The red queen in the kingdom of recombination

Thibault Latrille, Laurent Duret, Nicolas Lartillot

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

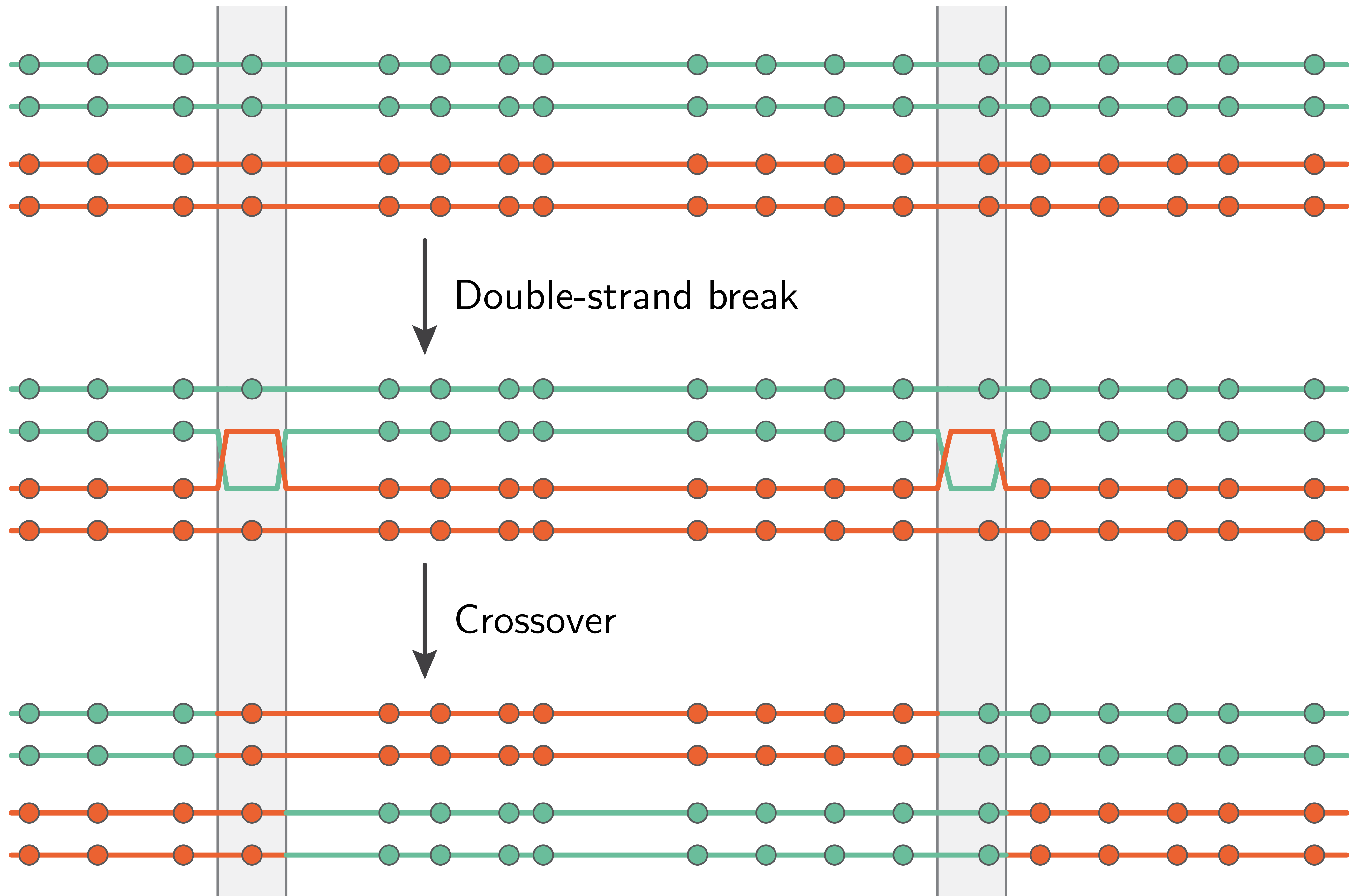July 4, 2016

Ad-hoc meeting, LBBE

CHAPTER 1

Once upon a time in the kingdom of recombination...
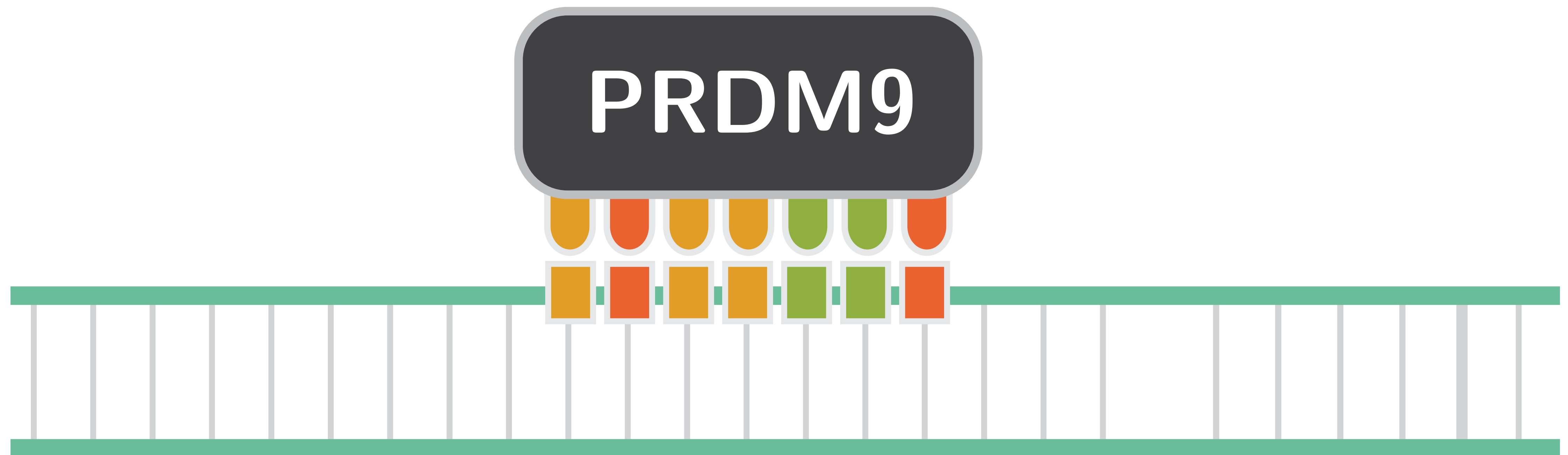
...there was recombination hotspots

...there was a zinc-finger protein

...there was erosion of hotspots

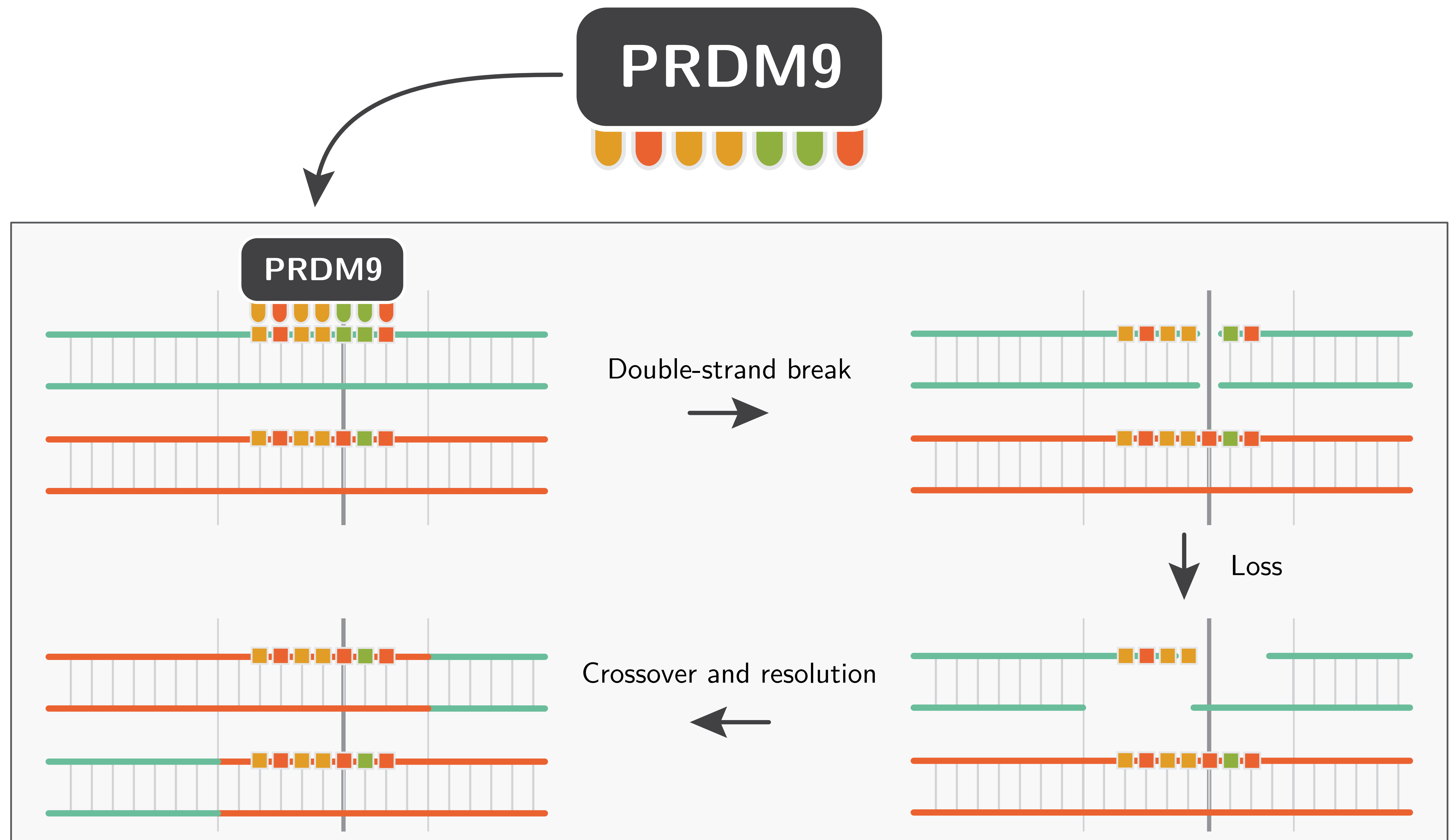# ...there was recombination hotspots

Double-strand break

Crossover

...there was a zinc-finger protein

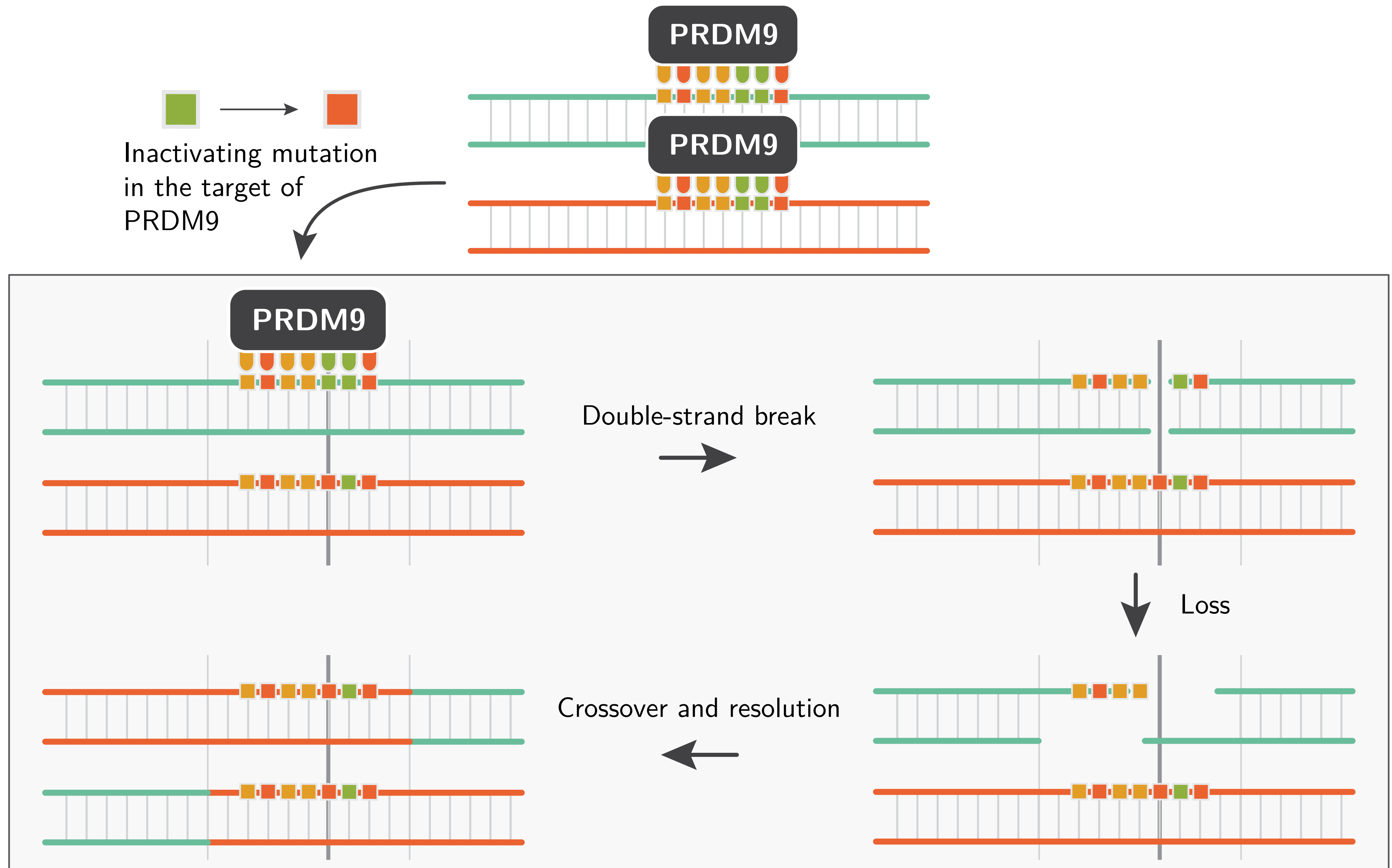PRDM9 targets sequences located at
recombination hotspots.

Parvanov *et al* (2010), Baudat *et al* (2010)

...there was a zinc-finger protein that induces a double strand break


Coop & Myers (2007), Myers *et al* (2010)

...there was erosion of hotspots

Inactivating mutation in the target of PRDM9

Double-strand break

Loss

Crossover and resolution

Coop & Myers (2007), Myers *et al* (2010)

The red queen own paradox: PRDM9 governs hotspots, but kills them slowly

PRDM9

Erosion of recombination hotspots by PRDM9

The red queen own paradox:
PRDM9 governs hotspots, but kills them slowly

PRDM9

Mutation

PRDM9

Erosion of recombination hotspots by PRDM9

Erosion of recombination hotspots by PRDM9
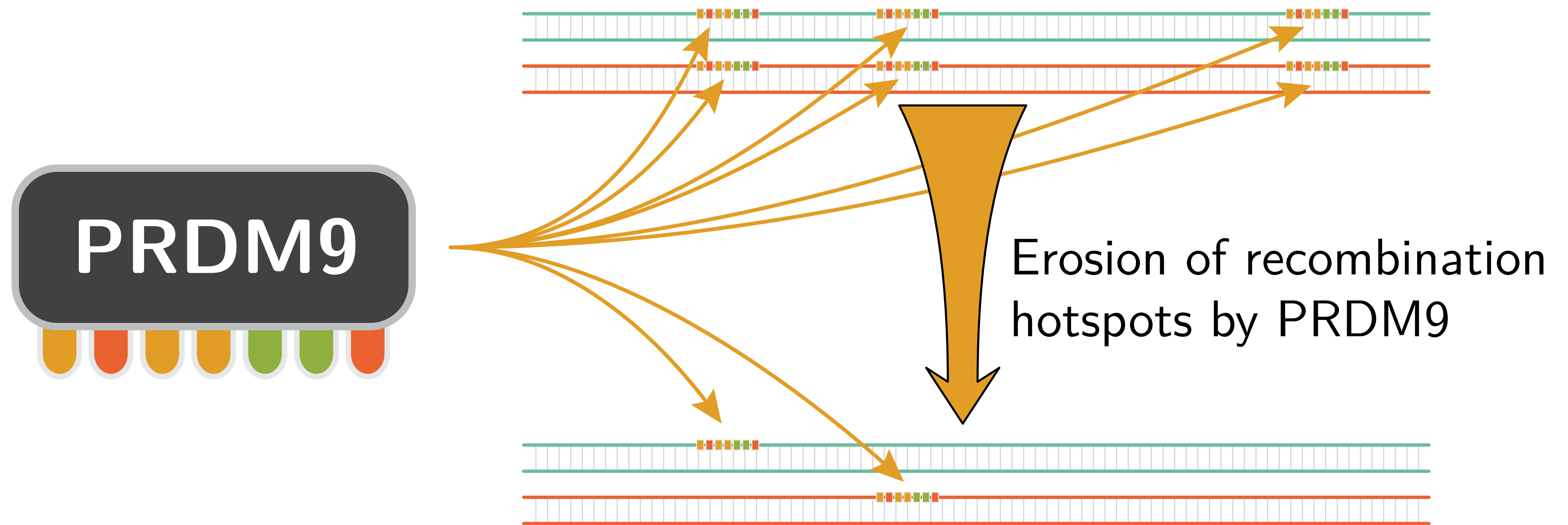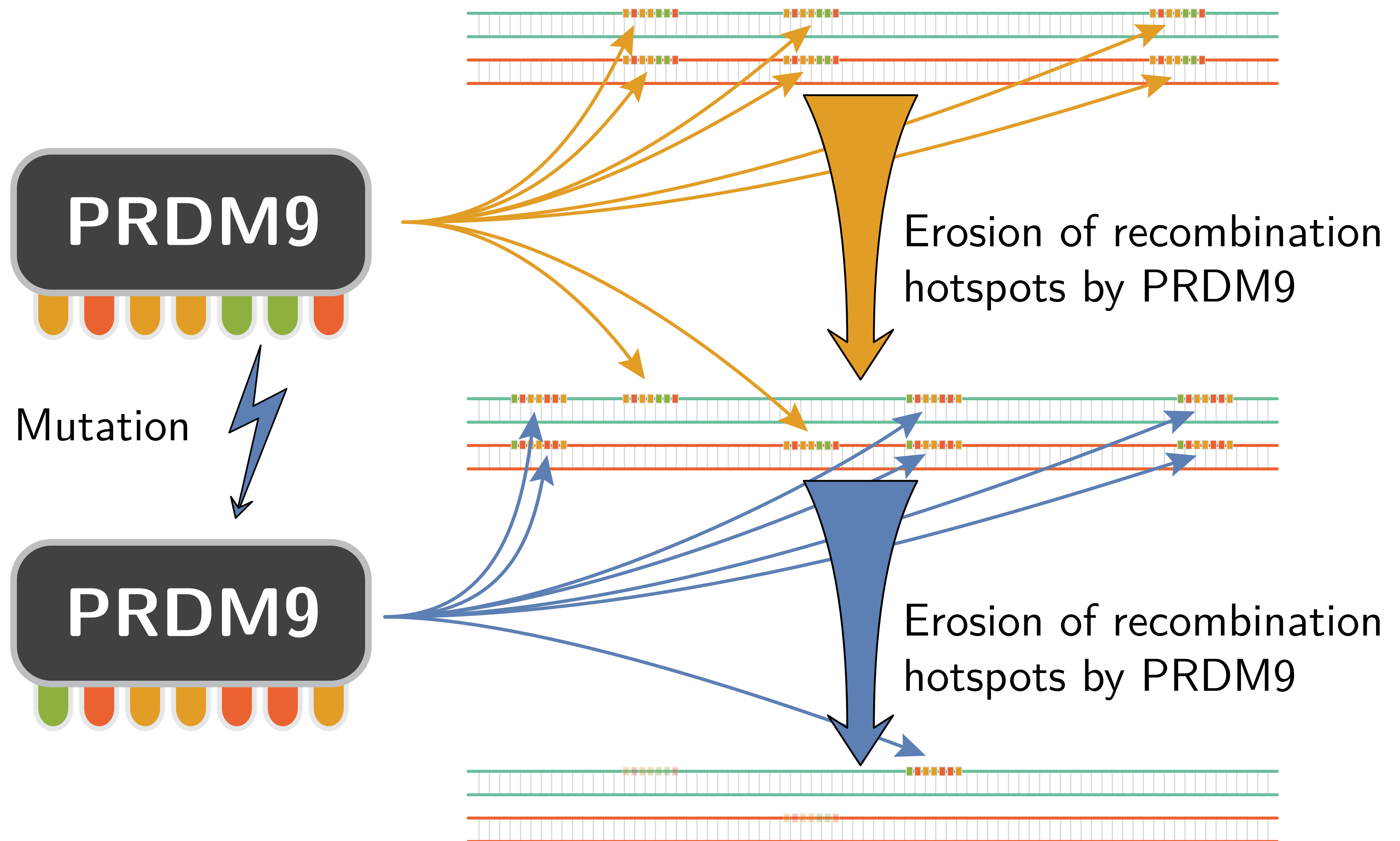
End of chapter 1

Red queen hypothesis :
In reference to an evolutionnary system, continuing adaptation is needed in order for a species to maintain its relative fitness among the system it is co-evolving with.
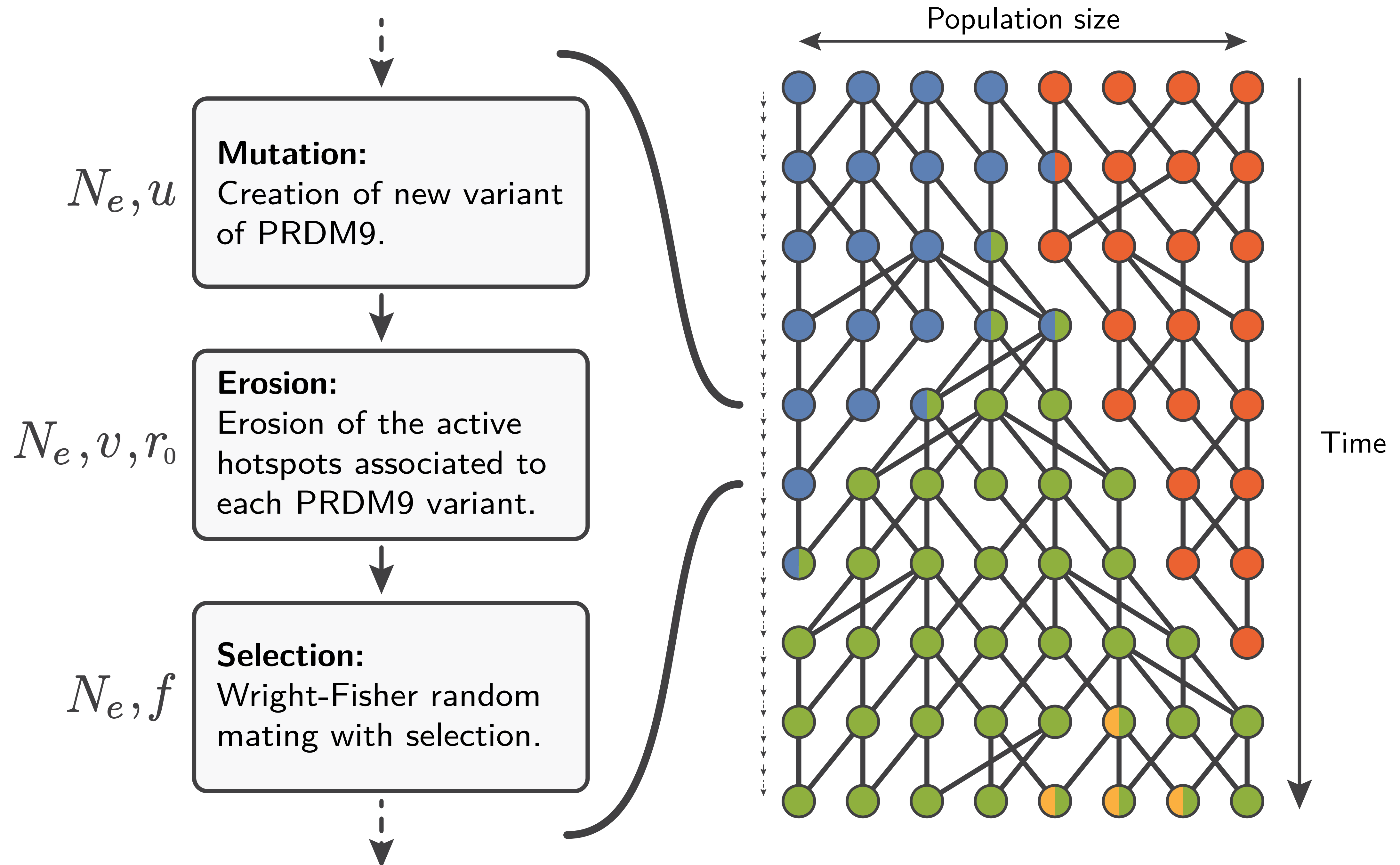
- Can we construct a population genetic model that exhibits red queen dynamic ?

- From qualitative explanation to a quantitative model.

CHAPTER 2

A population genetic model
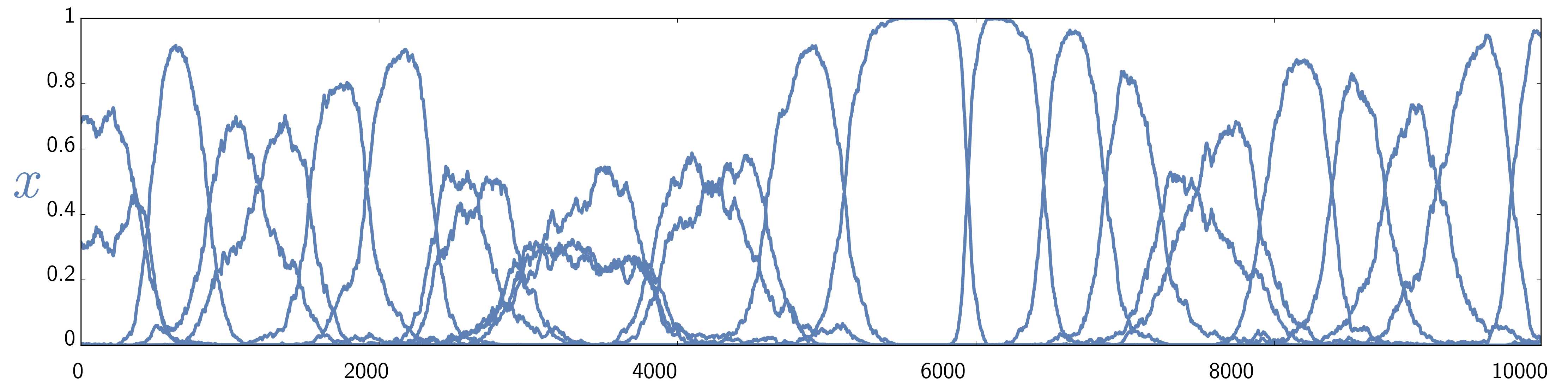# built to Her Majesty the red queen

- Stepping into the territory of Wright-Fisher simulations

- Peeking at simulated trajectory

- Exploration of parameters space

Stepping into the territory of Wright-Fisher simulations

N_e, u

Mutation:
Creation of new variant of PRDM9.

N_e, v, r_0

Erosion:
Erosion of the active hotspots associated to each PRDM9 variant.

N_e, f

Selection:
Wright-Fisher random mating with selection.
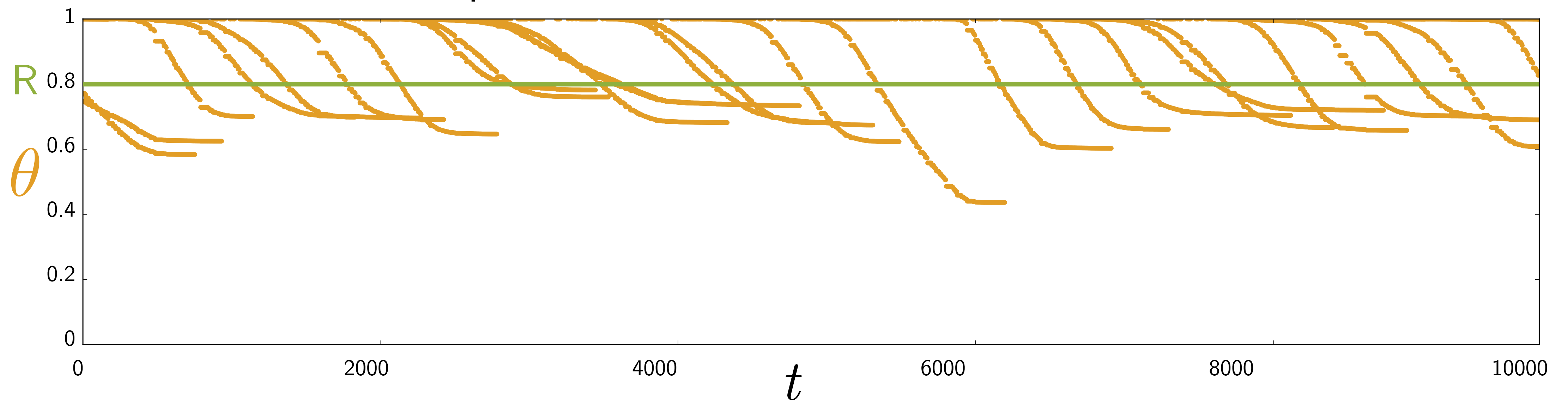
Population size

Time

Peeking at simulated trajectory
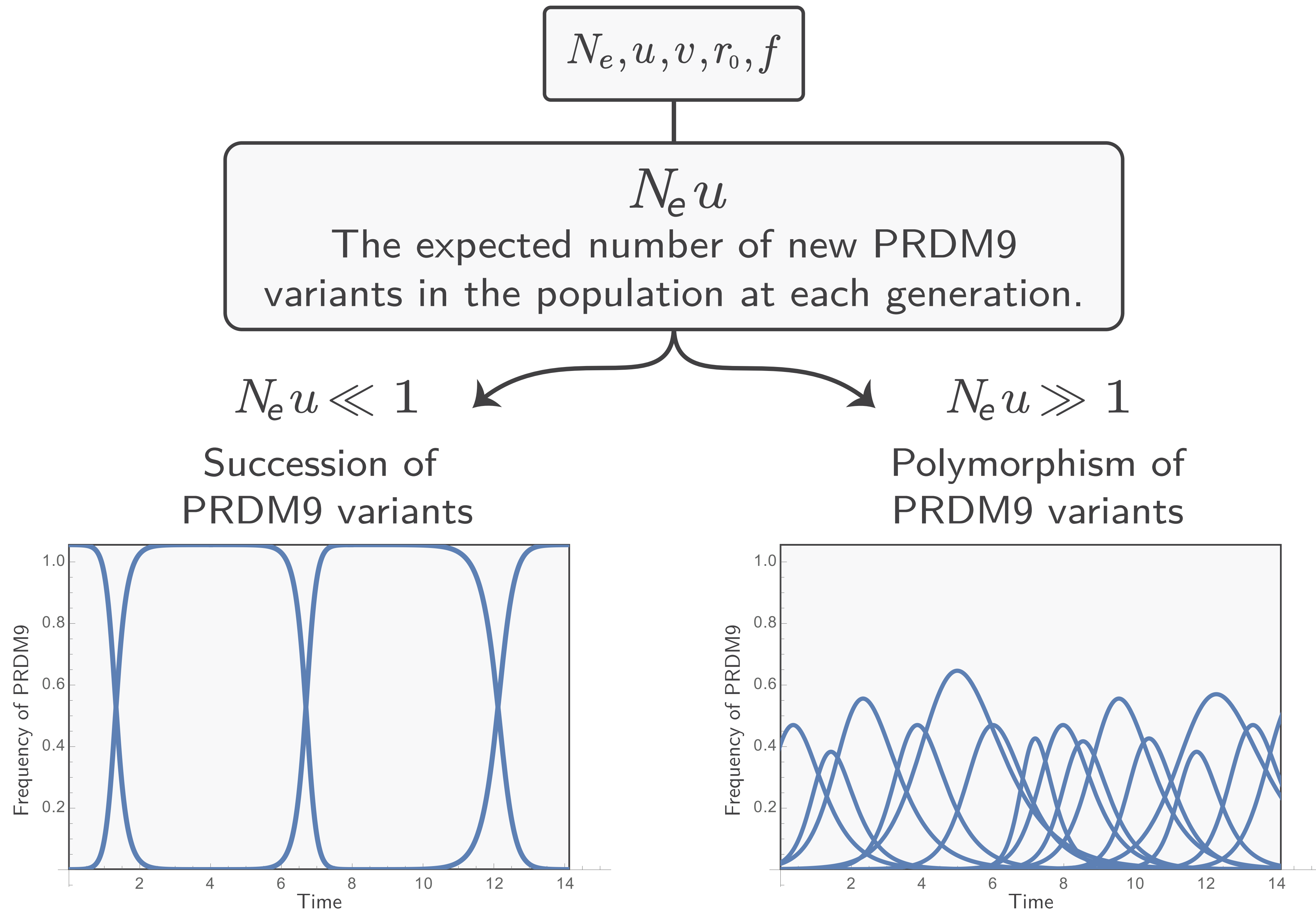
PRDM9 frequency over time for each alllele

hotspots recombination rate over time for each allele

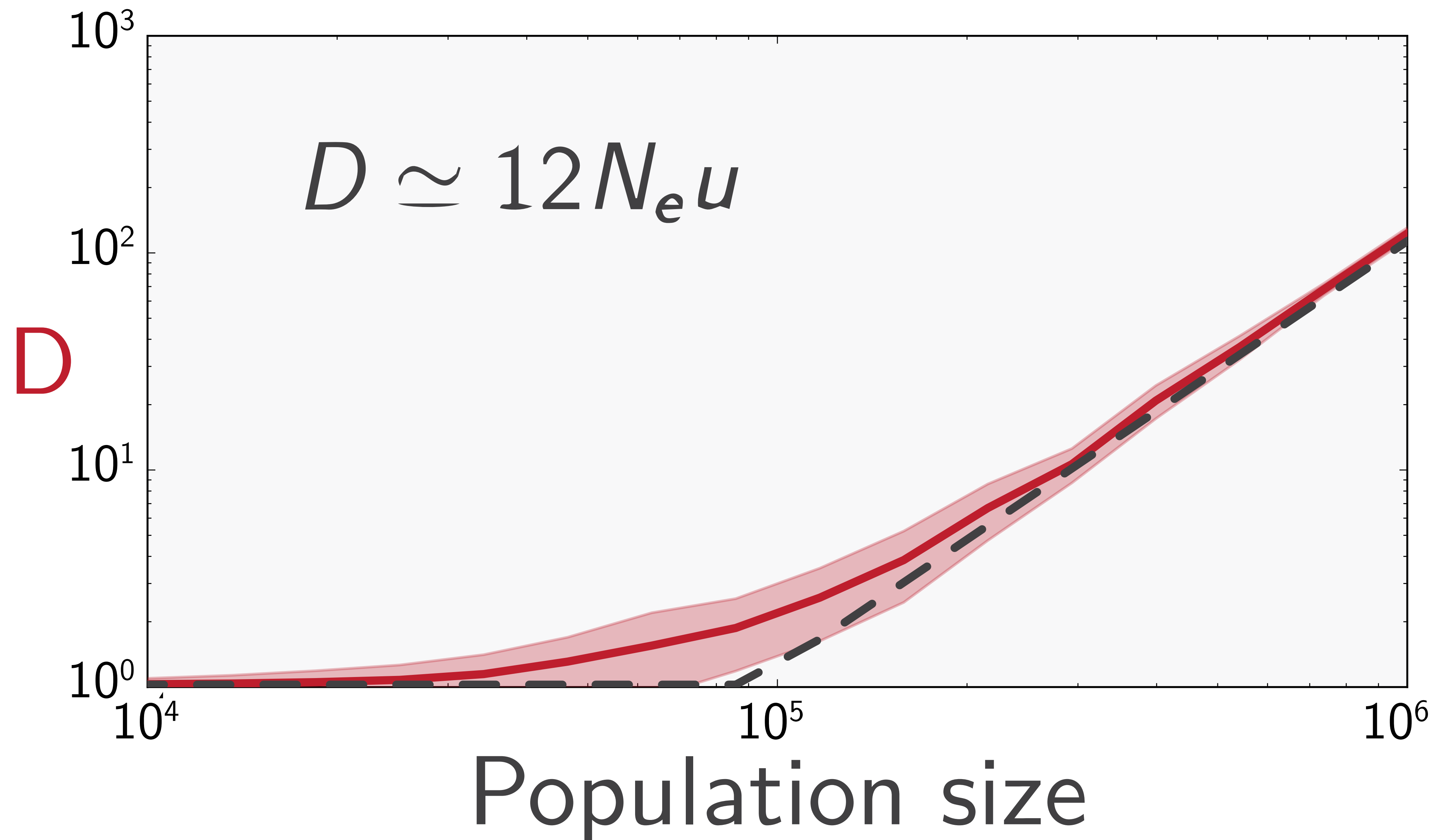# Succession or polymorphism of PRDM9 ?

$$N_e, u, v, r_0, f$$

$$N_e u$$

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

$N_e u \ll 1$

Succession of PRDM9 variants

$N_e u \gg 1$

Polymorphism of PRDM9 variants

Phase diagram of the red queen
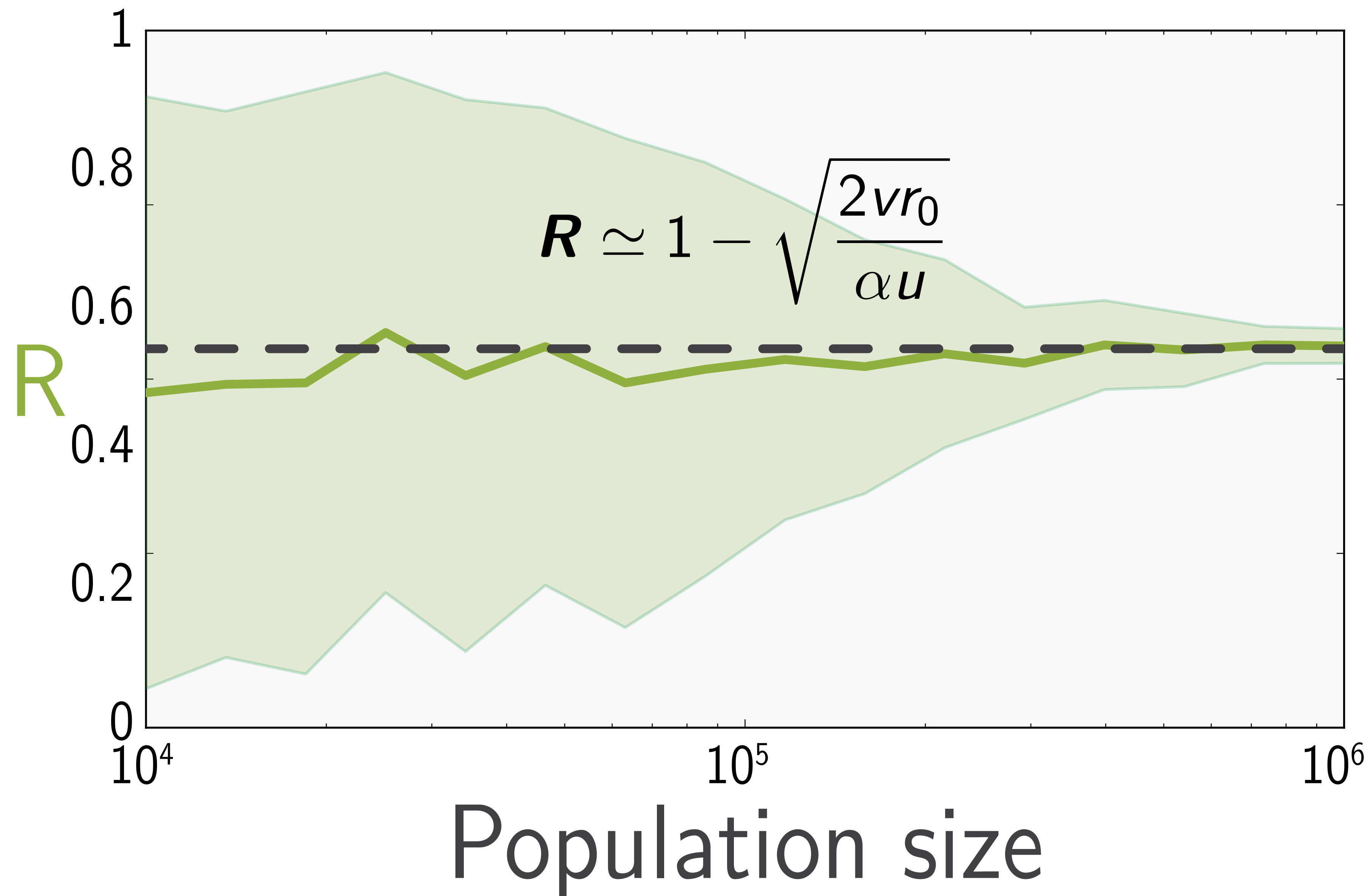# Round 1: Diversity of PRDM9 (D) *vs* Population size

$$D \simeq 12 N_e u$$

Phase diagram of the red queen
Round 2: Diversity of PRDM9 (D) *vs* Mutation rate
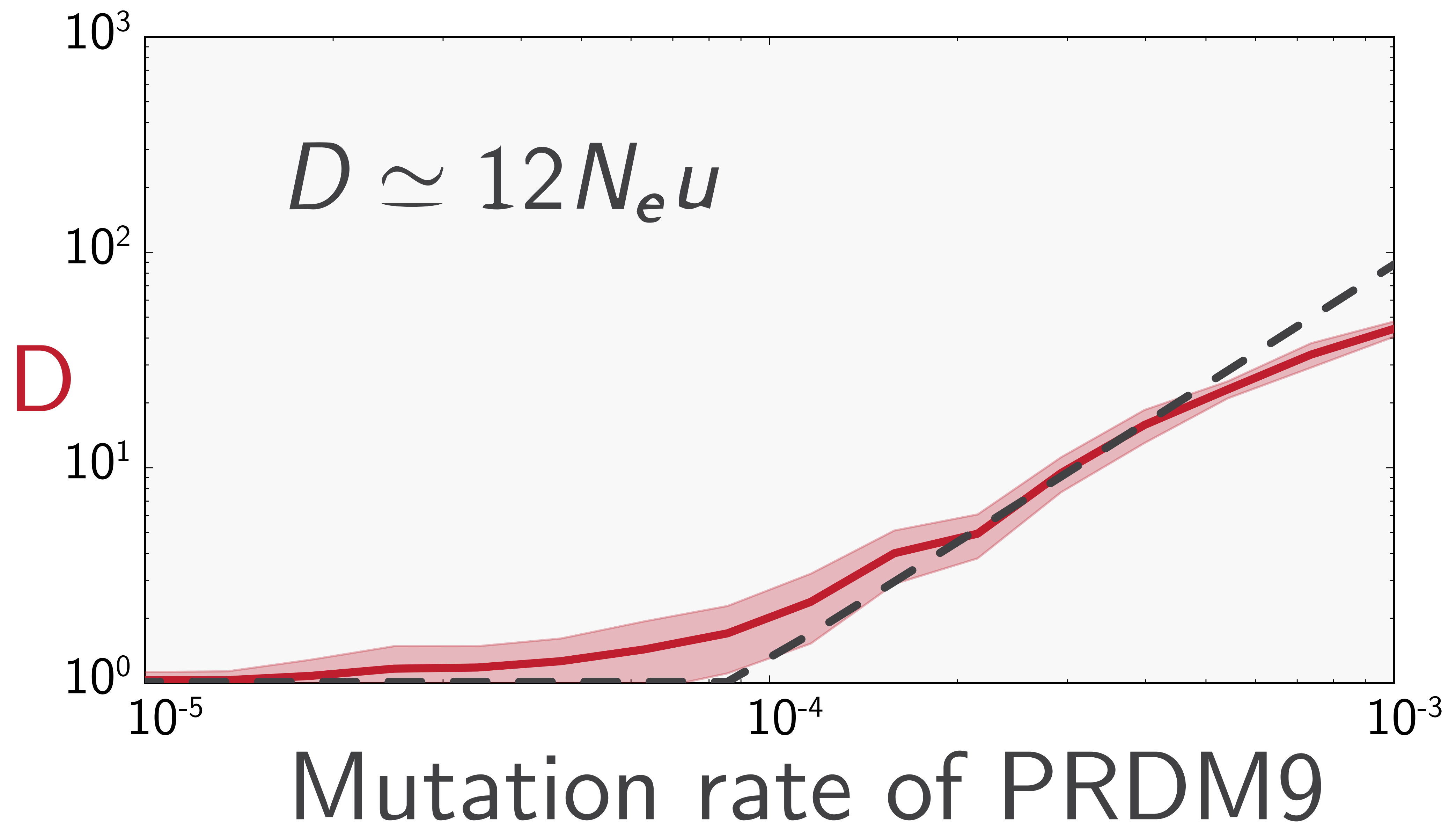
Phase diagram of the red queen
Round 3: Recombination rate (R) *vs* Population size

$$R \simeq 1 - \sqrt{\frac{2vr_0}{\alpha u}}$$

R

Population size

END OF CHAPTER 2

- What intuition can be derived from this model ?

- From a quantitative to an analytical model

CHAPTER 3

A journey to the sovereignty of mathematics

Coupled equations for all PRDM9 alleles

1. K_t is the number of PRDM9 alleles in the population.

2. $x_{i,t}$ is the frequency of the ith PRDM9 allele.

3. $\theta_{i,t}$ is the recombination rate associated to the ith PRDM9 allele.
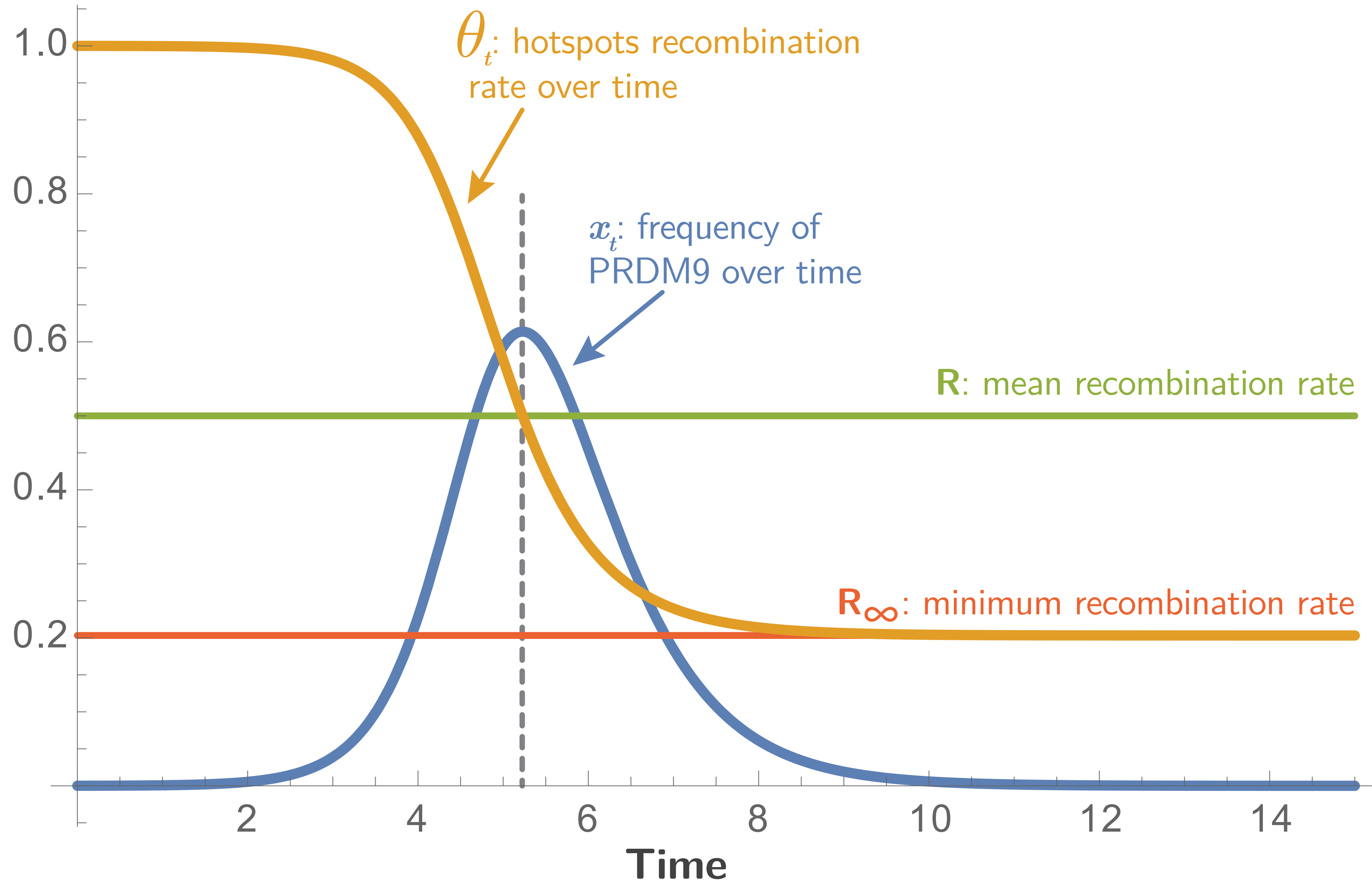
4. Assume there is no drift.

$$
\begin{cases}
\dfrac{\mathrm{d}x_{i,t}}{\mathrm{d}t} = \dfrac{\alpha}{R_t}\left(\theta_{i,t} - R_t\right)x_{i,t}, \ \forall i \in \{1, \ldots, K_t\} \\[2ex]
\dfrac{\mathrm{d}\theta_{i,t}}{\mathrm{d}t} = -\rho x_{i,t}\theta_{i,t}, \ \forall i \in \{1, \ldots, K_t\} \\[2ex]
R_t = \displaystyle\sum_{i \in K_t} x_{i,t}\theta_{i,t}
\end{cases}
$$

Decoupling the equations :
Mean field approximation in polymorphic regime

1. x_t is the frequency of PRDM9.

2. θ_t is the recombination rate associated to PRDM9.

3. Assume there is no drift.
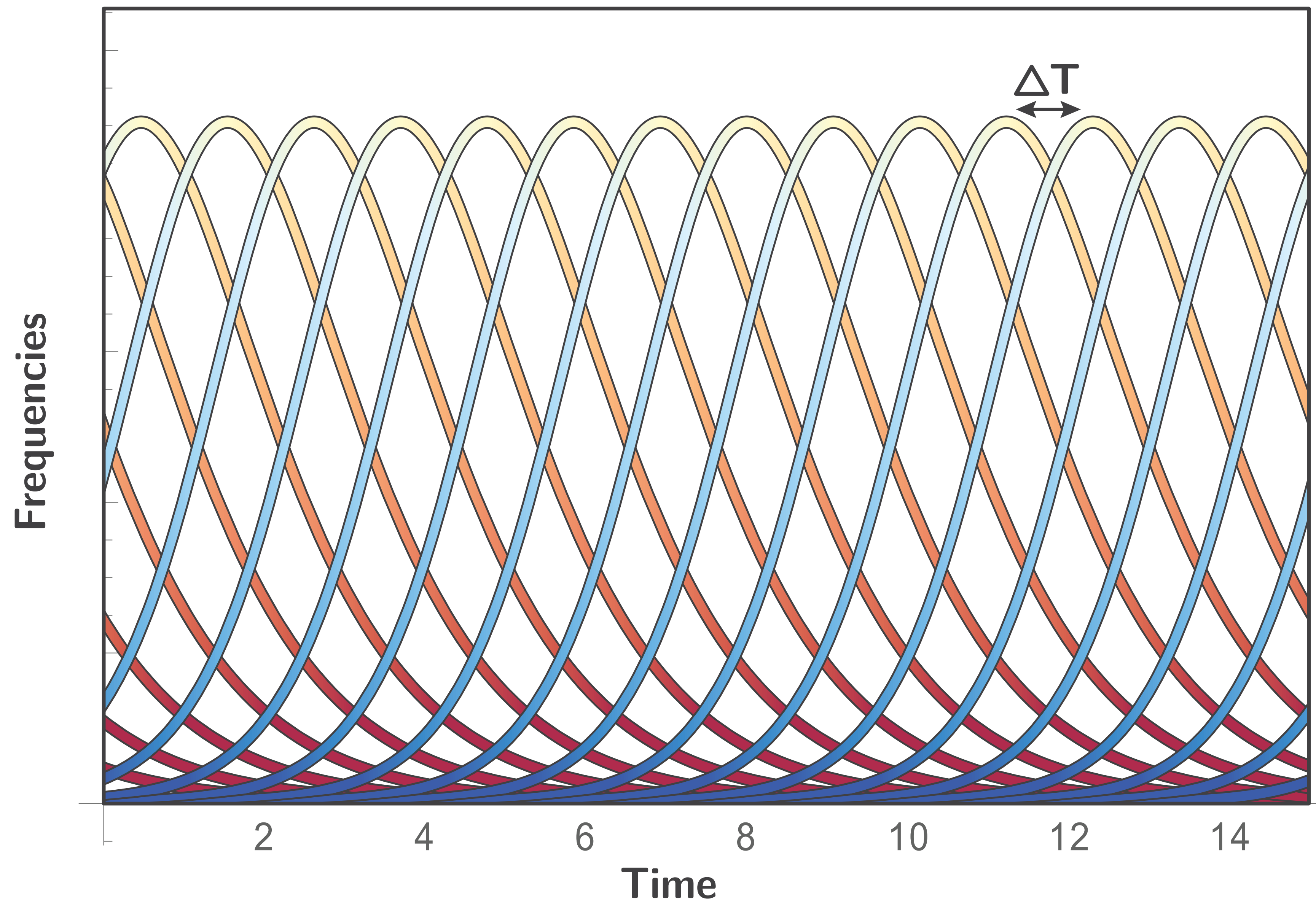
4. Approximate R_t as a constant parameter.

$$
\begin{cases}
\dfrac{\mathrm{d}x_t}{\mathrm{d}t} = \dfrac{\alpha}{R} \left(\theta_t - R \right) x_t \\[2mm]
\dfrac{\mathrm{d}\theta_t}{\mathrm{d}t} = -\rho x_t \theta_t
\end{cases}
$$

Numerical resolution of the equations for a single allele

θ_t: hotspots recombination rate over time

x_t: frequency of PRDM9 over time

R: mean recombination rate

$\mathbf{R_\infty}$: minimum recombination rate

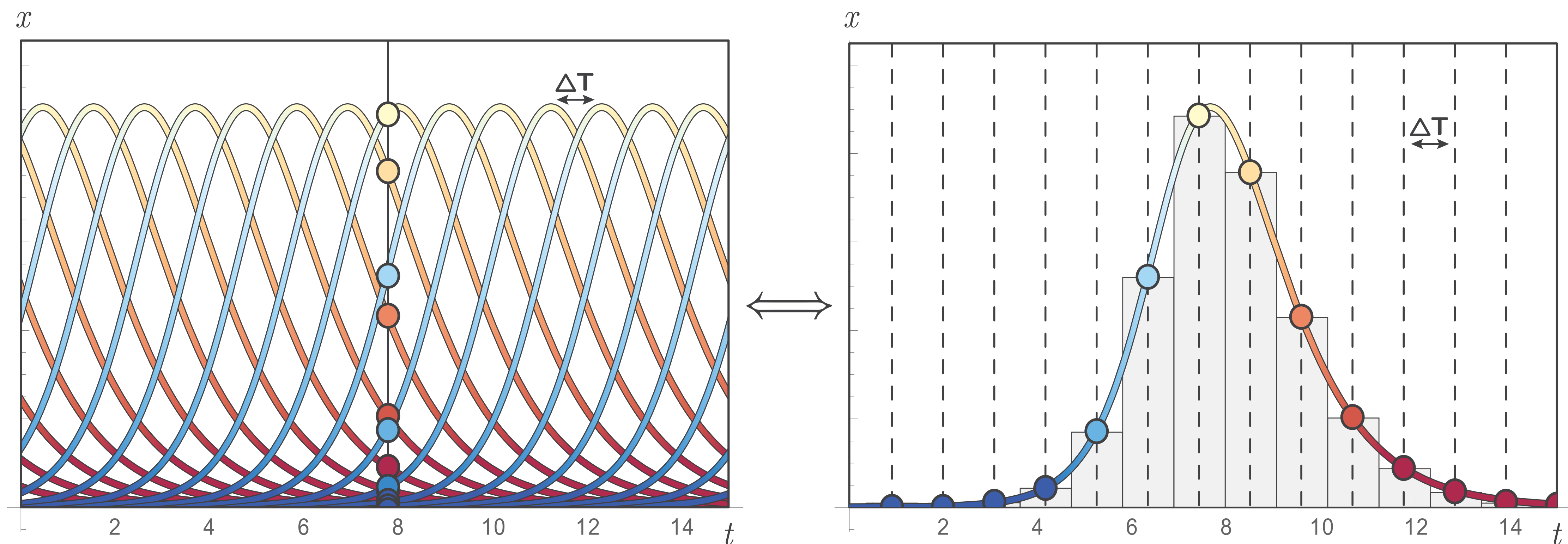$$0 \simeq 1 - \mathbf{R_\infty} + \mathbf{R} \log(\mathbf{R_\infty})$$

And back to polymorphism using delayed equations

$$\Delta T = \frac{1}{\tau_{\text{mutation}} p_{\text{invasion}}} = \left(2 \alpha N_{\text{e}} u \frac{1 - R}{R} \right)^{-1}$$

Approximate sum of alleles as an integral of a single allele

$$\Delta T = \Delta T \sum_{i \in K_t} x_{i,t} \simeq \int_0^\infty x_t \mathrm{d}t = \frac{1 - R_\infty}{\rho R}$$
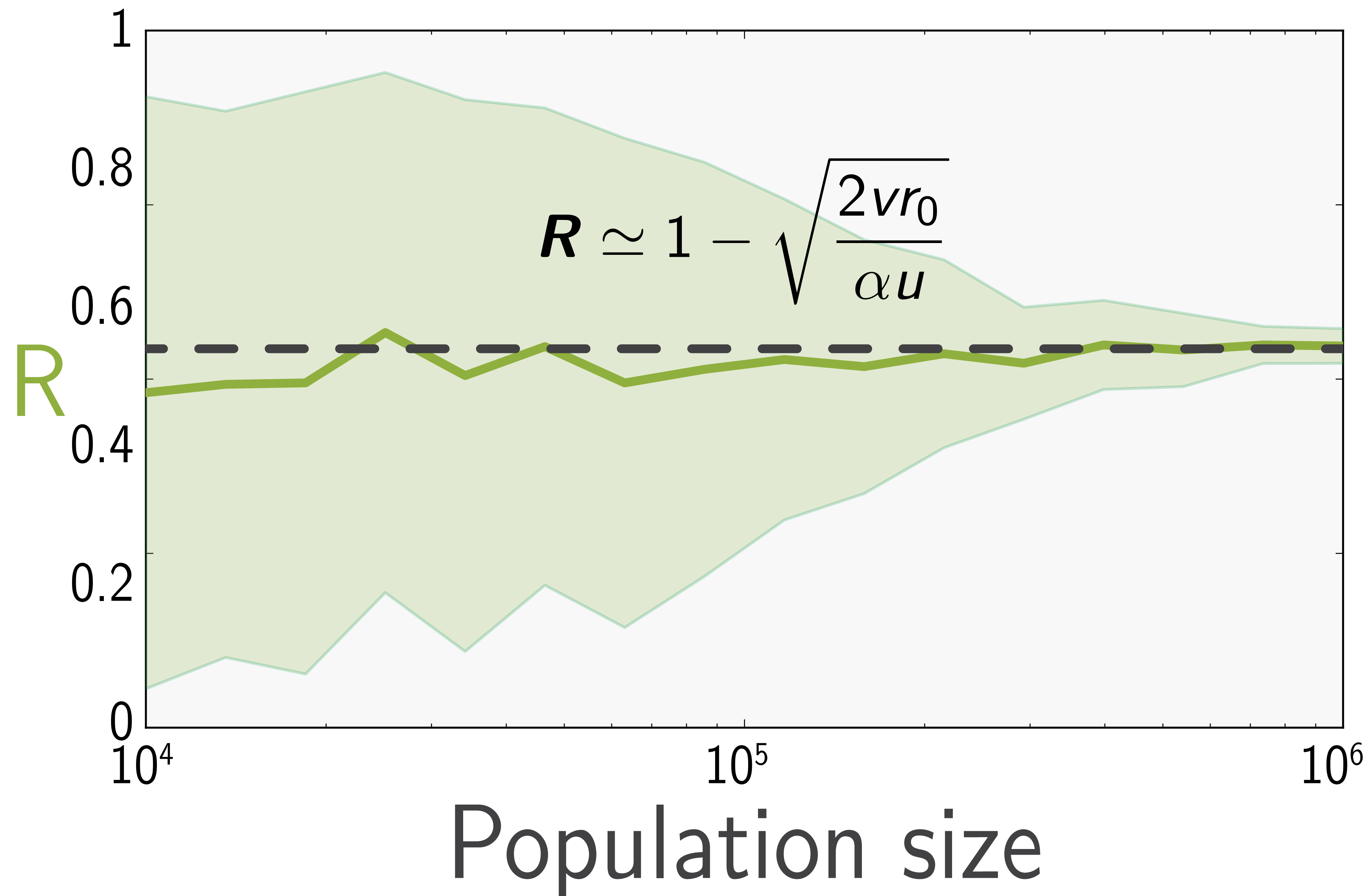
Approximation for the mean recombination rate (R)

$$\begin{cases} \Delta T \simeq \dfrac{1 - R_\infty}{\rho R} \\[2em] \Delta T \simeq \left(2\alpha N_{\mathrm{e}} u \dfrac{1 - R}{R} \right)^{-1} \\[2em] 0 \simeq 1 - R_\infty + R \log(R_\infty) \end{cases}$$

$$\Rightarrow \begin{cases} \dfrac{(1 - R)(1 - R_\infty)}{R^2} \simeq \dfrac{2 v r_0}{\alpha u} \\[1.5em] 0 \simeq 1 - R_\infty + R \log(R_\infty) \end{cases}$$
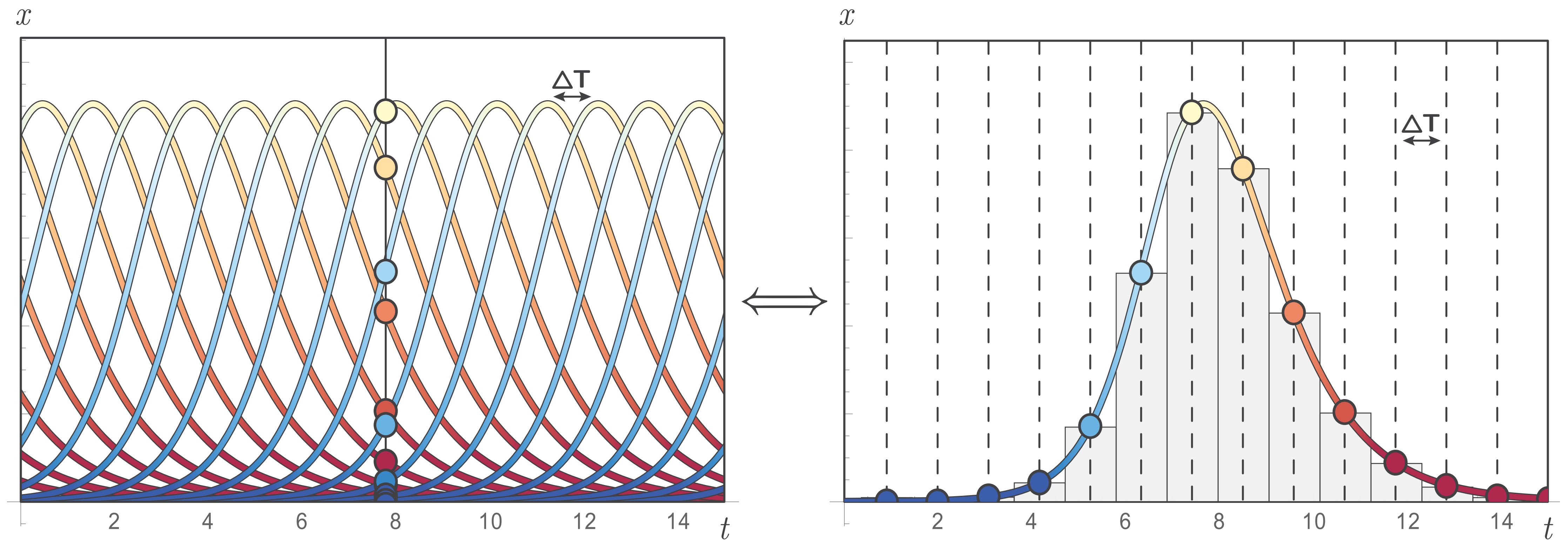
$$\Rightarrow R \simeq 1 - \sqrt{\dfrac{2 v r_0}{\alpha u}}$$
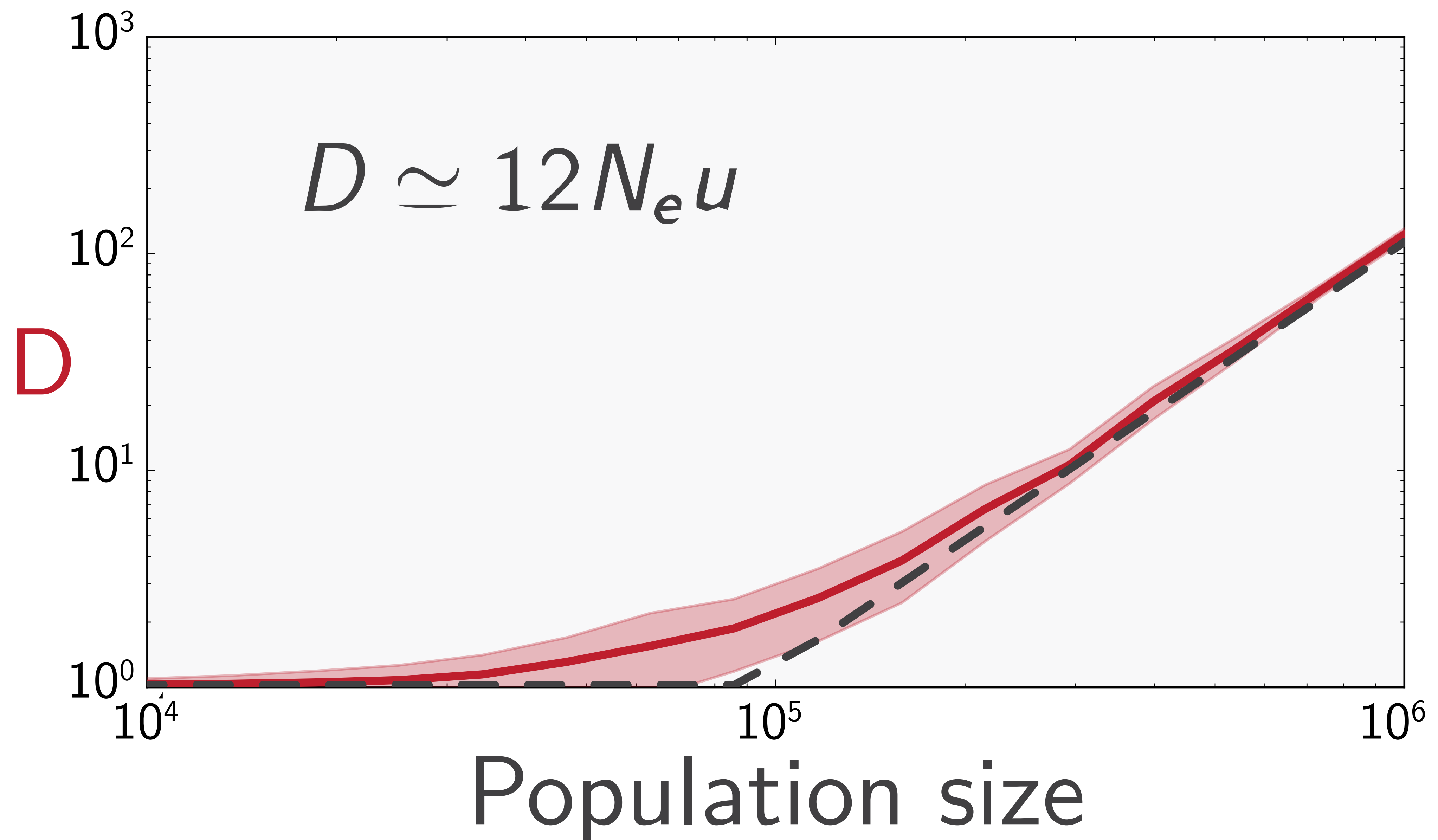
Recombination rate (R) *vs* Population size

$$R \simeq 1 - \sqrt{\frac{2vr_0}{\alpha u}}$$
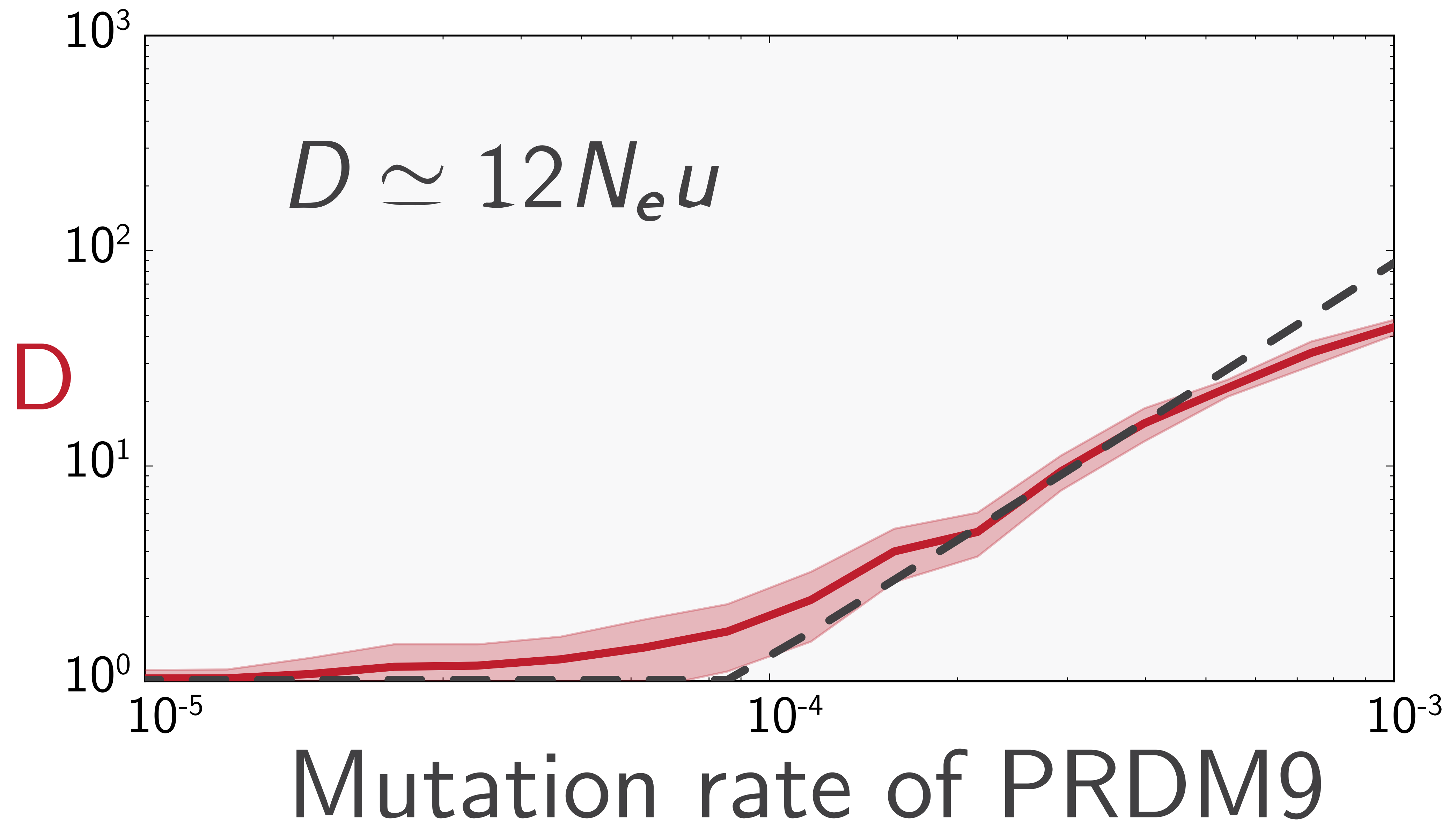
Approximation of D, the diversity at the PRDM9 locus

$$D = \left(\sum_{i \in K_t} x_{i,t}^2 \right)^{-1} = \left(\int_0^\infty x_t^2 \mathrm{d}t \right)^{-1} \int_0^\infty x_t \mathrm{d}t \simeq 12 N_\mathrm{e} u$$

Diversity of PRDM9 (D) *vs* Population size

$$D \simeq 12 N_e u$$

D

Population size

Diversity of PRDM9 (D) *vs* Mutation rate

$$D \simeq 12 N_e u$$

D

Mutation rate of PRDM9

EPILOGUE

Conclusion:

- We can mathematicaly estimate mean recombination rate (R) directly from the parameters. It is independent of population size.

- We can mathematicaly estimate diversity of PRDM9 (D) directly from the parameters. It is increasing with regard to population size and mutation rate at the PRDM9 locus.

Perspective:

- What can we tell about turn-over ? The landscape of hotspots ?

- What about the data ?

THE END

Thank you for your attention

If you have any questions, feel free to answer them…