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.

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 $2N_{\mathrm{e}}v$
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 $2gx_i(t)$ for i^{th} allele of PRDM9.

Altogether, the activity induced by allele i decays as:

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

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

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

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

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

4. The mean fitness over the population is

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

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

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

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

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

2. $\boldsymbol{\tau}$ the mean time between two successive invasions.

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

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

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

4. $\boldsymbol{\tau}$ is also the inverse of the invasion rate:

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

5. Altogether,

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

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

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

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

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

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

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

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:

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

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

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

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

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

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:

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

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

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

$$\frac{vg}{u} \ll 1 \Rightarrow D \simeq 24 N_{\mathrm{e}} u$$

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

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 $\alpha \simeq 10^{-4}$.