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 Claude Bernard Lyon 1, LBBE, UMR 5558, CNRS, VAS, Villeurbanne, France

* nicolas.lartillot@univ-lyon1.fr

Supplementary Appendix S1 : Derivation of the analytical approximation to the equilibrium regime of the Red Queen

A   Introduction

In this document, we give the detailed derivation of the analytical approximation for the stationary regime of the Red Queen process. The derivation proceeds along similar lines as in Latrille *et al.* [1]. It relies on a self-consistent mean field argument. In brief, the derivation assumes the following hypotheses :

- Wright-Fisher model (*i.e.* non-overlapping generations) with mutation and selection

- Panmictic population (random mating)

- Constant population size N

- Highly polymorphic model

- Weak erosion

- Strong selection implying that genetic drift can be ignored

The first four hypotheses are entailed by the simulator. The last three are additional assumptions that are made in order to make the analytical derivation feasible. The analytical results will therefore be valid only in regimes in which these three assumptions are met.

Based on these hypotheses, we first express the frequency $f(t)$ and the proportion of active site , $\theta(t)$ through time of a typical *PRDM9* allele. We will then obtain expressions for the erosion level of an allele as a function of its age, and the mean erosion level of the population. Relying on the argument that the population itself is composed of such typical *PRDM9* alleles successively invading, with a mean time interval between successive invasions noted τ, itself depending on the mean fitness of the population, we can obtain a self-consistent expression for the mean erosion level, from where we can then express all summary statistics of interest.

In terms of notations, since we consider that all alleles follow the same trajectory over their existence, differing only in their arrival time in the population, all quantities will be expressed as a function of t (e.g. $f(t)$ for the frequency, or $\theta(t)$ for the proportion of active sites for the allele), t being understood as the time since the birth of the allele. This thus differs from the notations in the main text, where time is absolute and measured relative to a single origin (defined as the time when the simulation run has reached equilibrium, at which point the summary statistics start to be monitored), and quantities are explicitly indexed by the specific allele or allelic combination for which they are being computed (e.g. $f_{i,t}$ for the frequency of allele i at time t, or $w_{i,j,t}^{het}$ for the fitness of an individual with alleles i and j at time t).

In addition, below, we will consider a change of variable, expressing the same quantities as a function, not of time t, but as a function of the allele intrinsic age z (which is defined below and which is an increasing function of t). With a slight abuse of notations, the same symbol will be used as a function of t or z (e.g. $f(t)$ or $f(z)$, or even simply f), given that the meaning will be clear depending on the context.

B   Analytical developments

B.1   Studied model

First, assuming negligible random drift, the evolution of the frequency through time $f(t)$ of a typical allele is deterministic and is given by:

$$\frac{df}{dt} = \frac{w^*(t) - \bar{w}}{\bar{w}} f, \tag{1}$$

where $w^*(t)$ is the mean fertility of the allele at time t and $\bar{w}$ is the mean fertility of all *PRDM9* alleles in the population (weighted by their frequency). Of note, under the mean-field approximation considered here, $\bar{w}$ is, by definition, assumed independent of time. Furthermore, it is considered as an independent variable, which will be determined in a second step, based on a self-consistent argument.

Equation 1 means that selection will depend on differential erosion. In turn, erosion depends on the frequency trajectory of the allele. To decouple these effects, we start by observing that the fraction of active target sites for the allele varies approximately as :

$$\frac{d\theta}{dt} \approx -\rho f \theta \tag{2}$$

In this equation, ρ is the erosion rate per generation. It can be expressed as :

$$\rho = (2Nv)(2g) \tag{3}$$

where $g = \frac{d}{8h}$ is the mean gene conversion rate (*i.e.* the probability that a site undergoes a DSB and a repair). Thus, ρ is equal to :

$$\rho = \frac{Nvd}{2h} \tag{4}$$

Equation (2) expresses the fact that the rate of extinction of target sites is equal to the mutation rate at the level of the population $(2Nv)$ multiplied by the probability of fixation of the inactive mutant. Assuming strong gene conversion $(4Ng >> 1)$, this probability is well approximated by twice the conversion rate, *i.e.* $2fg$. Note that the rate of erosion given by (3) is only approximate. A more accurate expression accounting for the affinity distribution of the sites will be given below.

B.2   The intrinsic age of an allele (z)

Therefore,

$$z(t) = \rho \int_0^t f(u)\, du \tag{5}$$

can be seen as a measure of the cumulative erosion level of an allele. As such, it can be used as a measure of the intrinsic age of the allele. We can then express the equation describing the evolution of f and θ directly as a function of z, instead of t. This change of variable will entail the following factor :

$$\frac{dz}{dt} = \rho f. \tag{6}$$

As mentioned above, our derivation assumes weak erosion. Mathematically, this translates into the assumption that $z \ll 1$.

B.3   Frequency of an allele (f) depending on its age (z) : f(z)

Then, we want to express the evolution of the frequency of an allele in the population as a function on its intrinsic age z:

$$\frac{df}{dz} = \frac{df}{dt}\frac{dt}{dz} = \left[\frac{w^*(z) - \bar{w}}{\bar{w}}\right] \cdot \left[\frac{1}{\rho}\right] = \frac{1}{\rho}\left(\frac{w^*(z)}{\bar{w}} - 1\right) \tag{7}$$

However, $\frac{w^*}{\bar{w}}$ is extremely close to 1, so $\frac{w^*}{\bar{w}} - 1$ is close to 0. In addition, we know that, in the vicinity of 0, $x \approx ln(1+x)$. As a result, we can write

$$\frac{1}{\rho}\left(\frac{w^*(z)}{\bar{w}} - 1\right) \approx \frac{1}{\rho}ln\left(1 + \left(\frac{w^*(z)}{\bar{w}} - 1\right)\right) = \frac{1}{\rho}ln\left(\frac{w^*(z)}{\bar{w}}\right) \tag{8}$$

Working in the weak erosion limit ($z \ll 1$) allows us to linearize $ln(w)$ in the vicinity of 0 :

$$ln\left(\frac{w^*(z)}{\bar{w}}\right) = ln(w^*(z)) - ln(\bar{w}) \approx ln(w(0,0)) - \frac{\alpha}{2}(z+\bar{z}) - ln(w(0,0)) + \frac{\alpha}{2}(\bar{z}+\bar{z}) = -\frac{\alpha}{2}(z-\bar{z}) \tag{9}$$

where:

$$\alpha = \left|\frac{\partial ln(w)}{\partial z}\right|_{(z=0)} \tag{10}$$

is the slope at the origin of $ln(w)$. It depends on the mechanistic parameters, in a way that will be determined further below.

Then, by replacement in equation (7) :

$$\boxed{\frac{df}{dz} \approx -\frac{\alpha}{2\rho}(z - \bar{z})} \tag{11}$$

Integrating equation (11), with the constraint that $f(0) = 0$, gives :

$$f(z) = -\frac{\alpha}{4\rho}z[z - 2\bar{z}]. \tag{12}$$

The function f has the shape of a concave parabola and $f(z) = 0$ when $z = 0$ and $z = 2\bar{z}$.

3

B.4 Mean age of an allele in the population ($\bar{z}$) : a self-consistent derivation

Equations (11) and (12) depend on $\bar{z}$, and thus we now need to express $\bar{z}$ as a function of the model parameters.

The mean allele age of the population $\bar{z}$ is equal to :

$$\bar{z} = \sum_i (f_i z_i). \tag{13}$$

where, for simplicity, we have momentarily used explicit indexing of multiple alleles simultaneously segregating in the population at a given time.

Relying on a tiling argument (Latrille *et al.*, 2017 [1]) this can also be expressed as :

$$\bar{z} = \frac{1}{\tau} \int_0^{+\infty} f(t) z(t) \, dt. \tag{14}$$

Equation (14) expresses the idea that, at stationarity, the allele frequency distribution at a given time point (eq. (13)) is equivalent to the distribution of frequencies at which a typical allele has segregated over its entire life normalized by the mean waiting time τ between successive invasions (see Latrille *et al.*, 2017 figure 6 [1]).

We then do a change of variable from t to z in the integrand:

$$\bar{z} = \frac{1}{\tau} \int_0^{z(\infty)} z f \frac{dt}{dz} \, dz = \frac{1}{\tau} \int_0^{z(\infty)} z f \left(\frac{dz}{dt} \right)^{-1} dz, \tag{15}$$

then, we replace $\frac{dz}{dt}$ by its expression in equation (2) :

$$\bar{z} = \frac{1}{\tau} \int_0^{z(\infty)} z f \frac{1}{f\rho} \, dz = \frac{1}{\tau\rho} \int_0^{z(\infty)} z \, dz = \frac{1}{\tau\rho} \frac{z(\infty)^2}{2}. \tag{16}$$

Yet, we know that the representative curve of $z(t)$ is a concave parabola with $z(0) = 0$ and therefore $z(\infty) = 2\bar{z}$. Thus, if we repalace it in 16, we obtain :

$$\bar{z} = \frac{1}{\tau\rho} \frac{(2\bar{z})^2}{2} = \frac{2\bar{z}^2}{\tau\rho} \tag{17}$$

which gives, by isolation of $\bar{z}$:

$$\bar{z} = \frac{\rho\tau}{2}. \tag{18}$$

We now need to express τ, which is the inverse of the invasion rate of a new allele in the population. The rate of invasion is equal to the rate of mutation at the population level ($2Nu$) multiplicated by the invasion probability. Assuming strong selection, this probability is well approximated by $2s_0$.

$$\tau^{-1} = (2Nu).(2s_0) \tag{19}$$

where u is the mutation rate at the *PRDM9* locus and s_0 is the selection coefficient of a new allele in the population. Based on equation (9), s_0 can be expressed as $s_0 = \frac{\alpha}{2}\bar{z}$. Thus :

$$\tau^{-1} = 4Nu\frac{\alpha}{2}\bar{z} = \mu\frac{\alpha}{2}\bar{z}, \tag{20}$$

Where $\mu = 4Nu$. We can also express *tau* in function of N, u, v, g and α (eq 6 in the main text) as :

$$\tau = \frac{1}{2N\sqrt{u\alpha vg}}. \tag{21}$$

If we replace this in equation (18) we finally obtain :

$$\bar{z} = \frac{\rho}{2\tau^{-1}} = \frac{\rho}{2}\frac{2}{\mu\alpha\bar{z}} = \frac{\rho}{\mu\alpha\bar{z}} \iff \bar{z}^2 = \frac{\rho}{\mu\alpha} \iff \boxed{\bar{z} = \sqrt{\frac{\rho}{\mu\alpha}}} \tag{22}$$

which can also be expressed in function of v, g, u and α (eq 4 in the main text) as :

$$\boxed{\bar{z} = \sqrt{\frac{vg}{u\alpha}}} \tag{23}$$

We thus recover the main result of the derivation given in Latrille *et al.* [1], although now, the compound parameters ρ and α depend on the mecanistic details of our model. We have already seen that $\rho = \frac{Nvd}{2h}$. On the other hand, we still need to express α.

B.5   Slope at the origin of the fertility rate : α

As mentioned above, α is the slope of $ln(w)$ at the origin (*i.e.* for two new alleles of age $z_1 = z_2 = 0$):

$$\alpha = \left| \frac{\partial \ln(w(z_1, z_2))}{\partial z_1} \right|_{(z_1=0, z_2=0)} = \left| \frac{1}{w}\frac{\partial w(z_1, z_2)}{\partial z_1} \right|_{(z_1=0, z_2=0)} = \left| \frac{\partial w(z_1, z_2)}{\partial z_1} \right|_{(z_1=0, z_2=0)}, \tag{24}$$

since $w \approx 1$.

Thus, in order to find an explicit expression for α, we need to express w according to the parameters.

Assuming that gametes are not limiting, the fitness of an individual is equal to the rate of success of meiosis, which is itself equal to the probability of having at least one DSB in a symmetrical bound site. So, we can write w as 1 - the probability of having no DSB in symmetrical bound site (1 - probability of failure of the meiosis). The number of DSBs in symmetrical bound sites is approximately Poisson of mean $dq(z_1, z_2)$, where d is the mean number of DSBs and $q(z_1, z_2)$ is the probability that a DSB occurs in a symmetrically bound site in an individual heterozygous for two *PRDM9* alleles of age z_1 and z_2. Thus, the probability of 0 DSB in a symmetrically bound site is $e^{-dq(z_1, z_2)}$. So, $w(z_1, z_2)$ can be expressed as :

$$w(z_1, z_2) = 1 - e^{-dq(z_1, z_2)} \tag{25}$$

Substituting in equation (24), we obtain :

$$\alpha = \left| \frac{\partial w(z_1, z_2)}{\partial z_1} \right|_{(z_1=0, z_2=0)} = \left| d\, \frac{\partial q(z_1, z_2)}{\partial z_1} \right|_{(z_1=0, z_2=0)} e^{-dq(z_1=0, z_2=0)} \right| \tag{26}$$

We see here that, in order to obtain an explicit formula for α, it is necessary to express $q(z_1, z_2)$ as a function of the model parameters and then compute its derivative as a function of z.

B.6   Probability of symmetrical binding : q

For a given site of affinity y, the probability that PRDM9 is bound is given by $x = \frac{cy}{1+cy}$ where $c = 1$ or $c = 2$. The conditional probability of symmetrical binding at that site (conditional on at least one of the four chromatids being bound) is then equal to :

$$q = \frac{2x^2 - x^3}{x} \tag{27}$$

5

This expression is valid for a single site. To compute the mean probability of symmetrical binding over the genome, we need to average the numerator and the denominator separately over the affinity distribution across sites. Of note this distribution itself depends on the age z of the allele, and the mean over the distribution of a given function $B(y)$ is noted $<B>_z$. Also, the mean q over the genome depends on the *PRDM9* genotype of this individual. In the case of a homozygote, thus possessing twice the same *PRDM9* allele of age z, q is equal to:

$$q^{hom}(z) = \frac{2<x^2>_z - <x^3>_z}{<x>_z} \tag{28}$$

If the individual is heterozygous for *PRDM9* with two alleles of age z_1 and z_2, q is equal to:

$$q^{het}(z_1, z_2) = \frac{2<x^2>_{z_1} - <x^3>_{z_1} + 2<x^2>_{z_2} - <x^3>_{z_2}}{<x>_{z_1} + <x>_{z_2}} \tag{29}$$

The mean over the affinity distribution can be more precisely expressed as :

$$<B>_z = \int B(y)\theta_y(z)\varphi(y)\,dy. \tag{30}$$

In this equation $\varphi(y)$ is the affinity distribution of an allele at birth and $\theta_y(z)$ is the fraction of active sites with a given affinity y recognised by an allele of age z. Thus, $\theta_y(z)\varphi(y)dy$ is the total number of target sites still active with an affinity $y^{\pm}dy$.

The integrals of the form (30) can be obtained numerically. However, they depend on $\theta_y(z)$, which we therefore need to determine.

B.7 Fraction of active sites with a given affinity y recognised by an allele of an age z : $\theta_y(z)$

Here our aim is to compute more precisely the proportion of target sites of a given affinity y that are still active, for an allele of age z. We note this quantity $\theta_y(z)$

By an argument similar to that used for equation (3):

$$\frac{d\theta_y}{dt} = -(2Nv).(2fg_y^{het})\theta_y \tag{31}$$

where $g_y^{het} = g_y^{het}(z_1, z_2)$ is the gene conversion rate at sites of affinity y in a genotype (z_1, z_2). Of note, we consider only heterozygotes for *PRDM9* since we work under the assumption of a highly polymorphic regime. In turn :

$$g_y^{het} = d.\frac{\frac{cy}{1+cy}}{4h\left[\left\langle\frac{cy}{1+cy}\right\rangle_{z_1} + \left\langle\frac{cy}{1+cy}\right\rangle_{z_2}\right]} \tag{32}$$

The fraction on the right-hand side of equation (32) is the proportion of sites of affinity y among all sites that are bound by either one of the two *PRDM9* alleles.

So, if we replace it in equation (31), we obtain :

$$\frac{d\theta_y}{dt} = -2Nv.d\frac{\frac{cy}{1+cy}}{4h\left[\left\langle\frac{cy}{1+cy}\right\rangle_{z_1} + \left\langle\frac{cy}{1+cy}\right\rangle_{z_2}\right]}.2f(t)\theta_y(z) = -Nv\frac{d}{h}.\frac{\frac{cy}{1+cy}}{\left[\left\langle\frac{cy}{1+cy}\right\rangle_{z_1} + \left\langle\frac{cy}{1+cy}\right\rangle_{z_2}\right]}.f(t)\theta_y(z). \tag{33}$$

If we suppose that $< \frac{cy}{1+cy} >_{z_1}$ and $< \frac{cy}{1+cy} >_{z_2}$ don't vary too much as a function of z ($\approx < \frac{cy}{1+cy} >_0$), then we have:

$$\frac{d\theta_y}{dt} = -Nv\frac{d}{h}\frac{\frac{cy}{1+cy}}{2\left\langle \frac{cy}{1+cy} \right\rangle_0}.f(t)\theta_y(z) \tag{34}$$

Then, by setting $\rho = \frac{Nvd}{2h}$ we can re-express the equation (37) as

$$\frac{d\theta_y}{dt} = -\rho\frac{\frac{cy}{1+cy}}{\left\langle \frac{cy}{1+cy} \right\rangle_0}.f(t)\theta_y(z) \tag{35}$$

And at this stage we can pose $\gamma(y) = \frac{\frac{cy}{1+cy}}{\left\langle \frac{cy}{1+cy} \right\rangle_0}$, which gives the following equation:

$$\frac{d\theta_y}{dt} = -\rho f(t)\gamma(y)\theta_y. \tag{36}$$

Now, we can replace $\rho f(t)$ by $\frac{dz}{dt}$:

$$\frac{d\theta_y}{dt} = -\frac{dz}{dt}\gamma(y)\theta_y \Leftrightarrow \frac{d\theta_y(z)}{dz} = -\gamma(y)\theta_y \tag{37}$$

Finally if we integrate this equation, we obtain the fraction of active sites with an affinity y recognized by an allele of age z :

$$\theta_y(z) = \theta_y(0)e^{-\gamma(y)z} \tag{38}$$

By definition, $\theta_y(0)$ is the fraction of active sites with an affinity y recognized by an allele of age 0. Thus $\theta_y(0) = 1$, and as a result, equation (34) becomes

$$\boxed{\theta_y(z) = e^{-\gamma(y)z}} \tag{39}$$

Under the condition of weak erosion, we can simplify equation (39) which gives

$$\theta_y(z) \approx 1 - \gamma(y)z \tag{40}$$

and by averaging over the affinity, we get

$$< \theta_y(z) > \approx 1 - z \tag{41}$$

B.8 Expression of α

Thanks to this equation, we are now able to express the successive moments $< x^m >_z$ for $m = 1, 2$ and 3.

$$< x^m >_z = \int \left(\frac{cy}{1+cy}\right)^m \theta_y(z)\varphi(y)\,dy = \int \left(\frac{cy}{1+cy}\right)^m e^{-\gamma(y)z}\varphi(y)\,dy \tag{42}$$

To express α, we need $q(0,0)$ and $\frac{\partial q(z_1,z_2)}{\partial z_1}\Big|_{z_1=0,z_2=0}$.

For $q(0,0)$ we have the following expression

$$q(0,0) = \frac{2 < x^2 >_0 - < x^3 >_0 + 2 < x^2 >_0 - < x^3 >_0}{< x >_0 + < x >_0} = \frac{2 < x^2 >_0 - < x^3 >_0}{< x >_0} \tag{43}$$

with $< B >_0 = \int B(y)\varphi(y)$.

And for $\frac{\partial q(z_1, z_2)}{\partial z_1}\Big|_{z_1=0, z_2=0}$, we first need to determine the general expression of $\frac{\partial <B>_z}{\partial z}$ and then express all the moments $<x^n>$ that we need.

$$\frac{\partial <B>_z}{\partial z} = \int B(y)\varphi(y)(-\gamma(y))e^{\gamma(y)z}\, dy = -<\gamma(y)B>_z \tag{44}$$

And with $\gamma(y) = \frac{\frac{cy}{1+cy}}{\left\langle \frac{cy}{1+cy}\right\rangle_0} = \frac{x}{<x>_0}$ we obtain :

$$\frac{\partial <B>_z}{\partial z} = -\frac{<xB>_z}{<x>_0} \tag{45}$$

So

$$\frac{\partial <B>_z}{\partial z}\Bigg|_{z=0} = -\frac{<xB>_0}{<x>_0} \tag{46}$$

In particular, with $B = x^n$ we have

$$\frac{\partial <x^n>_z}{\partial z}\Bigg|_{z=0} = -\frac{<x^{n+1}>_0}{<x>_0} \tag{47}$$

If we come back to the expression of $\frac{\partial q(z_1, z_2)}{\partial z_1}\Big|_{z_1=0, z_2=0}$, we obtain

$$\beta = \frac{\partial q(z_1, z_2)}{\partial z_1}\Bigg|_{z_1=0, z_2=0} = \frac{\left[-2\frac{<x^3>_0}{<x>_0} + \frac{<x^4>_0}{<x>_0}\right]2<x>_0 + \frac{<x^2>_0}{<x>_0}(4<x^2>_0 - 2<x^3>_0)}{4<x>_0^2} \tag{48}$$

Replacing in the equation (26) and after some simplifications, we obtain the final expression of α

$$\alpha = \left| d\left[\frac{-2<x^3>_0 + <x^4>_0 + \frac{2<x^2>_0^2}{<x>_0} - \frac{<x^3>_0<x^2>_0}{<x>_0}}{2<x>_0^2}\right]e^{-d\frac{2<x^2>_0 - <x^3>_0}{<x>_0}}\right| \tag{49}$$

which can also be written as follows

$$\boxed{\alpha = \left|d\beta e^{-dq}\right|} \tag{50}$$

C   Summary statistics

Based on these analytical developments, we can now obtain analytical expressions for the summary statistics.

C.1   Diversity : D

The *PRDM9* diversity, written D, is defined as

$$D = \left(\sum_i f_i^2\right)^{-1} \tag{51}$$

We now need to calculate $D^{-1} = \sum_i f_i^2$. For that, we can use again the tiling principle :

$$D^{-1} \approx \frac{1}{\tau}\int_0^\infty f^2(t)\, dt = \frac{1}{\tau}\int_0^{z(\infty)} f^2(z)\frac{dt}{dz}\, dz = \frac{1}{\tau}\int_0^{z(\infty)} f^2(z)\frac{1}{\rho f(z)}\, dz = \frac{1}{\rho\tau}\int_0^{z(\infty)} f(z)\, dz \tag{52}$$

Replacing f(z) by its expression from equation (12) and integrating over z gives :

$$\frac{1}{\rho\tau}\int_0^{z(\infty)} -\frac{\alpha}{4\rho}z(z-2\bar{z})\,dz = -\frac{\alpha}{4\rho^2\tau}\left[\int_0^{z(\infty)} z^2\,dz - 2\bar{z}\int_0^{z(\infty)} z\,dz\right] - \frac{\alpha}{4\rho^2\tau}\frac{z(\infty)^3}{3} + \frac{\alpha\bar{z}}{2\rho^2\tau}\frac{z(\infty)^2}{2}$$

(53)

But we know that $\bar{z} = \frac{\rho\tau}{2}$ and $z(\infty) = \rho\tau$. So if we replace it in the equation (53)

$$-\alpha\left[\frac{\rho^3\tau^3}{12\rho^2\tau} - \frac{\rho^3\tau^3}{8\rho^2\tau}\right] = -\alpha\left[\frac{\rho\tau^2}{12} - \frac{\rho\tau^2}{8}\right] = -\alpha\left[\frac{2\rho\tau^2}{24} - \frac{3\rho\tau^2}{24}\right]$$

(54)

So we obtain

$$D^{-1} = \frac{\alpha\rho\tau^2}{24}$$

(55)

But, we also know that $\tau = \frac{2}{\mu\alpha\bar{z}}$ and $\bar{z} = \sqrt{\frac{\rho}{\mu\alpha}}$. So then $\tau = \frac{2}{\mu\alpha}\sqrt{\frac{\mu\alpha}{\rho}}$ and $\tau^2 = \frac{4}{\mu\alpha\rho}$. Which leads to the equation

$$D^{-1} = \frac{4\alpha\rho}{24\mu\alpha\rho} = \frac{1}{6\mu}$$

(56)

Where $\mu = 4Nu$, so finally we have

$$D^{-1} = \frac{1}{24Nu}$$

(57)

And thus

$$\boxed{D = 24Nu}$$

(58)

C.2   Mean age : $\bar{z}$

Using equation (50), $\bar{z}$ can be more directly expressed as :

$$\boxed{\bar{z}} = \sqrt{\frac{\rho}{\mu\alpha}} = \sqrt{\frac{Nvd}{2h}\frac{1}{4Nu}\frac{e^{dq}}{d\beta}} = \boxed{\sqrt{\frac{ve^{dq}}{8hu\beta}}}$$

(59)

with q given by the equation (43)

C.3   Mean activity : $<\bar{\theta}>$

If we start from the equation (39) replacing z by $\bar{z}$ we have :

$$\overline{\theta_y(z)} = \theta_y(\bar{z}) = e^{-\gamma(y)\bar{z}} \approx 1 + (-\gamma(y)\bar{z}) \approx 1 - \gamma(y)\bar{z}$$

(60)

Averaging over the affinity distribution :

$$\boxed{<\theta_y(\bar{z})> \approx 1 - \bar{z}}$$

(61)

C.4   Mean probability of symmetrical binding : $\bar{q}$

If we start from the equation (29) replacing z by $\bar{z}$ we have :

$$\boxed{\bar{q}} = q^{het}(\bar{z}, \bar{z}) = \frac{2<x^2>_{\bar{z}} - <x^3>_{\bar{z}} + 2<x^2>_{\bar{z}} - <x^3>_{\bar{z}}}{<x>_{\bar{z}} + <x>_{\bar{z}}} = \boxed{\frac{2<x^2>_{\bar{z}} - <x^3>_{\bar{z}}}{<x>_{\bar{z}}}}$$

(62)

The moments $<x^m>_{\bar{z}}$ are given by equation (42) an can be evaluated by numerical integration.

C.5 Mean fertility rate : $\bar{w}$

If we start from the equation (25) replacing z by $\bar{z}$ we have :

$$\boxed{\bar{w} = w(\bar{z}, \bar{z}) = 1 - e^{-dq(\bar{z}, \bar{z})}} \tag{63}$$

With $\bar{q} = \frac{2<x^2>_{\bar{z}} - <x^3>_{\bar{z}}}{<x>_{\bar{z}}}$.

D Perturbative development accounting for genetic dosage

The evolution of the frequency of an allele in the population as a function of its age has the same general expression as without dosage :

$$\frac{df}{dz} = \frac{1}{\rho}\left(\frac{w^*(z) - \bar{w}}{\bar{w}}\right) \tag{64}$$

However, for the fitness, we now account for the contribution of homozygotes:

$$w^*(z) = f(t)w^{hom}(z) + (1 - f(t))w^{het}(z, \bar{z}) \tag{65}$$

Linearizing in the vicinity of 0 for z:

$$w^*(z) = fw^{hom}(z) + (1-f)w^{het}(z, \bar{z}) \approx fw^{hom}(0)(1 - \alpha^{hom}z) + (1-f)w^{het}(0,0)(1 - \frac{\alpha^{het}}{2}(z + \bar{z})) \tag{66}$$

Our development is perturbative in the sense that it assumes that gene dosage has a weak impact, and this, because homozygotes are assumed to be rare, i.e. because f is small ($f << 1$). Combined with the weak erosion assumption ($\bar{z} << 1$), this means that we can ignore terms of the order of zf. Thus, equation (66) simplifies to :

$$w^*(z) \approx fw^{hom}(0) + (1-f)w^{het}(0,0) - w^{het}(0,0)\frac{\alpha^{het}}{2}(z + \bar{z}) \tag{67}$$

Averaging over the population gives the mean fitness:

$$\bar{w} \approx \bar{f}w^{hom}(0) + (1-\bar{f})w^{het}(0,0) - w^{het}(0,0)\frac{\alpha^{het}}{2}(\bar{z} + \bar{z}) \tag{68}$$

Finally :

$$\frac{df}{dz} = \frac{1}{\rho}\left(\frac{w^*(z) - \bar{w}}{\bar{w}}\right) \approx \frac{1}{\rho}\left[\frac{w^{hom}(0) - w^{het}(0,0)}{w^{het}(0,0)}(f - \bar{f}) - \frac{\alpha^{het}}{2}(z - \bar{z})\right] \tag{69}$$

If we express

$$\sigma(0) = \frac{w^{hom}(0) - w^{het}(0,0)}{w^{het}(0,0)} \tag{70}$$

we obtain :

$$\frac{df}{dz} \approx \frac{1}{\rho}\left[\sigma(0)(f - \bar{f}) - \frac{\alpha^{het}}{2}(z - \bar{z})\right] \tag{71}$$

Or, if we express this in terms of $\frac{dln(f)}{dt}$, it gives :

$$\boxed{\frac{d\ln f}{dt} \simeq -\frac{\alpha}{2}(z - \bar{z}) + \sigma_0(f - \bar{f})} \tag{72}$$

as it is expressed in eq 3 of the main text.

References

[1] T. Latrille, L. Duret, and N. Lartillot. "The Red Queen model of recombination hot-spot evolution : a theoretical investigation." In: *Philosophical Transactions of the Royal Society B : Biological Sciences* 372 (2017), p. 20160463. DOI: `http://dx.doi.org/10.1098/rstb.2016.0463`.