available from: https://www-science-org.docelec.univ-lyon1.fr/doi/10.1126/science.1182363 [Accessed on: 2022 Mar 18]

11. LANGE J, YAMADA S, TISCHFIELD SE, PAN J, KIM S, ZHU X, SOCCI ND, JASIN M et KEENEY S. The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. Cell. 2016 Oct 20 ; 167:695-708.e16. DOI : 10.1016/j.cell.2016.09.035. Available from: https://www.sciencedirect.com/science/article/pii/S0092867416313174 [Accessed on: 2022 Mar 18]

12. LI R, BITOUN E, ALTEMOSE N, DAVIES RW, DAVIES B et MYERS SR. A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination.

Nature Communications. 2019 Aug 29 ; 10:3900. DOI : 10.1038/s41467-019-11675-y. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6715734/ [Accessed on: 2022 Sep 21]

13. BOULTON A, MYERS RS et REDFIELD RJ. The hotspot conversion paradox and the evolution of meiotic recombination. Proceedings of the National Academy of Sciences. 1997 Jul 22 ; 94. Publisher : Proceedings of the National Academy of Sciences:8058-63. DOI : 10.1073/pnas.94.15.8058. Available from: https://www.pnas.org/doi/10.1073/pnas.94.15.8058 [Accessed on: 2022 Mar 18]

14. LESECQUE Y, GLÉMIN S, LARTILLOT N, MOUCHIROUD D et DURET L. The Red Queen Model of Recombination Hotspots Evolution in the Light of Archaic and Modern Human Genomes. PLOS Genetics. 2014 Nov 13 ; 10. Publisher : Public Library of Science:e1004790. DOI : 10.1371/journal.pgen.1004790. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004790 [Accessed on: 2022 Mar 18]

15. ÚBEDA F et WILKINS JF. The Red Queen theory of recombination hotspots. Journal of Evolutionary Biology. 2011 ; 24. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1420-9101.2010.02187.x:541-53. DOI : 10.1111/j.1420-9101.2010.02187.x. Available from: http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1420-9101.2010.02187.x [Accessed on: 2022 Mar 18]

16. AUTON A, FLEDEL-ALON A, PFEIFER S, VENN O, SÉGUREL L, STREET T, LEFFLER EM, BOWDEN R, ANEAS I, BROXHOLME J, HUMBURG P, IQBAL Z, LUNTER G, MALLER J, HERNANDEZ RD, MELTON C, VENKAT A, NOBREGA MA, BONTROP R, MYERS S, DONNELLY P, PRZEWORSKI M et MCVEAN G. A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. Science. 2012 Apr 13 ; 336. Publisher : American Association for the Advancement of Science:193-8. DOI : 10.1126/science.1216872. Available from: https://www-science-org.docelec.univ-lyon1.fr/doi/full/10.1126/science.1216872 [Accessed on: 2022 Mar 17]

17. JEFFREYS AJ, COTTON VE, NEUMANN R et LAM KWG. Recombination regulator PRDM9 influences the instability of its own coding sequence in humans. Proceedings of the National Academy of Sciences. 2013 Jan 8 ; 110. Publisher : Proceedings of the National Academy of Sciences:600-5. DOI : 10.1073/pnas.1220813110. Available from: https://www.pnas.org/doi/10.1073/pnas.1220813110 [Accessed on: 2022 Mar 18]

18. BUARD J, RIVALS E, SEGONZAC DDd, GARRES C, CAMINADE P, MASSY Bd et BOURSOT P. Diversity of Prdm9 Zinc Finger Array in Wild Mice Unravels New Facets of the Evolutionary Turnover of this Coding Minisatellite. PLOS ONE. 2014 Jan 13 ; 9. Publisher : Public Library of Science:e85021. DOI : 10.1371/journal.pone.0085021. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0085021 [Accessed on: 2022 Mar 18]

19. LATRILLE T, DURET L et LARTILLOT N. The Red Queen model of recombination hot-spot evolution : a theoretical investigation. Philosophical Transactions of the Royal Society B : Biological Sciences. 2017 Dec 19 ; 372. Publisher : Royal Society:20160463. DOI : 10.1098/rstb.2016.0463. Available from: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2016.0463 [Accessed on: 2022 Mar 17]

20. SMAGULOVA F, BRICK K, PU Y, CAMERINI-OTERO RD et PETUKHOVA GV. The evolutionary turnover of recombination hot spots contributes to speciation in mice. Genes & Development. 2016 Feb 1 ; 30:266-80. DOI : 10.1101/gad.270009.115. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4743057/ [Accessed on: 2022 Mar 18]

21. BAKER CL, KAJITA S, WALKER M, SAXL RL, RAGHUPATHY N, CHOI K, PETKOV PM et PAIGEN K. PRDM9 Drives Evolutionary Erosion of Hotspots in Mus musculus through Haplotype-Specific Initiation of Meiotic Recombination. PLOS Genetics. 2015 Jan 8 ; 11. Publisher : Public Library of Science:e1004916. DOI : 10.1371/journal.pgen.1004916. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004916 [Accessed on: 2022 Mar 17]

22. GREGOROVÁ S et FOREJT J. PWD/Ph and PWK/Ph Inbred Mouse Strains of Mus
cyrchar
cyrt. musculus Subspecies-a Valuable Resource of Phenotypic Variations and Genomic Polymorphisms. Folia biologica. 2000 Feb 1 ; 46:31-41

23. GREGOROVA S, GERGELITS V, CHVATALOVA I, BHATTACHARYYA T, VALISKOVA B, FOTOPULOSOVA V, JANSA P, WIATROWSKA D et FOREJT J. Modulation of Prdm9-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. eLife. 2018 ; 7:e34282. DOI : 10.7554/eLife.34282. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5902161/ [Accessed on: 2022 Mar 18]

24. BAUDAT F, IMAI Y et MASSY B de. Meiotic recombination in mammals : localization and regulation. Nature Reviews Genetics. 2013 Nov ; 14. Number : 11 Publisher : Nature Publishing Group:794-806. DOI : 10.1038/nrg3573. Available from: http://www.nature.com/articles/nrg3573 [Accessed on: 2022 Mar 18]

25. BORDE V et MASSY B de. Programmed induction of DNA double strand breaks during meiosis : setting up communication between DNA and the chromosome structure. Current Opinion in Genetics & Development. Genome architecture and expression 2013 Apr 1 ; 23:147-55. DOI : 10.1016/j.gde.2012.12.002. Available from: https://www.sciencedirect.com/science/article/pii/S0959437X12001463 [Accessed on: 2022 Mar 18]

26. BAKER Z, PRZEWORSKI M et SELLA G. Down the Penrose stairs : How selection for fewer recombination hotspots maintains their existence. Pages : 2022.09.27.509707 Section : New Results. 2022 Sep 28. DOI : 10.1101/2022.09.27.509707. Available from: https://www.biorxiv.org/content/10.1101/2022.09.27.509707v1 [Accessed on: 2022 Oct 10]

27. BAKER CL, PETKOVA P, WALKER M, FLACHS P, MIHOLA O, TRACHTULEC Z, PETKOV PM et PAIGEN K. Multimer Formation Explains Allelic Suppression of PRDM9 Recombination Hotspots. PLoS Genetics. 2015 Sep 14 ; 11:e1005512. DOI : 10.1371/journal.pgen.1005512. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4569383/ [Accessed on: 2022 Mar 18]

28. SALCEDO T, GERALDES A et NACHMAN MW. Nucleotide Variation in Wild and Inbred Mice. Genetics. 2007 Dec ; 177:2277-91. DOI : 10.1534/genetics.107.079988. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2219490/ [Accessed on: 2022 Mar 18]

29. BRICK K, SMAGULOVA F, KHIL P, CAMERINI-OTERO RD et PETUKHOVA GV. Genetic recombination is directed away from functional genomic elements in mice. Nature. 2012 May ; 485. Number : 7400 Publisher : Nature Publishing Group:642-5. DOI : 10.1038/nature11089. Available from: http://www.nature.com/articles/nature11089 [Accessed on: 2022 Mar 18]

30. KAUPPI L, BARCHI M, LANGE J, BAUDAT F, JASIN M et KEENEY S. Numerical constraints and feedback control of double-strand breaks in mouse meiosis. Genes & Development. 2013 Apr 15 ; 27:873-86. DOI : 10.1101/gad.213652.113. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3650225/ [Accessed on: 2022 Mar 18]

31. UCHIMURA A, HIGUCHI M, MINAKUCHI Y, OHNO M, TOYODA A, FUJIYAMA A, MIURA I, WAKANA S, NISHINO J et YAGI T. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. Genome Research. 2015 Aug ; 25:1125-34. DOI : 10.1101/gr.186148.114. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4509997/ [Accessed on: 2022 Mar 18]

32. BAKER CL, WALKER M, KAJITA S, PETKOV PM et PAIGEN K. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. Genome Research. 2014 May ; 24:724-32. DOI : 10.1101/gr.170167.113. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4009602/ [Accessed on: 2022 Mar 18]

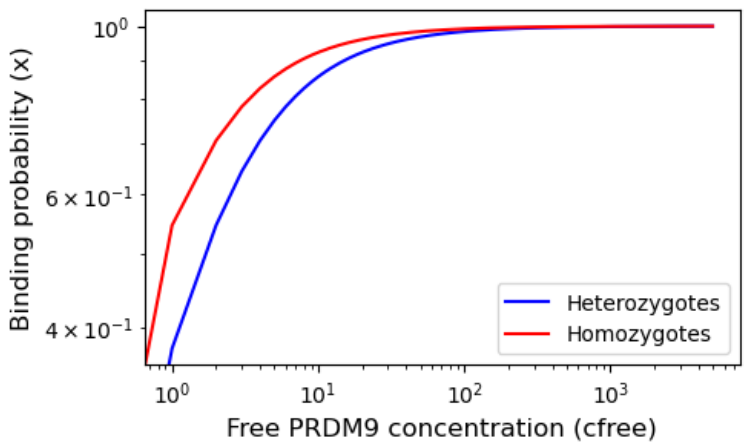
33. HAYASHI K, YOSHIDA K et MATSUI Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature. 2005 Nov ; 438. Number : 7066 Publisher : Nature Publishing Group:374-8. DOI : 10.1038/nature04112. Available from: http://www.nature.com/articles/nature04112 [Accessed on: 2022 Mar 18]

34. WU H, MATHIOUDAKIS N, DIAGOURAGA B, DONG A, DOMBROVSKI L, BAUDAT F, CUSACK S, MASSY B de et KADLEC J. Molecular Basis for the Regulation of the H3K4 Methyltransferase Activity of PRDM9. Cell Reports. 2013 Oct 17 ; 5:13-20. DOI : 10.1016/j.celrep.2013.08.035. Available from: https://www.sciencedirect.com/science/article/pii/S2211124713004762 [Accessed on: 2023 Feb 1]

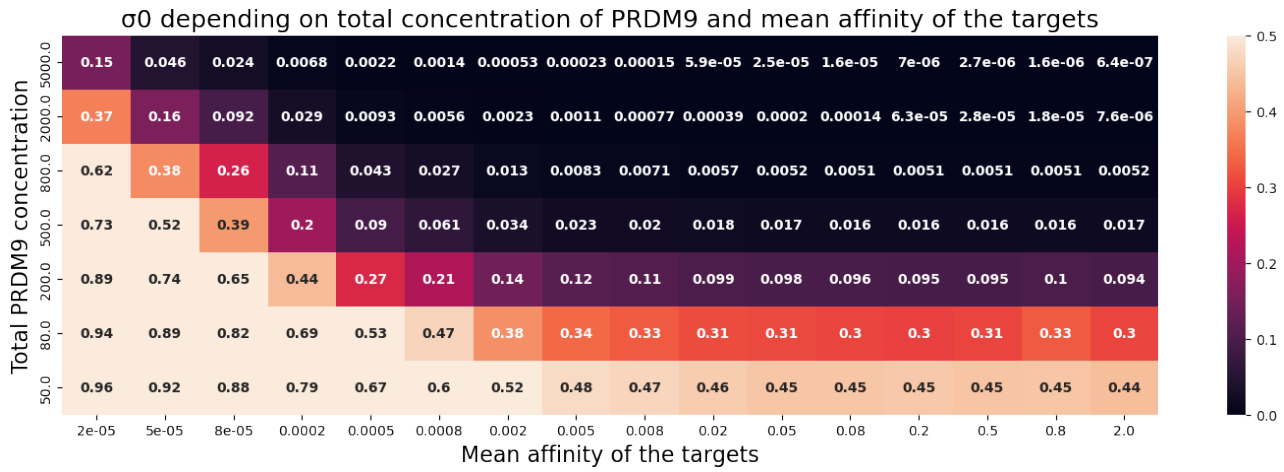
35. ERAM MS, BUSTOS SP, LIMA-FERNANDES E, SIARHEYEVA A, SENISTERRA G, HAJIAN T, CHAU I, DUAN S, WU H, DOMBROVSKI L, SCHAPIRA M, ARROWSMITH CH et VEDADI M. Trimethylation of Histone H3 Lysine 36 by Human Methyltransferase PRDM9 Protein. The Journal of Biological Chemistry. 2014 Apr 25 ; 289:12177-88. DOI : 10.1074/jbc.M113.523183. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4002121/ [Accessed on: 2023 Feb 1]

36. GREY C, BARTHÈS P, FRIEC GCL, LANGA F, BAUDAT F et MASSY Bd. Mouse PRDM9 DNA-Binding Specificity Determines Sites of Histone H3 Lysine 4 Trimethylation for Initiation of Meiotic Recombination. PLOS Biology. 2011 Oct 18 ; 9. Publisher : Public Library of Science:e1001176. DOI : 10.1371/journal.pbio.1001176. Available from: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001176 [Accessed on: 2023 Feb 1]

Supplemental figures



Supplemental Figure 1. Binding probability depending on the free *PRDM9* concentration for homozygotes heterozygotes ($y = 0.6$)**.** The binding probability for homozygotes is always higher than for heterozygotes but the difference decreases when the free *PRDM9* concentration increases



Supplemental Figure 2. Homozygotes selection coefficient associated to genetic dosage (σ_0) depending on the total *PRDM9* concentration in the cell and the mean affinity of the target sites. The σ_0 is all the greater as the concentration of *PRDM9* and the average affinity is low