The Red-Queen model of recombination hot-spot evolution: a theoretical investigation.

Thibault Latrille, Laurent Duret, Nicolas Lartillot

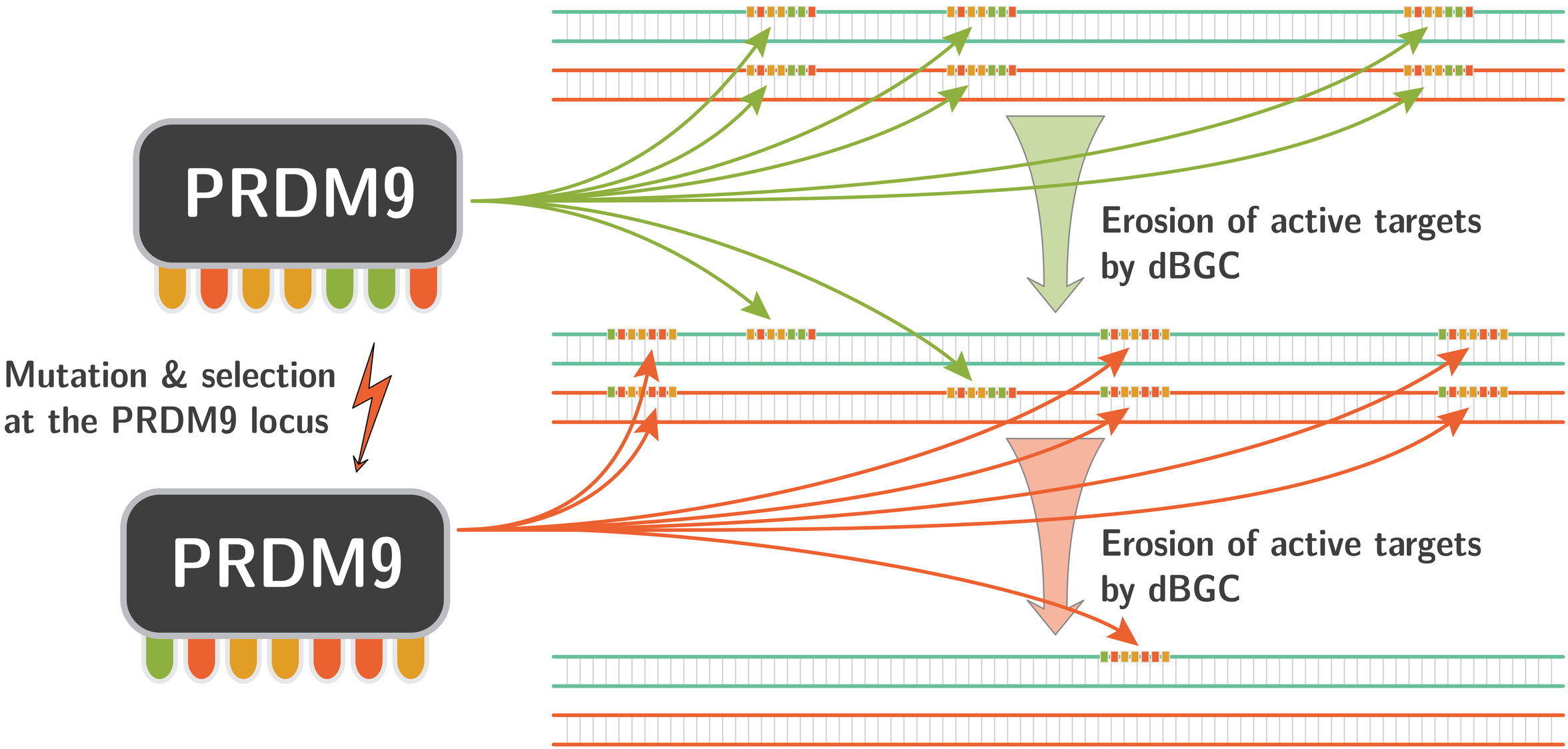
Laboratoire de Biométrie et Biologie Évolutive (LBBE), UMR CNRS 5558, Lyon

March 29, 2017

ANR Dasire meeting, Montpellier

The Red Queen of recombination

- dBGC results in accumulation of inactive hot spots.
- Mutation and positive selection at the PRDM9 locus leads to recruitment of new hot spots.



From a qualitative to a quantitative model of the Red Queen

- dBGC destroy active hot spots, new PRDM9 recruits new hot spots.
- Intra-genomic conflict, two processes are chasing after one another.
- PRDM9 locus is known to be highly polymorphic.

How does the Red Queen unfolds in the polymorphic case?

What is the influence of positive selection on PRDM9 diversity?

What is the results of changes in population size on hot spots activity? On PRDM9 diversity? On turn-over rate?

How to quantify the effects of dBGC and mutation rate of PRDM9 on the Red Queen dynamic?

A population genetic model of the Red Queen of recombination

1. Wright-Fisher simulator: mutation, erosion & selection.
2. Results of simulations: mean activity & diversity.
3. Succession or polymorphic regime?
4. Analytical approximation in succession and polymorphism.
5. Empirical calibration of the model.

Mutation at the PRDM9 locus

1. The locus of PRDM9 mutates at constant rate u per generation.
2. Each mutation produces a new functional PRDM9 allele.
3. The number of targets is the same for each allele.
4. There is no overlap between the targets of distinct PRDM9 alleles.
5. $K(t)$ denotes the number of PRDM9 alleles in the population.
6. $n_i(t)$ is the number of copies of the i^{th} allele in the population.
7. $\boldsymbol{x_i(t)} = n_i(t)/2N_{\mathrm{e}}$ is the frequency of allele i at time t.

Erosion at the targets of PRDM9

1. The recombination activity induced by an allele is maximal at the birth of this allele .
2. Erosion is modelled implicitly, by tracking over time the fraction of active targets associated with each allele, $r_i(t)$.
3. v is the mutation rate at the target sites.
4. The rate of inactivating mutations per target at the level of the population is $2N_e v$
5. g is the rate of conversion of active targets by the inactive mutant in an heterozygous individual.
6. Under strong dBGC, the fixation probability of inactive mutant equal to $2g x_i(t)$ for i^{th} allele of PRDM9.

Altogether, the activity induced by allele i decays as:

$$\frac{\mathrm{d}r_i(t)}{\mathrm{d}t} = -\rho x_i(t) r_i(t), \text{ where } \rho = 4N_e vg$$

Selection at the PRDM9 locus

1. The fitness of an individual with genotype (i, j) is:

$$\omega_{i,j}(t) = f\left(\frac{r_i(t) + r_j(t)}{2}\right)$$

2. f is assumed to be an increasing function, $f(x) = x^{\alpha}$.
3. The average fitness induced by allele i over the population is then

$$\omega_i(t) = \sum_{j=1}^{K_t} x_j(t)\omega_{i,j}(t)$$

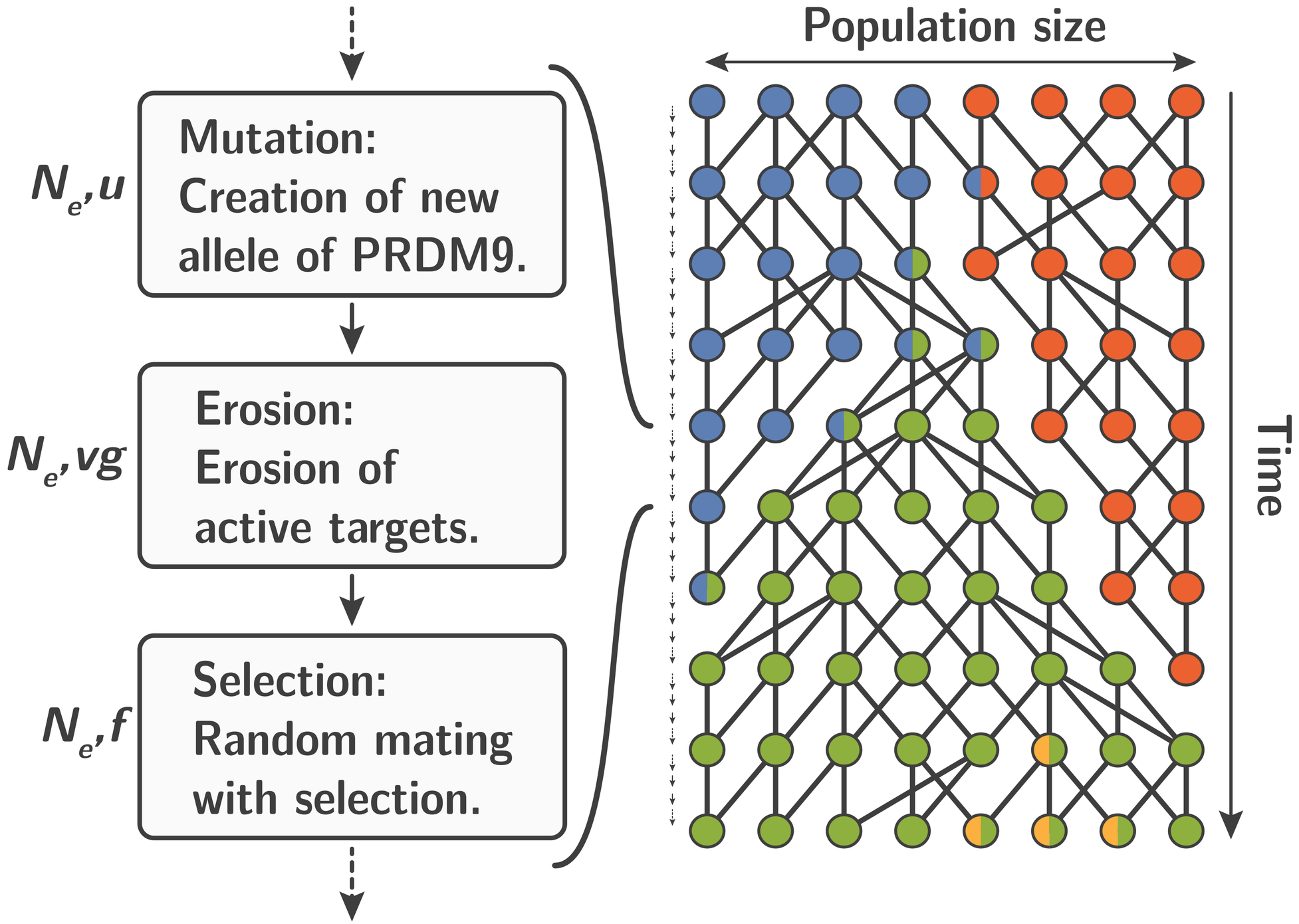
4. The mean fitness over the population is

$$\overline{\omega(t)} = \sum_{i=1}^{K_t} x_i(t)\omega_i(t)$$

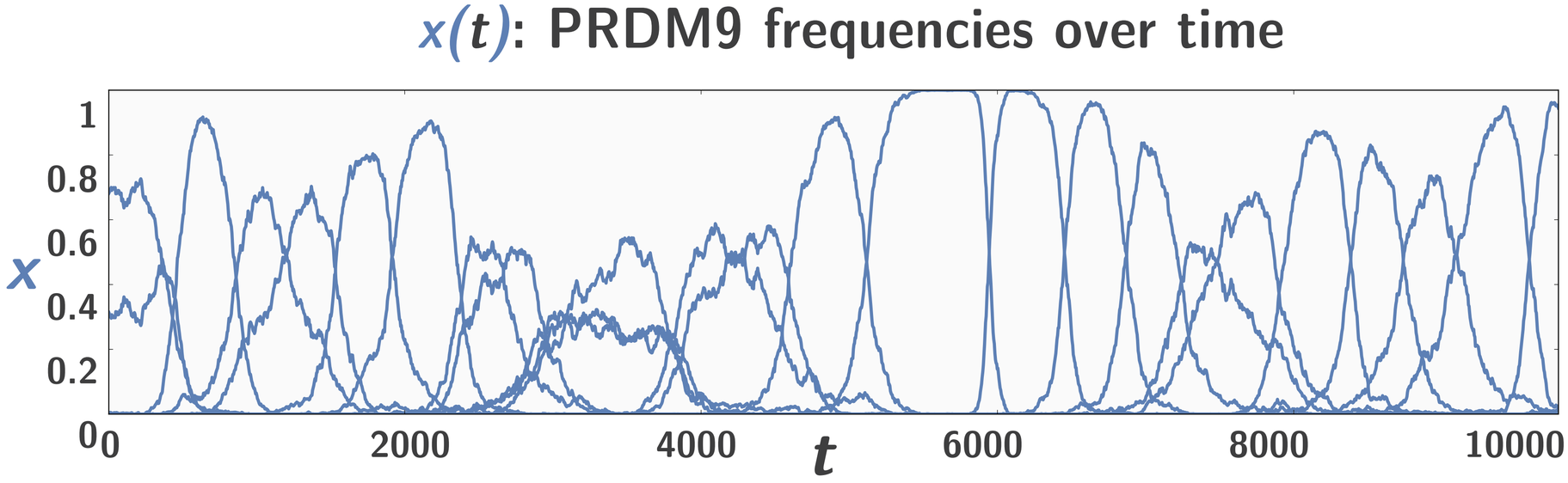
5. The probability for the i^{th} allele to be picked up at the next generation is:

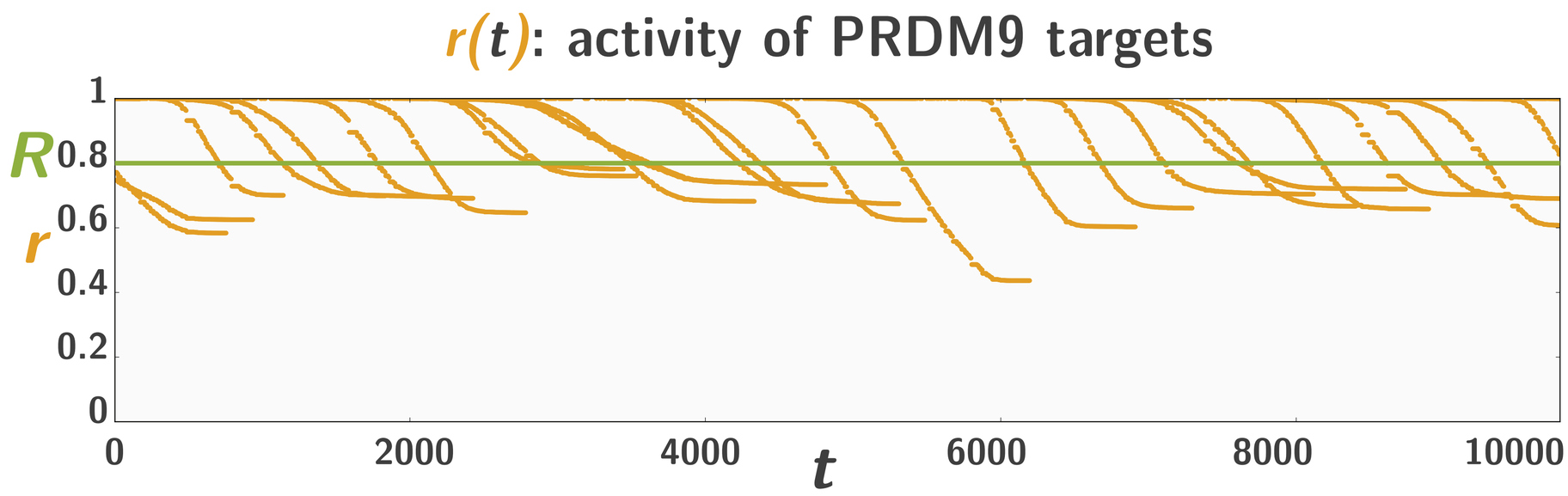
$$p_i(t+1) = x_i(t)\frac{\omega_i(t)}{\overline{\omega(t)}}$$

Wright-Fisher simulator at the PRDM9 locus



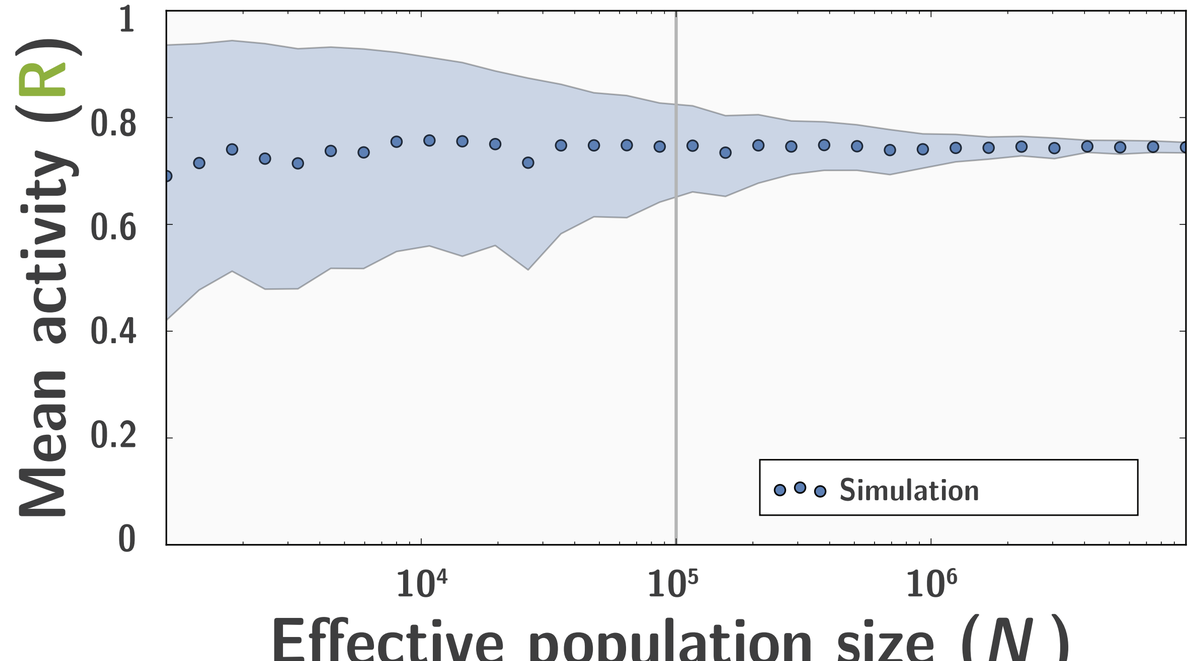
Trajectory of a typical simulation

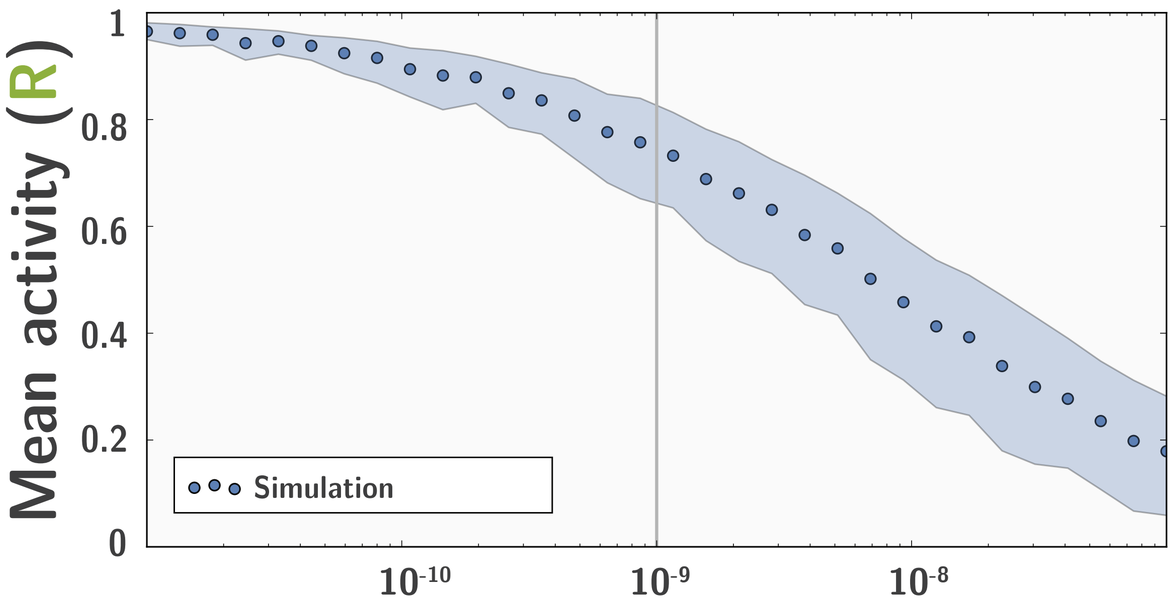


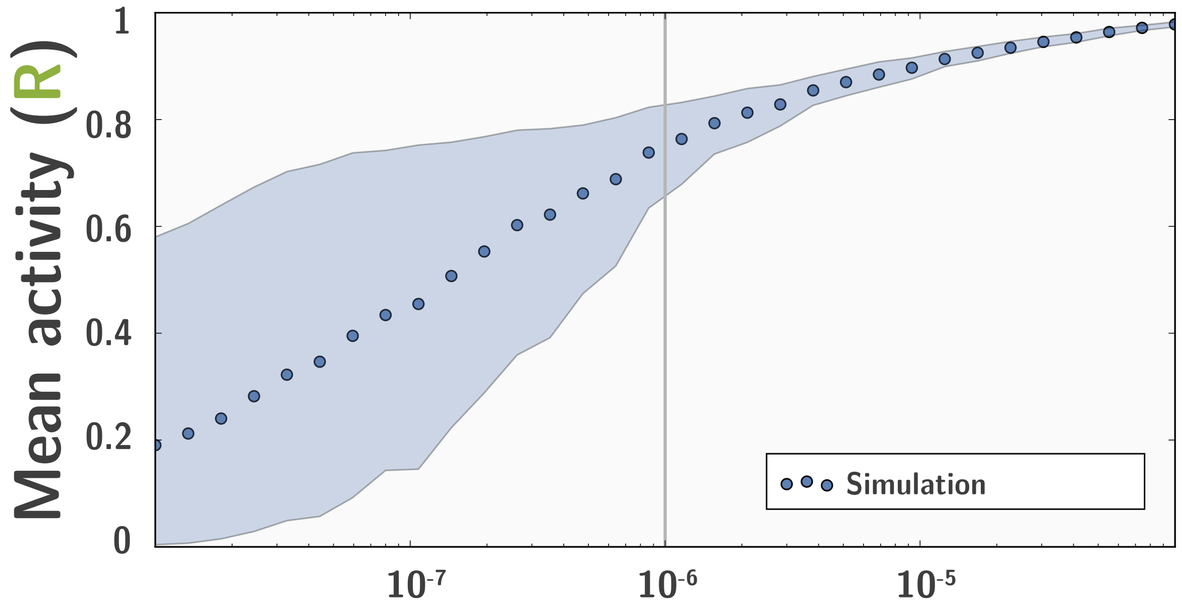


Mean activity of the targets as function of parameters of the model

$$R = \left\langle \sum_i x_i(t) r_i(t) \right\rangle$$

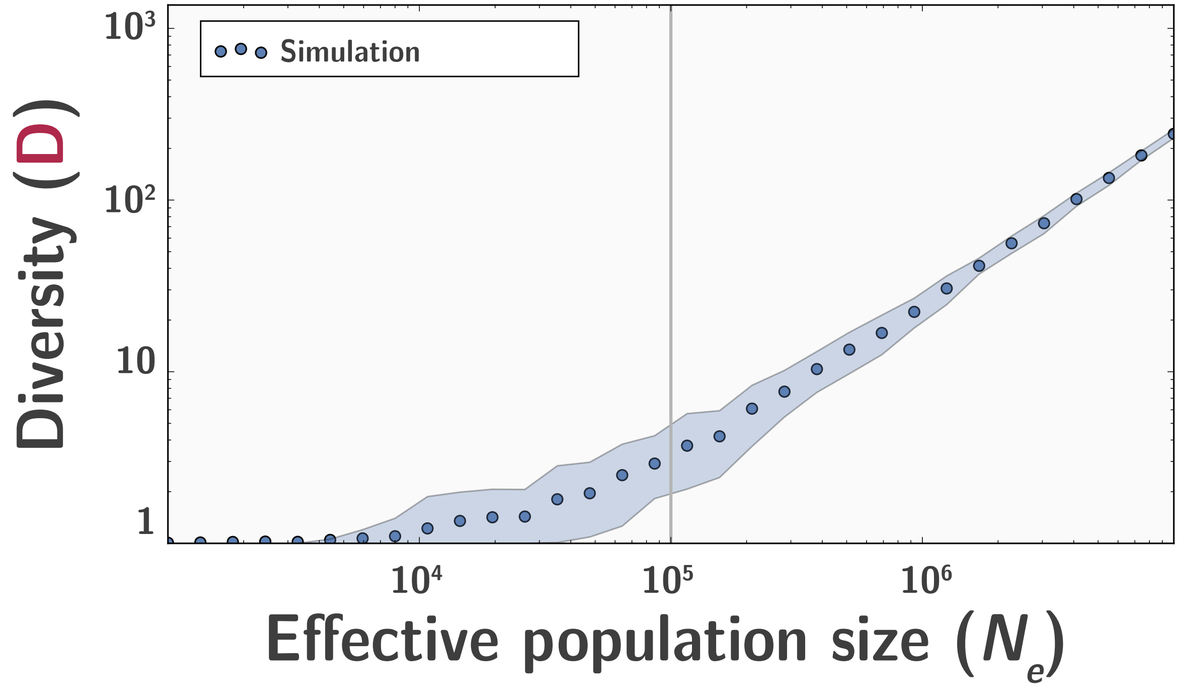


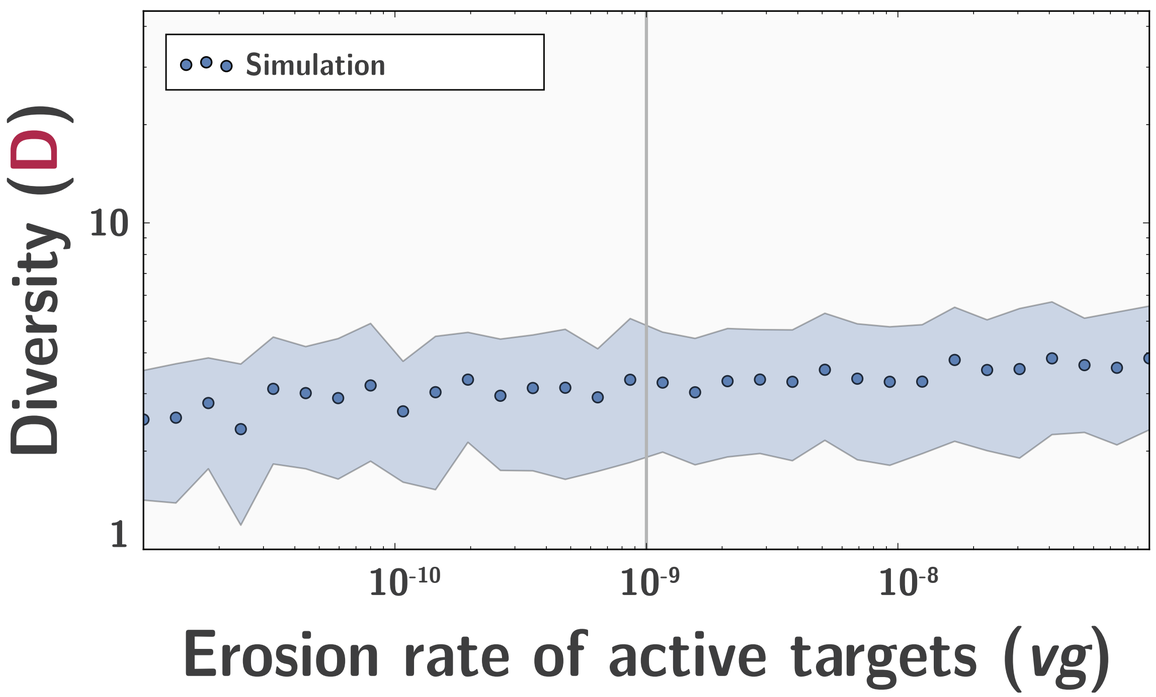


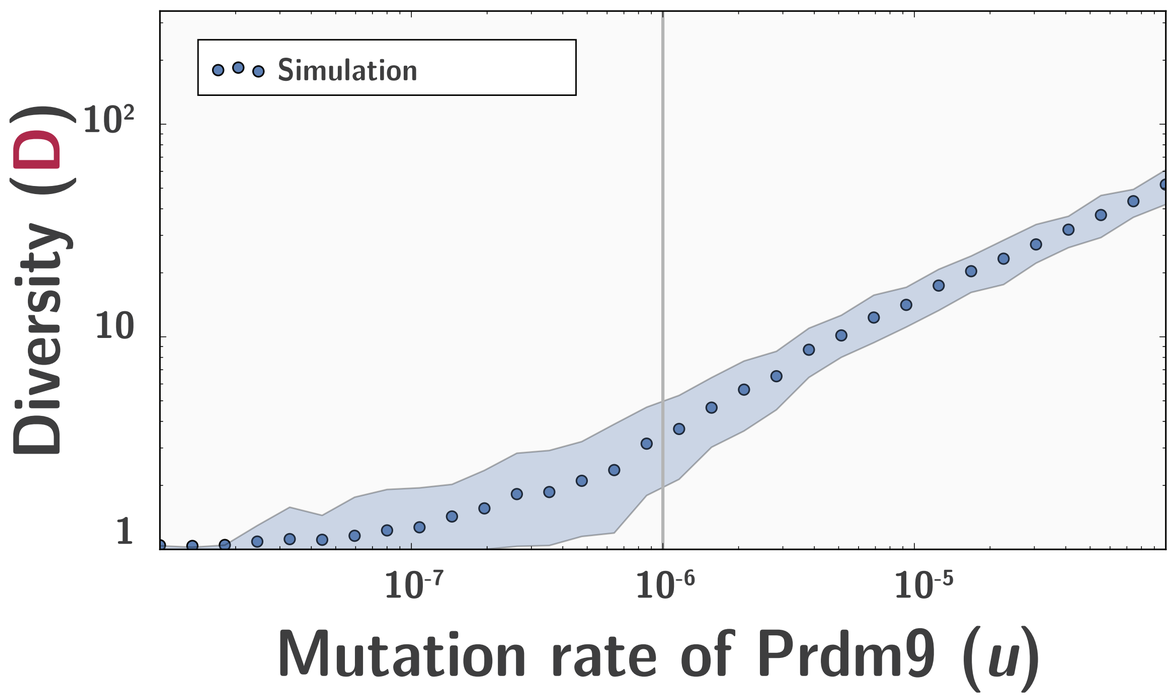


Diversity at the PRDM9 locus as function of parameters of the model

$$D = \left\langle \frac{1}{\sum_i x_i(t)^2} \right\rangle$$







Succession or polymorphism of PRDM9?

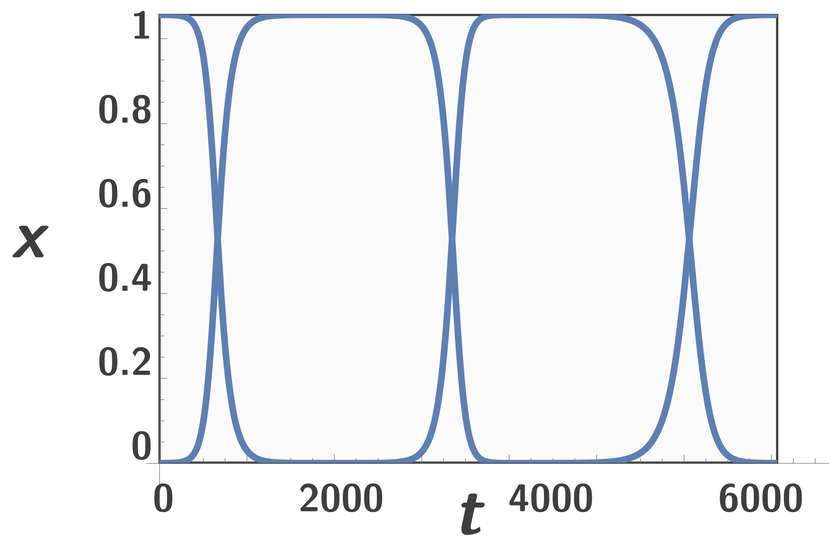
N_e, u, vg, f

$N_e u$

The expected number of new PRDM9 alleles in the population at each generation.

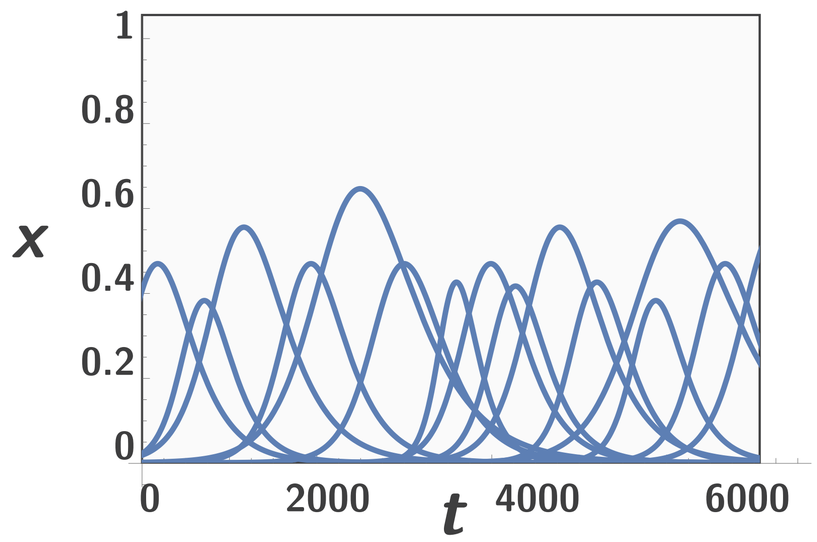
$N_e u \ll 1$

Succession of PRDM9 alleles



$N_e u \gg 1$

Polymorphism of PRDM9 alleles



Self-consistent equation in succession regime

1. $r(t)$ is the activity of targets for the current PRDM9 allele.

$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = -\rho r(t) \Rightarrow r(t) = e^{-\rho t}, \text{ where } \rho = 4N_e vg$$

2. τ the mean time between two successive invasions.

$$\Rightarrow R = \frac{1}{\tau}\int_0^{\tau} r(t)\mathrm{d}t = \frac{1}{\tau}\int_0^{\tau} e^{-\rho t}\mathrm{d}t = \frac{1-e^{-\rho\tau}}{\rho\tau}$$

3. s_0 is the selection coefficient experienced by a new allele

$$s_0 \simeq \frac{f'(r(t))}{f(r(t))}\frac{1-r(t)}{2} \simeq \frac{f'(R)}{f(R)}\frac{1-R}{2}$$

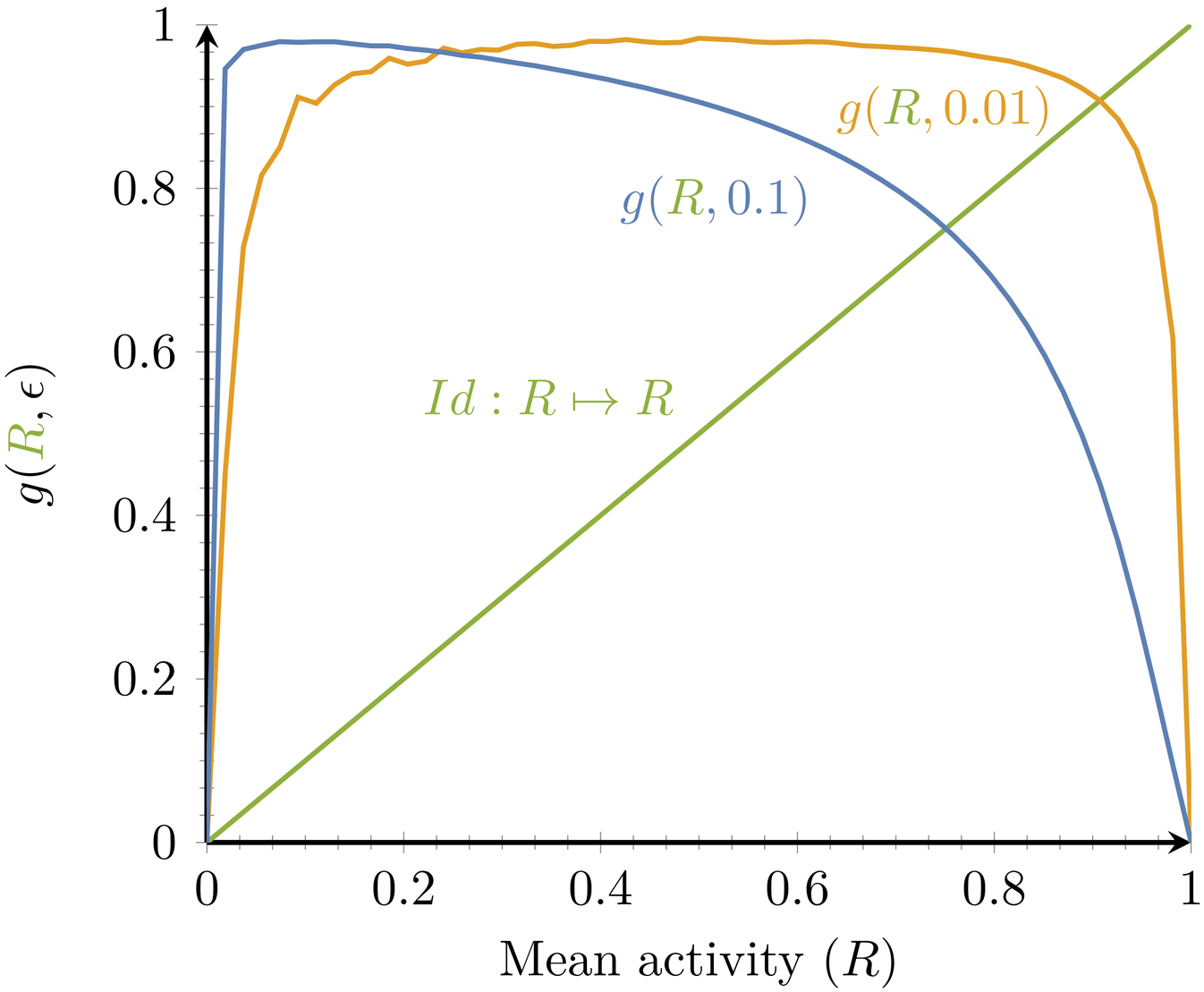
4. τ is also the inverse of the invasion rate:

$$\tau = \frac{1}{\mu s_0} \simeq \frac{1}{\mu}\frac{f(R)}{f'(R)}\frac{2}{1-R}, \text{ where } \mu = 4N_e u$$

5. Altogether,

$$R = g\left(R, \frac{\rho}{\mu}\right) = g\left(R, \frac{vg}{u}\right)$$

Solving numerically the self-consistent equation



Self-consistent solution as a fixed point, in the form $R = g(R, \epsilon)$

Succession or polymorphism of PRDM9?

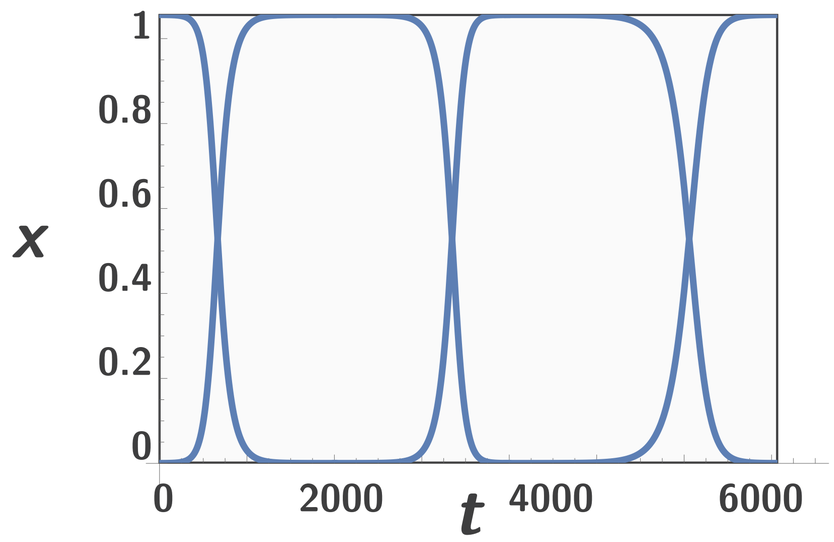
N_e, u, vg, f

$N_e u$

The expected number of new PRDM9 alleles in the population at each generation.

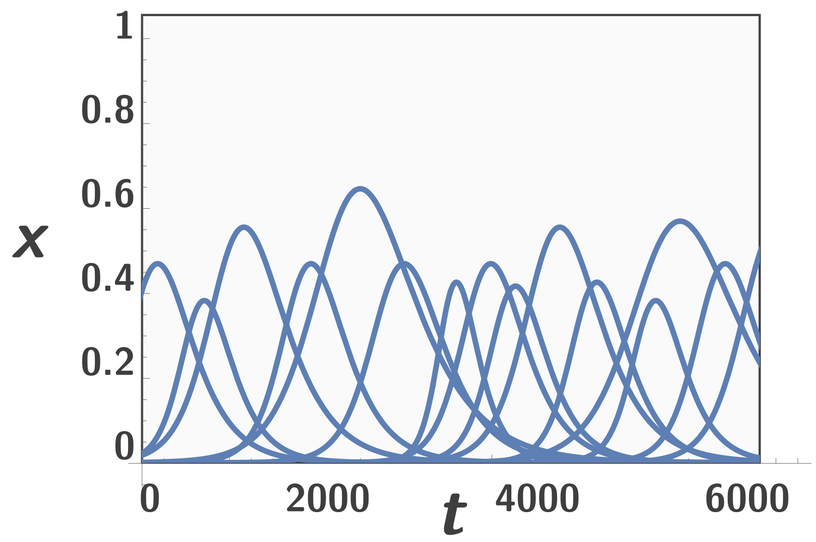
$N_e u \ll 1$

Succession of PRDM9 alleles



$N_e u \gg 1$

Polymorphism of PRDM9 alleles



Decoupling the equations: Mean field approximation in polymorphic regime

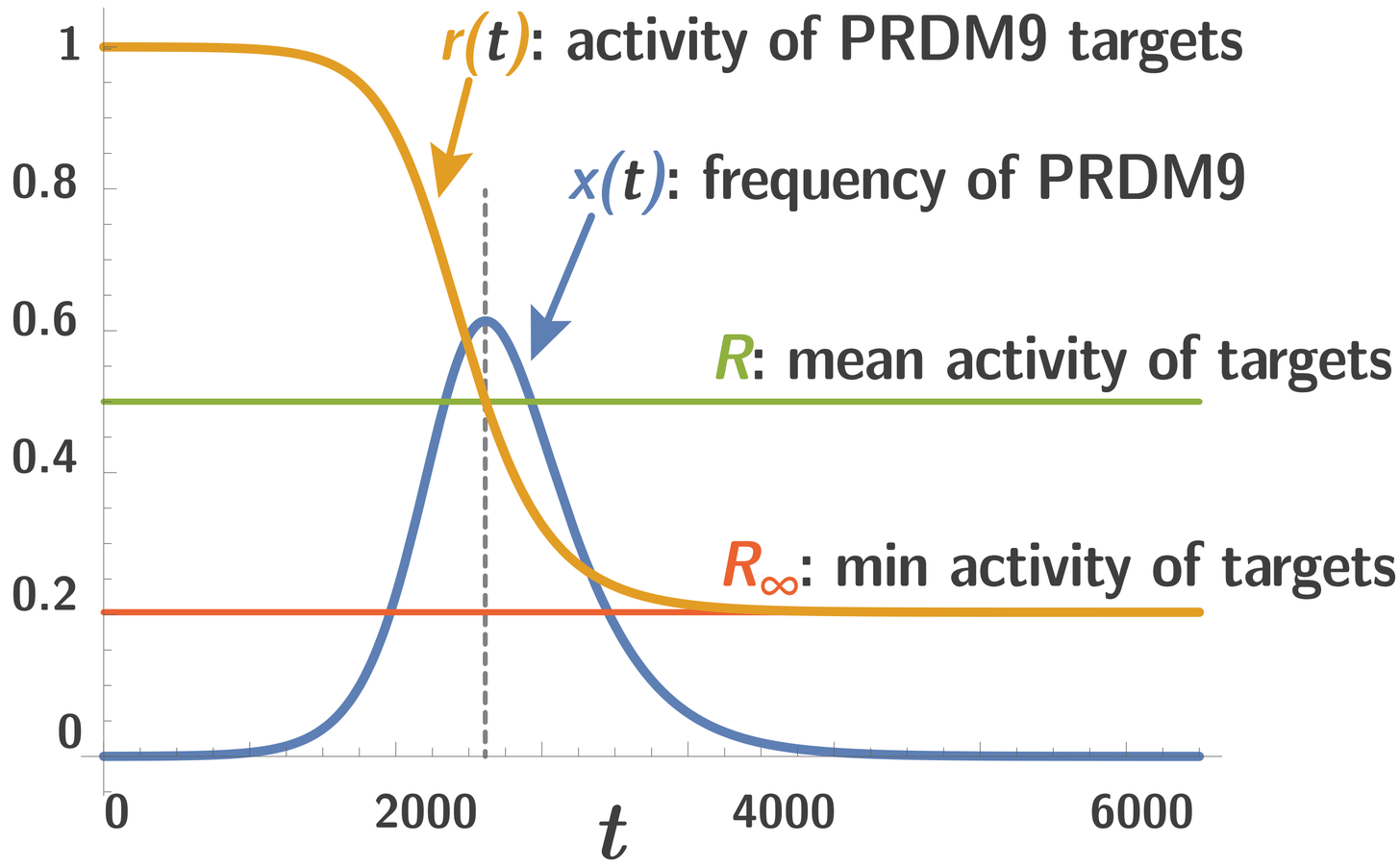
1. $x_i(t)$ is the frequency of the i^{th} PRDM9 allele.
2. $r_i(t)$ is the target's activity associated to the i^{th} PRDM9 allele.
3. Strong selection (no drift).

$$\begin{cases} \dfrac{\mathrm{d}x_i(t)}{\mathrm{d}t} = \dfrac{f'(R(t))}{2f(R(t))}\left(r_i(t) - R(t)\right)x_i(t) \\ \dfrac{\mathrm{d}r_i(t)}{\mathrm{d}t} = -\rho x_i(t) r_i(t), \text{ where } \rho = 4N_e vg \\ R(t) = \sum_i x_i(t) r_i(t) \end{cases}$$

4. $R(t)$ approximated as a constant parameter R (mean-field):

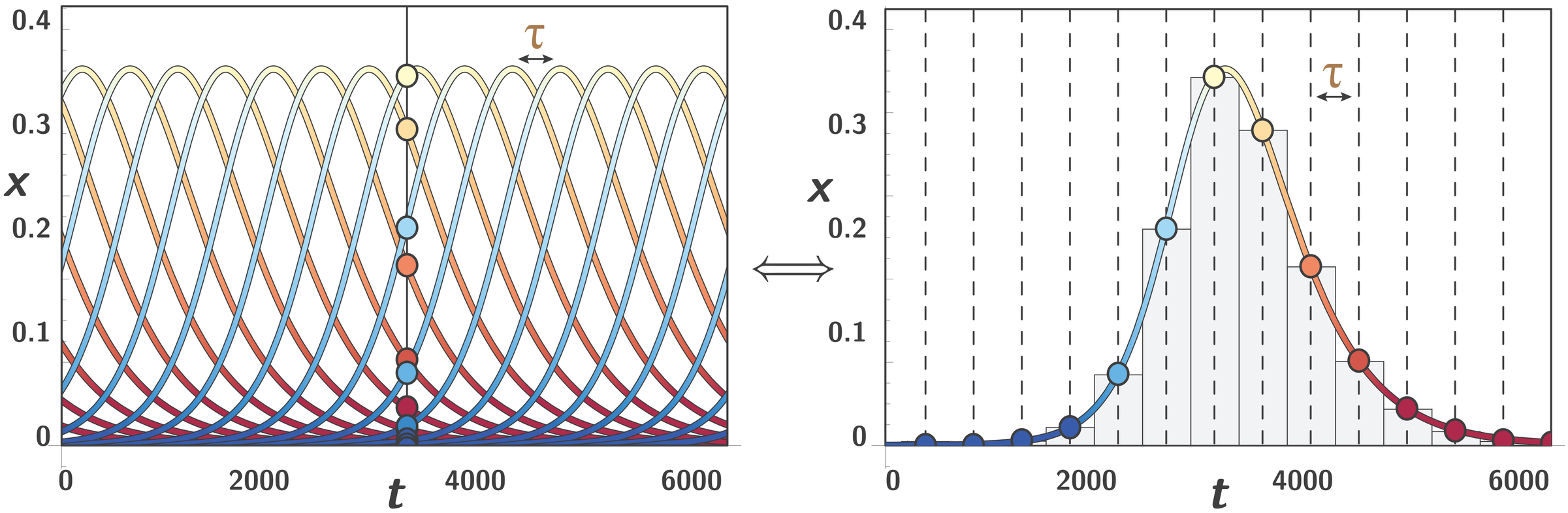
$$\begin{cases} \dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = \dfrac{f'(R)}{2f(R)}\left(r(t) - R\right)x(t) \\ \dfrac{\mathrm{d}r(t)}{\mathrm{d}t} = -\rho x(t) r(t) \end{cases}$$

Numerical resolution of the equations for a single allele



$$\begin{cases} \dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = \dfrac{f'(R)}{2f(R)}\,(r(t) - R)\,x(t) \\ \dfrac{\mathrm{d}r(t)}{\mathrm{d}t} = -\rho x(t) r(t) \end{cases} \Rightarrow \begin{cases} x(r) = \dfrac{f'(R)}{2\rho f(R)}\,[1 - r + R\ln(r)] + x_{initial} \\ 0 = 1 - R_\infty + R\ln(R_\infty) \end{cases}$$

Approximate sum of alleles as an integral of a single allele: tilling argument



$$\sum_i x_i(t) = 1 \Leftrightarrow \tau = \int_0^\infty x(t)\mathrm{d}t$$

Decoupling the equations: Mean field approximation in polymorphic regime

1. We have a relation between R_∞ and R:

$$0 = 1 - R_\infty + R\ln(R_\infty)$$

2. From the tilling argument, we also have:

$$\tau = \int_0^\infty x(t)\mathrm{d}t = \frac{1 - R_\infty}{\rho R} \Leftrightarrow R = \frac{1 - e^{-\rho\tau}}{\rho\tau}, \text{ where } \rho = 4N_e vg$$

3. As in succesion regime, τ is also the inverse of the invasion rate :

$$\tau \simeq \frac{1}{\mu}\frac{f(R)}{f'(R)}\frac{2}{1-R}, \text{ where } \mu = 4N_e u$$

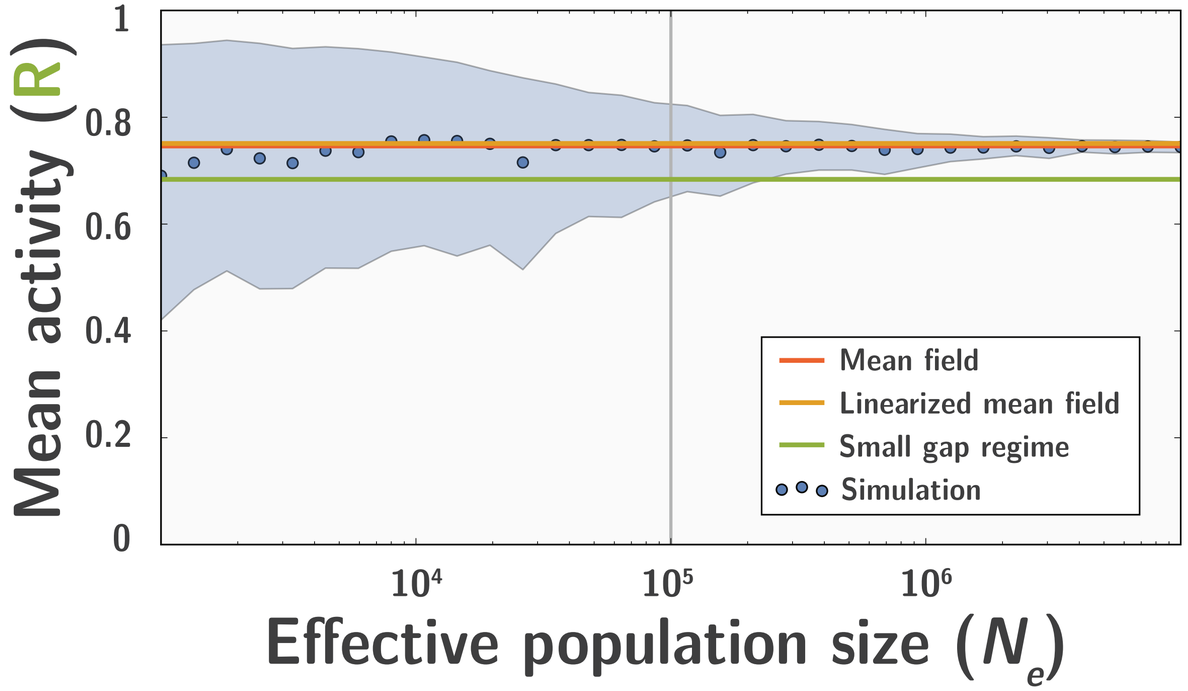
4. Altogether, we get the exact same equation as in succession regime:

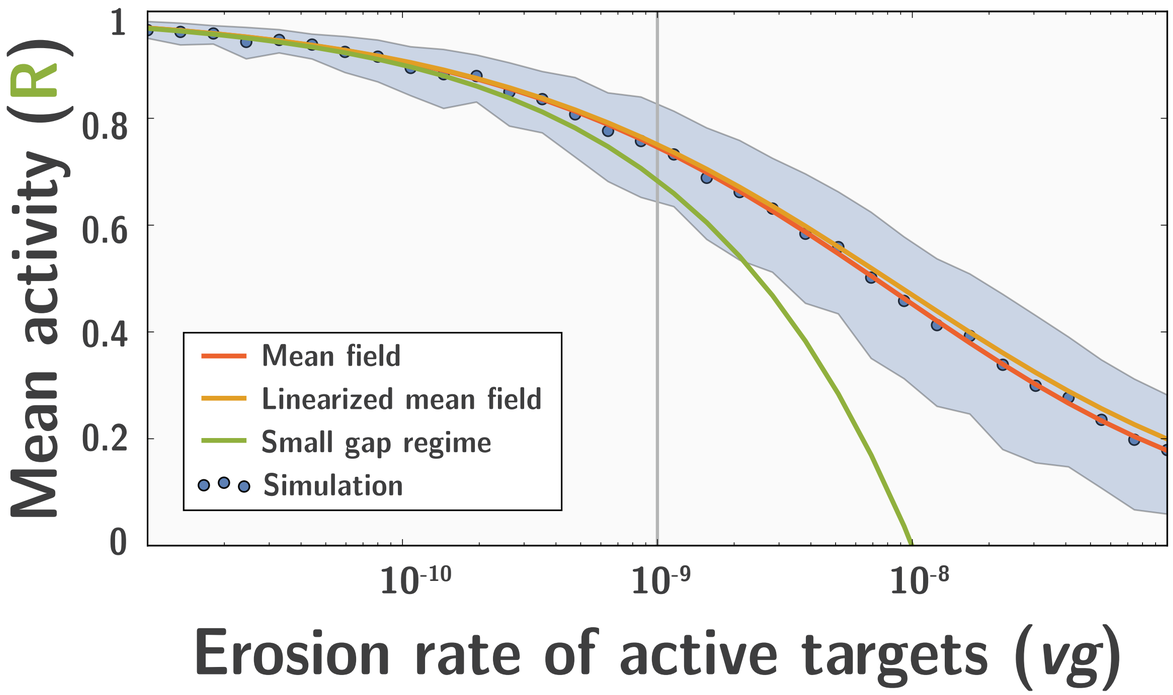
$$R = g\left(R, \frac{\rho}{\mu}\right) = g\left(R, \frac{vg}{u}\right)$$

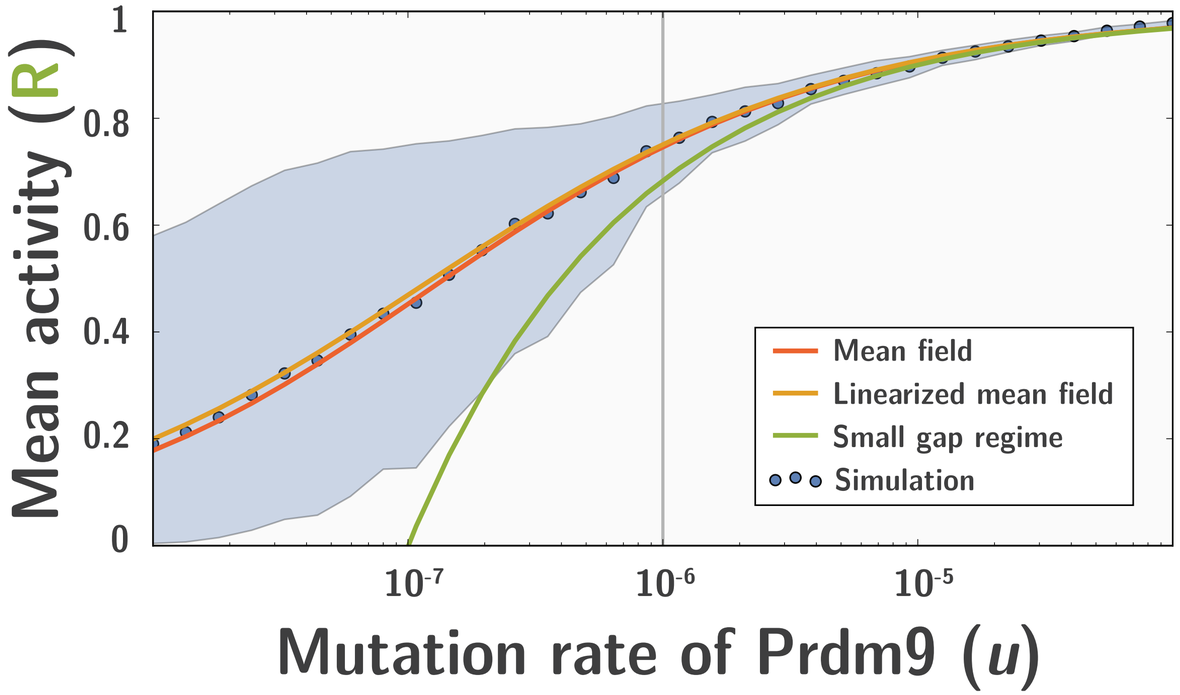
Mean activity ot the targets as function of parameters of the model

$$R = \left\langle \sum_i x_i(t) r_i(t) \right\rangle$$

$$\frac{vg}{u} \ll 1 \Rightarrow 1 - R \propto \sqrt{\frac{vg}{u}}$$



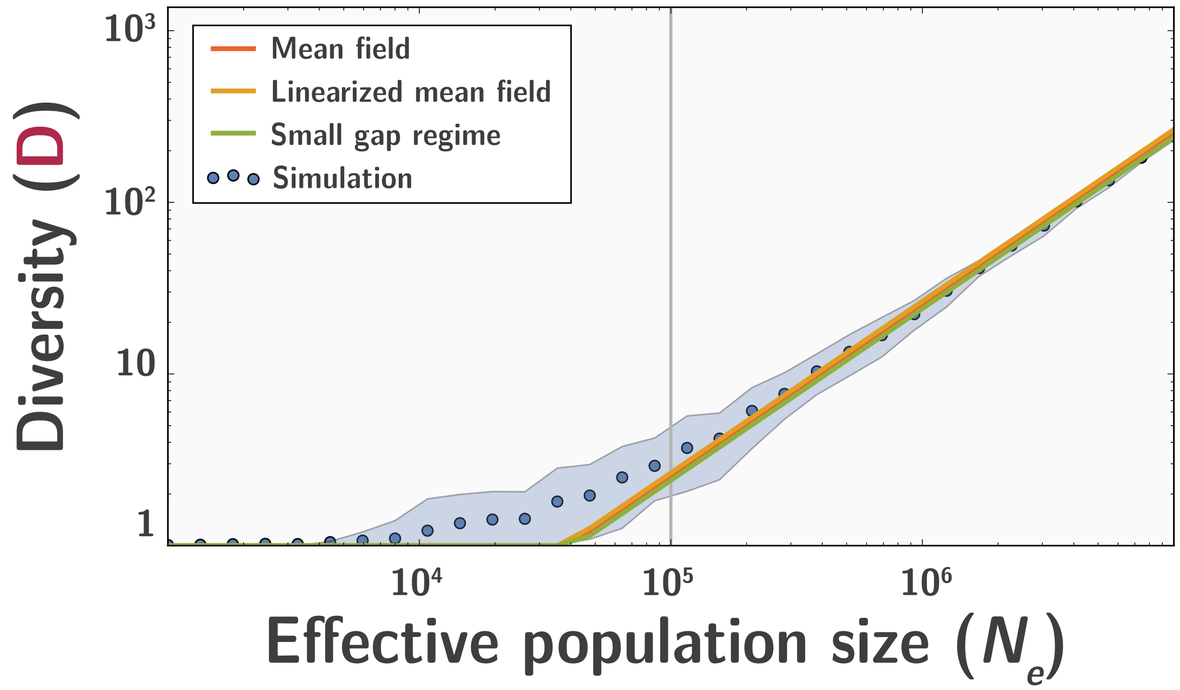


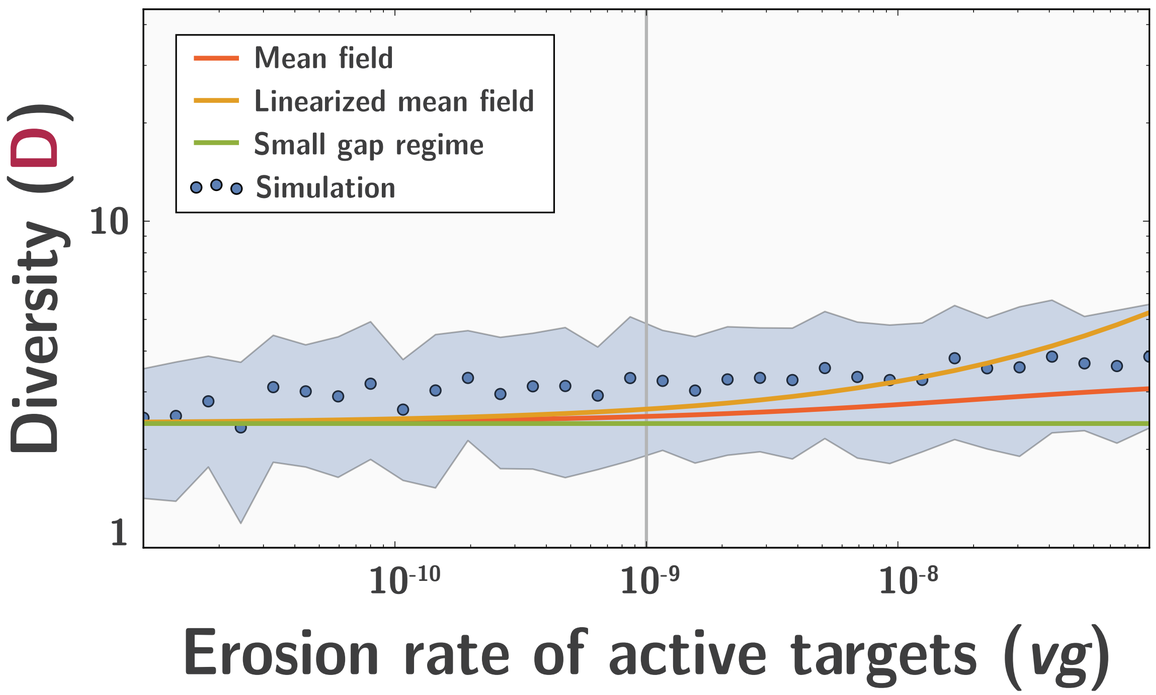


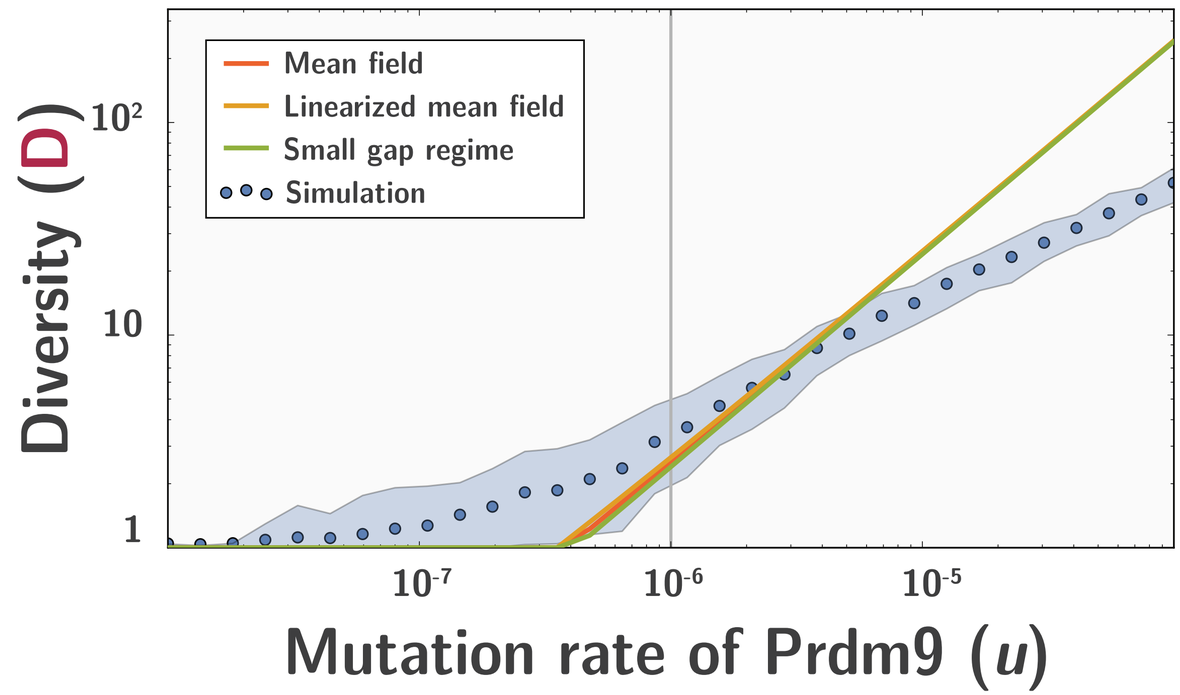
Diversity at the PRDM9 locus as function of parameters of the model

$$D = \left\langle \frac{1}{\sum_i x_i(t)^2} \right\rangle$$

$$\frac{vg}{u} \ll 1 \Rightarrow D \simeq 24 N_e u$$







Allowing for variation among hot spots in recombination rate

1. Recombination rates across hot spots vary according to a gamma distribution of mean 1 and shape parameter a:

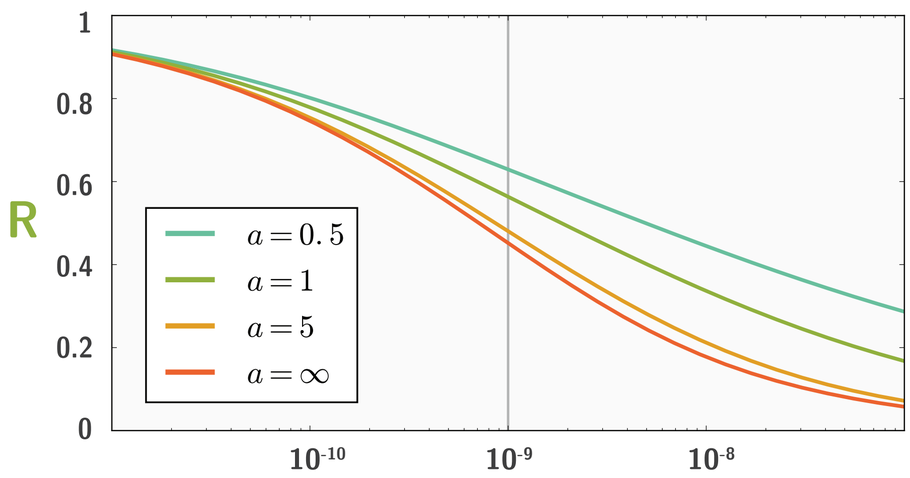
$$p(c) = \frac{b^a}{\Gamma(a)} c^{a-1} e^{-ac}$$

2. The rate of erosion for the fraction of hot spots recombining at rate c decays at a rate proportional to c:

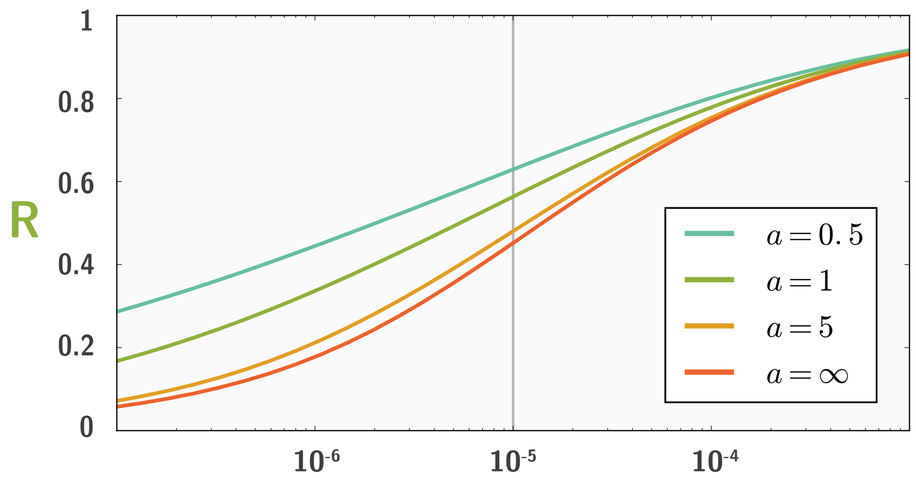
$$\frac{\mathrm{d}r_{i,c}(t)}{\mathrm{d}t} = -\rho x_i(t) c r_{i,c}(t)$$

3. The fraction of active targets in the population is then:

$$R = \frac{1}{\rho\tau} \frac{a}{(a-1)} \left[1 - \left(\frac{a}{a+\rho\tau}\right)^{a-1}\right]$$



Erosion rate of active targets (vg)



Mutation rate of Prdm9 (u)

Empirical calibration in the mouse

1. $\boldsymbol{D} \simeq 7$, estimated between 5 to 10.
2. $\boldsymbol{R} \simeq 0.5$, since the major allele eroded 50% of it's targets.
3. $S = 4N_e s_0 \gg 1$, suggested by the presence of strong positive selection acting on the Zn-finger array of PRDM9.
4. $N_e \simeq 10^5$, ranging from $N_e = 5.10^4$ to $N_e = 5.10^5$.
5. $v \simeq 10^{-7}$, assuming a point mutation rate of 10^{-8} and 10 inactivating mutations per target.
6. N_e and v are known. 3 parameters left to estimate: u, g and α.

Mutation rate of PRDM9 (u)	Erosion rate of targets (vg)	Fitness parameter (α)	$\epsilon = \frac{vg}{u}$	Mean fraction of active targets ($\boldsymbol{R}$)	Diversity at PRDM9 locus ($\boldsymbol{D}$)	Scaled selection coefficient (S)	Turn-over time (T)
3×10^{-6}	3×10^{-10}	1×10^{-4}	1×10^{-4}	0.6	9.9	26	6.4×10^4
3×10^{-6}	3×10^{-11}	1×10^{-4}	1×10^{-5}	0.82	8.2	8.6	1.6×10^5
3×10^{-7}	3×10^{-11}	1×10^{-4}	1×10^{-4}	0.6	1	26	6.5×10^4
3×10^{-6}	3×10^{-11}	1×10^{-5}	1×10^{-5}	0.6	9.9	2.6	6.4×10^5

Table: Fitness function is a power law, $f(x) = x^\alpha$

7. $u \simeq 3.10^{-6}$, $g \simeq 3.10^{-3}$ and $\alpha \simeq 10^{-4}$.

CONCLUSION

- Succession ($N_e u \ll 1$) and polymorphic ($N_e u \gg 1$) regimes
- Red Queen accelerates in the succession regime, spreads over more weaker hotspots in polymorphic regime.
- Diversity (**D** $\propto 24 N_e u$) and depletion (1-**R** $\propto (vg/u)^{1/2}$) are uncoupled.
- Mean activity (**R**) is solely governed by the strength of the two arms vg and u, and not on population size. The system will respond to forcing in one arm by balancing with the second arm.
- Under arbitrary fitness model, the positive selection exerced on PRDM9 is not diversying, **D** is seamingly neutral!
- Empirical calibration requires a high dBGC (g=3.10^{-3}) and high mutation rate of PRDM9 (u=3.10^{-6})

DISCUSSION & PERSPECTIVES

- **D** predicted by our model is supposed to correspond to the functional diversity thus potentially lower than the raw diversity.

- **R** might be less extrem than 50% measured specifically for the major PRDM9 alleles.

- Our fitness model ignors the role of recombination as a dissipator of linkage disequilibrium.

- Why PRDM9 happens to have such a high mutation rate?

- Hybrid sterility phenotype is directly related to the asymmetry in the patterns of PRDM9 binding along chromosomes in F1 hybrid.

- The Red Queen process then supposedly plays role in the creation of genetic barriers between subspecies. We could certainly extend our model in a metapopulation context.

The end

Thank you for your attention