How so diverse?

Self-incompatibility in flowering plants:
a case-study for modelling the role of chance in evolution,
from the beginnings of population genetics to today

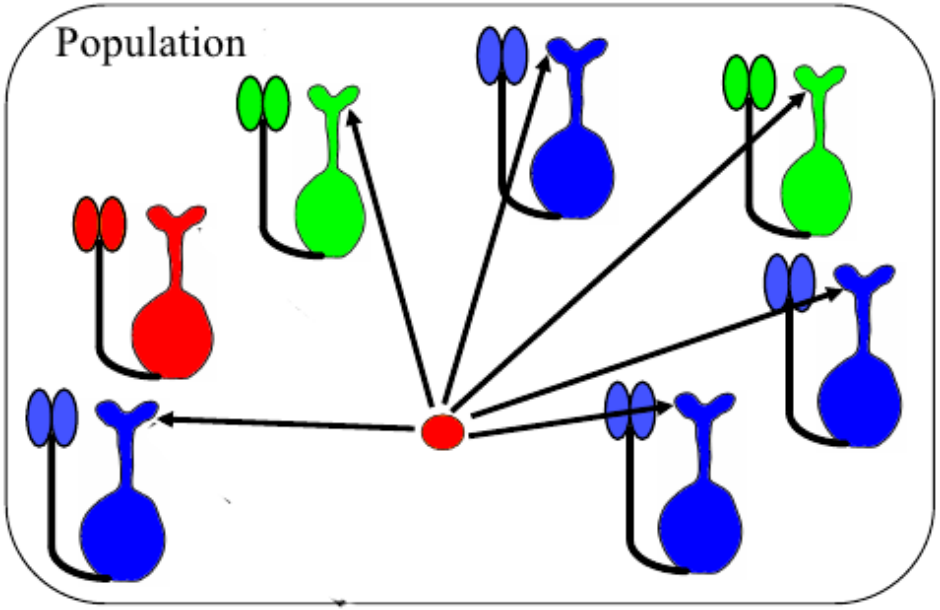
Sylvain Billiard

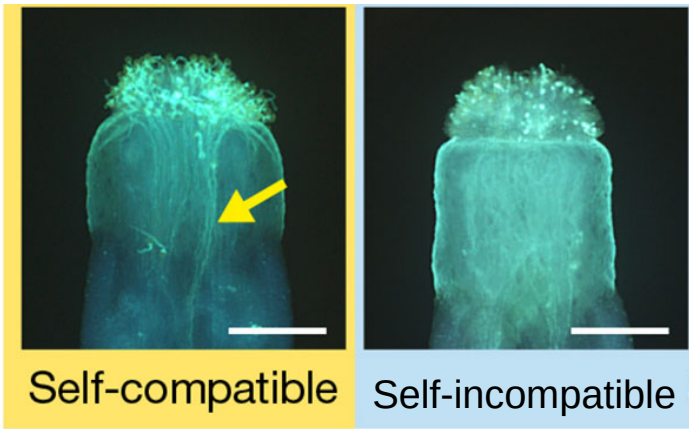
Probabilty meets Biology Workshop — Bath — 2023

Université de Lille - France - Lab. Evo-Eco-Paléo - UMR CNRS 8198

Wild Cherry (*Prunus avium*)

Self-incompatibility: who can mate with whom?





Suwabe et al. 2020

- Brassicaceae (cabbage, *Arabidopsis*, rapeseed, ...)
- Asteraceae (salad, chicory, chrysanthemum, …)
- Oleaceae (olive tree, ash tree, lilac, …)
- Rosaceae (cherries, apples, …)
- Solanaceae (tobacco, tomato, potato, ...)

A huge surprising diversity

Oenothera organensis

45 S-alleles ~ 5000 individuals

(Emerson 1938, 1939)

Wild cherry *(Prunus avium)*

22 S-alleles ~ 500 individuals

(Stoeckel et al. 2011)

Wild tomato *(Solanum chilense)*

34 S-alleles ~ 100 sampled individuals

(Igic et al. 2007)

Arabidopsis halleri

66 S-alleles ~ 900 sampled individuals

(V. Castric, Pers. Comm.)

Self-incompatibility and stochastic models: a long history

Oenothera organensis

45 S-alleles ~ 5000 individuals

(Emerson 1938, 1939)

The birth of Population genetics stochastic models

- Fisher (1930), Wright(1937, 1938, 1945), Kolmogorov: Fokker-Planck equation
- Malécot (1945), Moran (1962), Ewens, Kimura, etc.: Markov Process

An early application to Self-Incompatibility systems

- Wright(1939)

One of the first quantitative prediction ever in population genetics

Natural selection

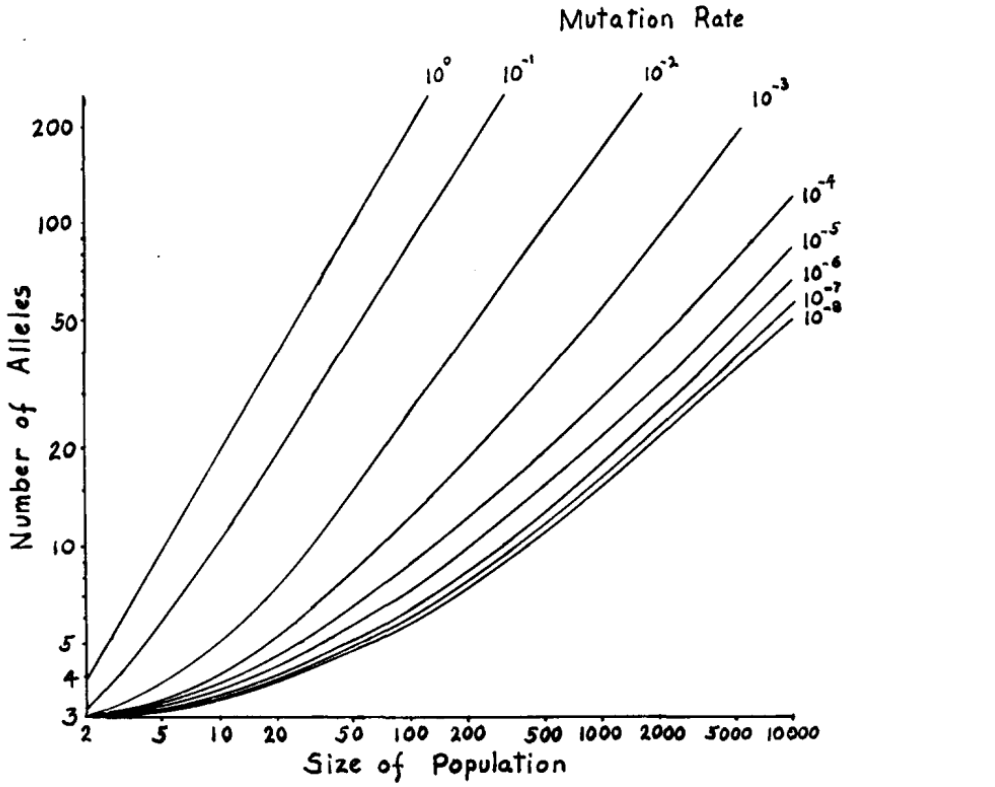
$$\Delta q = \frac{q(1-q)(1+R)}{2[R+q(1-R)]} - [1+u+m(1-q_t)]\,q + (v+mq_t)(1-q)$$

Mutation and Immigration

$$\sigma^2_{\Delta q} = \frac{q(1-q)}{2N}.$$

Binomial random sampling ("genetic drift)

$$\phi(q) = C[R+q(1-R)]^{2N(1+R)/(1-R)} q^{4N(v+mq_t)-1}(1-q)^{4N[1+u+m(1-q_t)]-1}$$



Self-incompatibility: an archetype for stochastic models in population genetics (long story short)

Haters gonna hate

- Fisher (1958), Wright (1960)

A surge from the shoulders of Giants

- Wright(1964, 1966)
- Ewens (1964, 1966), Kimura and Crown (1964), Mayo (1966), Moran (1962), Yokoyama and Nei (1979)

But incomplete

- Moran (1962): lack of rigour as "*the probabilistic model has not been specified*" and other criticisms
- Wright (1964): not important as " *My paper was directed at giving an admittedly approximate solution of a biological problem*"

From micro to macro: from scratch (Czuppon and Billiard 2022)

A Moran's model

Rates and approximations

Birth rate

$$T_{ij}^{+} = \Bigg(\underbrace{\frac{1}{2}\sum_{k\neq i,j} A_{jk}\frac{p_i}{1-p_j-p_k}}_{i\text{ pollen fertilization}} + \underbrace{\frac{1}{2}\sum_{k\neq i,j} A_{ik}\frac{p_j}{1-p_i-p_k}}_{j\text{ pollen fertilization}} \Bigg)_{\text{non}-\{ij\}\text{replacement.}} \underbrace{\frac{N-A_{ij}}{N}}$$

Death rate

$$T_{ij}^{-} = \underbrace{\left(N - \frac{1}{2}\sum_{k\neq i,j} A_{ik}\frac{p_j}{1-p_i-p_k} - \frac{1}{2}\sum_{k\neq j,i} A_{jk}\frac{p_i}{1-p_j-p_k} \right)}_{\text{non}-\{ij\}\text{ reproduction}} \underbrace{\frac{A_{ij}}{N}}_{\{ij\}\text{ replacement}} .$$

Approximations of the stationary distributions: Ornstein-Uhlenbeck for *M* different S-alleles.

(Diffusion approximation on frequencies, N large, Central Limit Theorem)

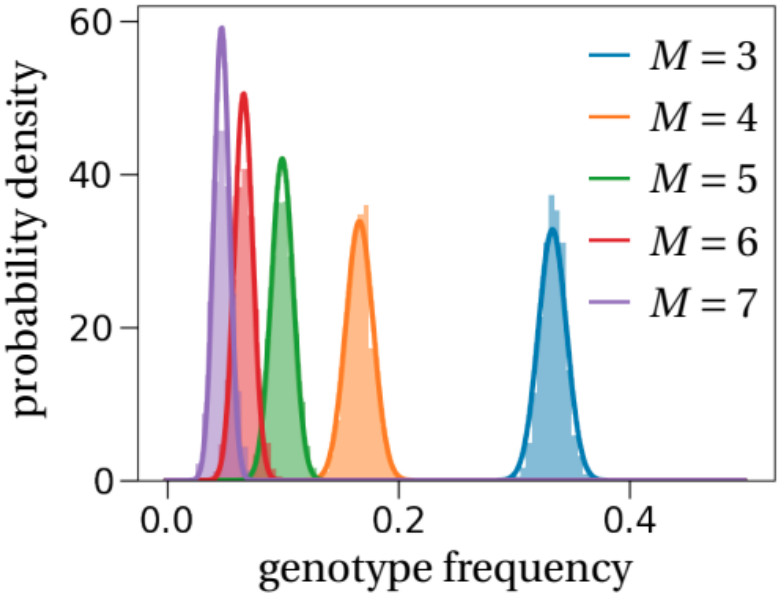
Genotypic frequencies

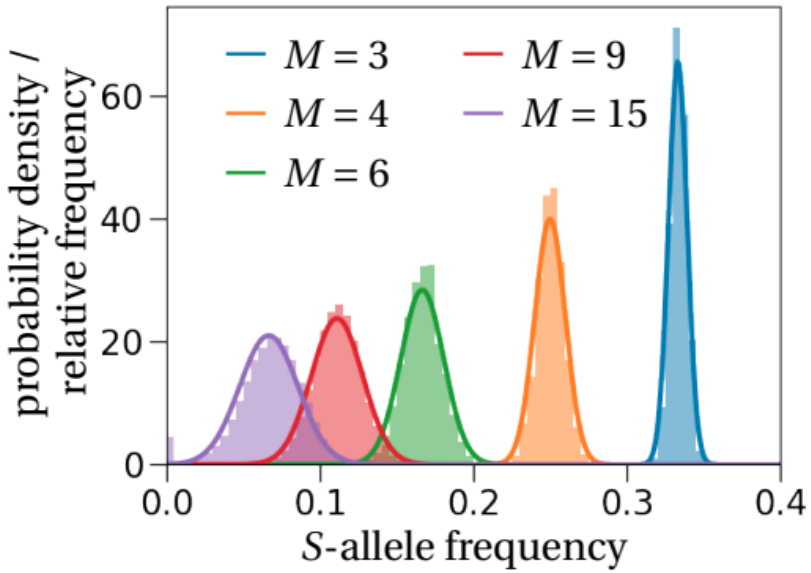
$$da_{ij}(t) = \mu_{ij}(t)dt + \sqrt{\frac{\sigma_{ij}^2(t)}{N}}dW_t^{ij}, \quad \longrightarrow \quad \psi_{ij}^{*} \sim \begin{cases} \mathcal{N}\left(\frac{2}{M(M-1)}, \frac{4(M-2)(M+1)}{3NM^2(M-1)^2}\right), & M=3, \\ \mathcal{N}\left(\frac{2}{M(M-1)}, \frac{2(M-2)(M+1)}{NM^2(M-1)^2}\right), & M\geq 4, \end{cases}$$

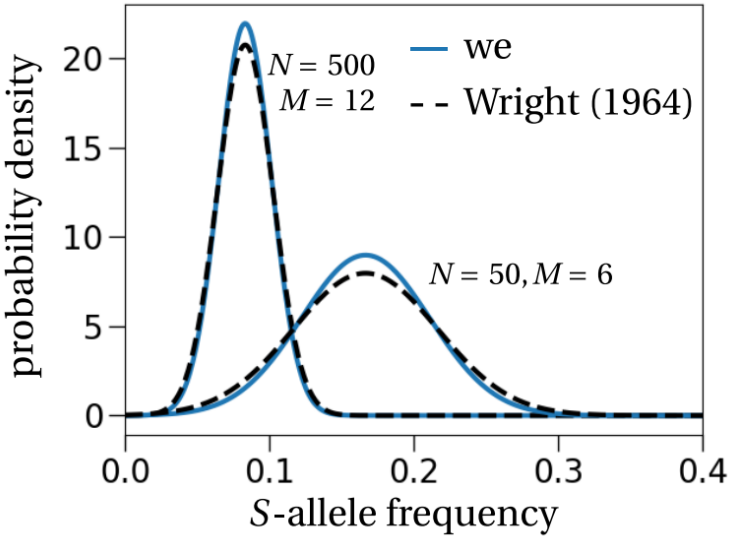
Allelic frequencies

$$\mathrm{d}p_i(t) = \mu_i(t)dt + \sqrt{\frac{\rho_i^2(t)}{N}}dW_t^{i}, \quad \longrightarrow \quad \psi \sim \mathcal{N}\left(\frac{1}{M}, \frac{(M-2)^3}{NM^2(2M-3)}\right),$$

Stationary distributions







Rates and approximations

Birth rate

$$T_{ij}^{+} = \left(\underbrace{\frac{1}{2}\sum_{k\neq i,j} A_{jk}\frac{p_i}{1-p_j-p_k}}_{i\,\text{pollen fertilization}} + \underbrace{\frac{1}{2}\sum_{k\neq i,j} A_{ik}\frac{p_j}{1-p_i-p_k}}_{j\,\text{pollen fertilization}} \right)_{\text{non}-\{ij\}\,\text{replacement.}} \underbrace{\frac{N-A_{ij}}{N}}$$

Death rate

$$T_{ij}^{-} = \underbrace{\left(N - \frac{1}{2}\sum_{k\neq i,j} A_{ik}\frac{p_j}{1-p_i-p_k} - \frac{1}{2}\sum_{k\neq j,i} A_{jk}\frac{p_i}{1-p_j-p_k} \right)}_{\text{non}-\{ij\}\,\text{reproduction}} \underbrace{\frac{A_{ij}}{N}}_{\{ij\}\,\text{replacement}}.$$

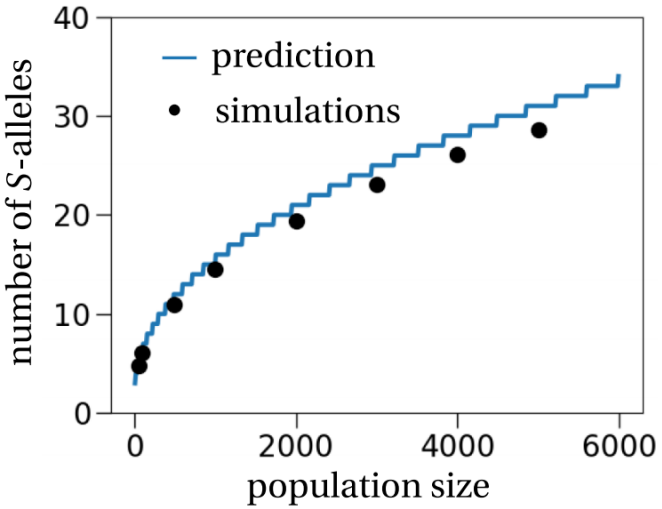
Approximations of the expected number of S-alleles.

Allelic frequencies

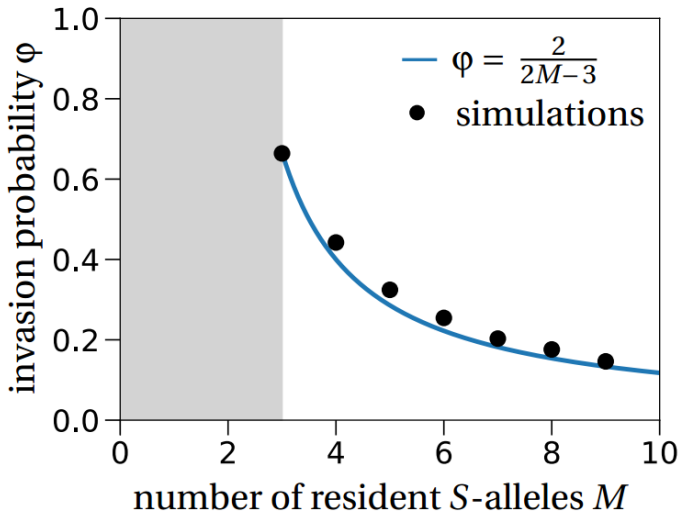
$$dp_i(t) = \mu_i(t)dt + \sqrt{\frac{\rho_i^2(t)}{N}}dW_t^i, \longrightarrow \psi \sim \mathcal{N}\left(\frac{1}{M}, \frac{(M-2)^3}{NM^2(2M-3)}\right),$$

$$u = \underbrace{\frac{M}{N}}_{T^{-}\times M} \times \underbrace{\int_{-\infty}^{1/2N} \psi(x)dx}_{S-\text{allele at freq.}1/2N},$$

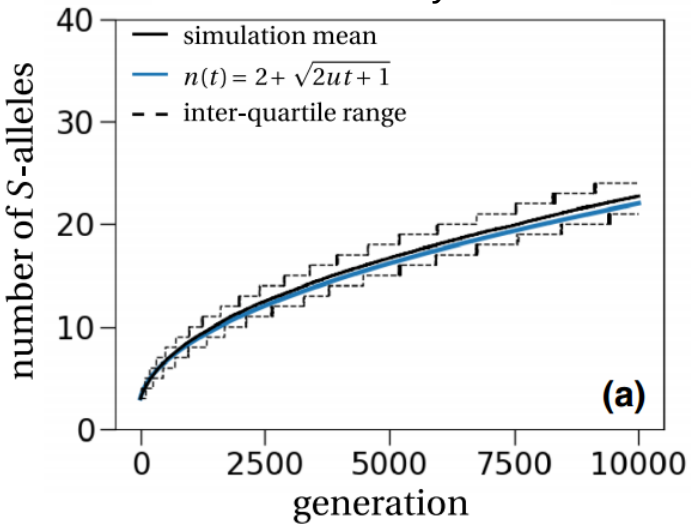
Expected diversity



Invasion probability



Diversification dynamics



Possible extensions and questions to address here in Bath

1. Towards the stationary distribution of the number of S-alleles M
2. Individual plants distributed in a continuous space and dispersal
3. Joint inference of the number of S-alleles $\widehat{M}$ and population size $\widehat{N}$
4. S-alelle emergence and the origin of the mutation rate u

1. Stationary distribution of the number of S-alleles M
A Moran's model with mutation

Birth rate

$$T_{ij}^{+} = \left(\underbrace{\frac{1}{2}\sum_{k\neq i,j} A_{jk}\frac{p_i}{1-p_j-p_k}}_{i\text{ pollen fertilization}} + \underbrace{\frac{1}{2}\sum_{k\neq i,j} A_{ik}\frac{p_j}{1-p_i-p_k}}_{j\text{ pollen fertilization}} \right)_{\text{non}-\{ij\}\text{replacement.}} \underbrace{\frac{N-A_{ij}}{N}}$$

Death rate

$$T_{ij}^{-} = \underbrace{\left(N - \frac{1}{2}\sum_{k\neq i,j} A_{ik}\frac{p_j}{1-p_i-p_k} - \frac{1}{2}\sum_{k\neq j,i} A_{jk}\frac{p_i}{1-p_j-p_k} \right)}_{\text{non}-\{ij\}\text{ reproduction}} \underbrace{\frac{A_{ij}}{N}}_{\{ij\}\text{ replacement}} .$$

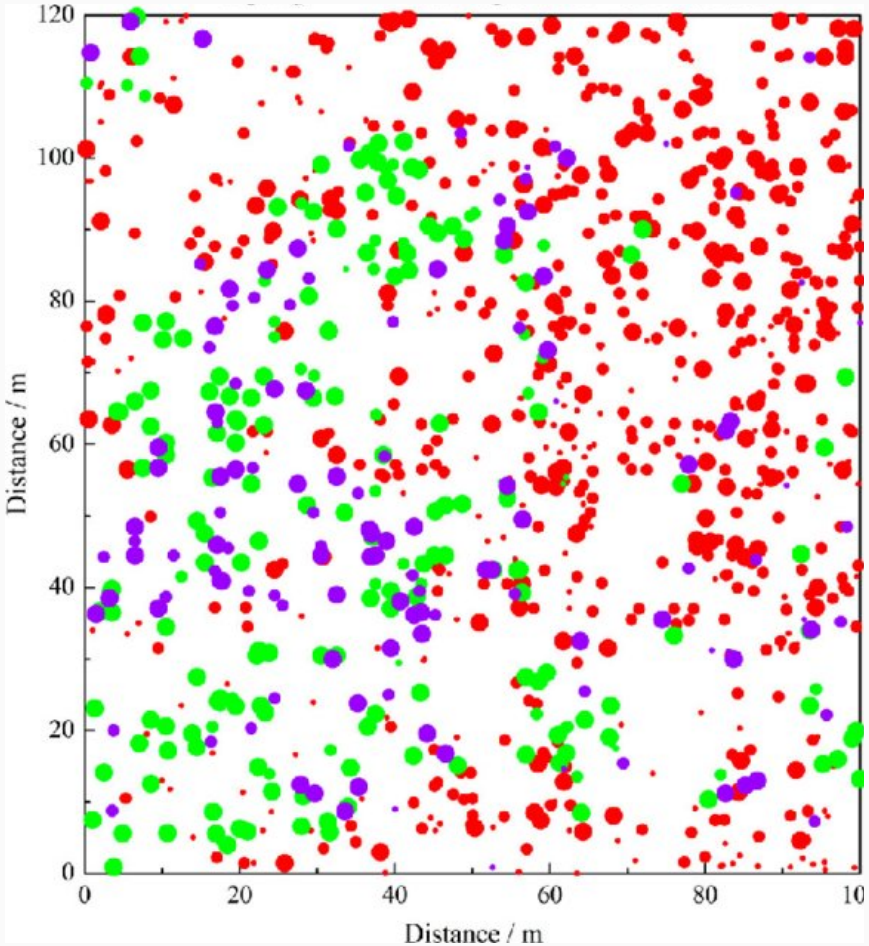
+ mutation / immigration

- mutation / immigration

Convergence?

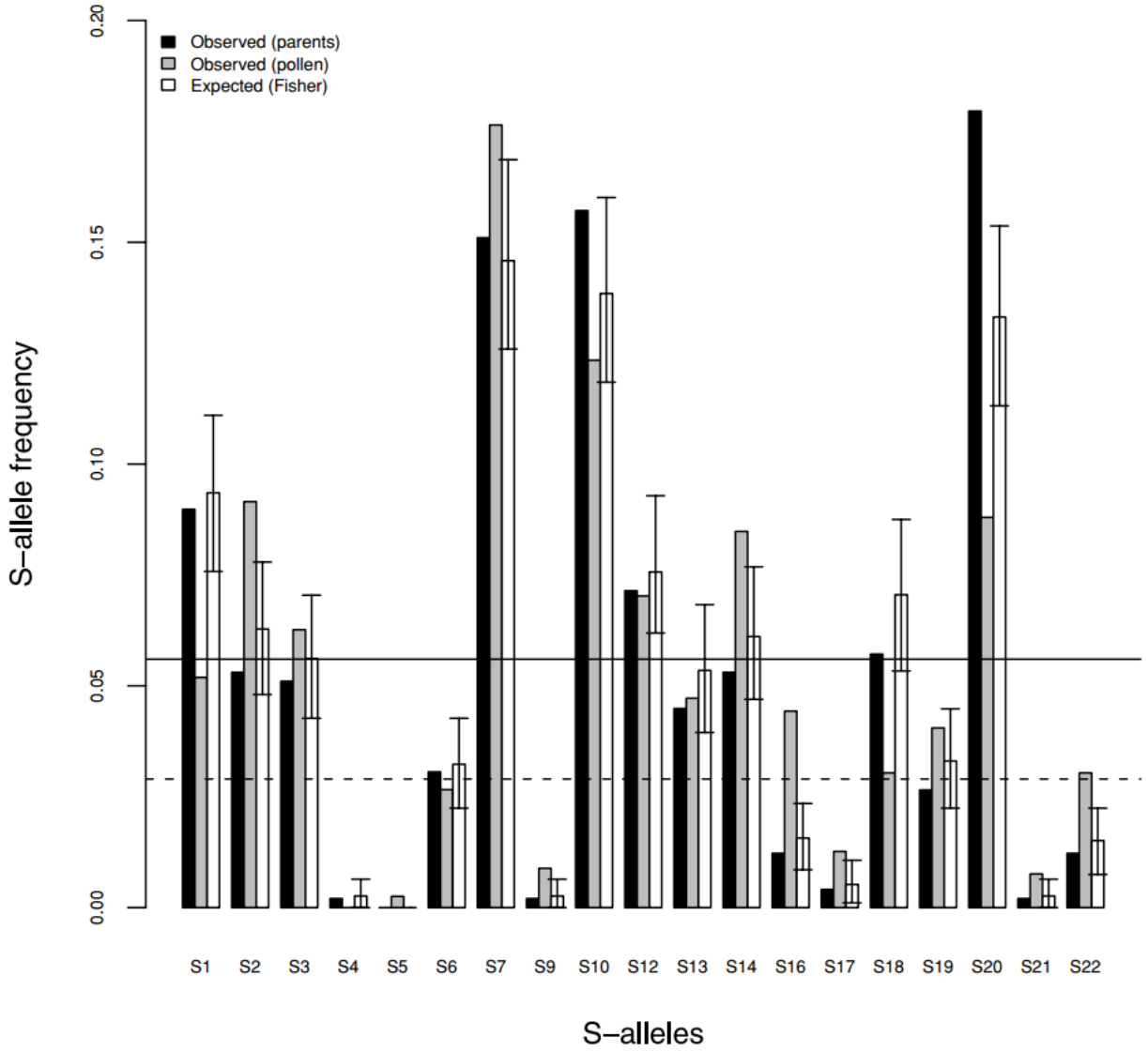
$$u = \underbrace{\frac{M}{N}}_{T^{-}\times M} \times \underbrace{\int_{-\infty}^{1/2N} \psi(x)\mathrm{d}x}_{S-\text{allele at freq.}1/2N} ,$$

2. Continuous spatial distribution



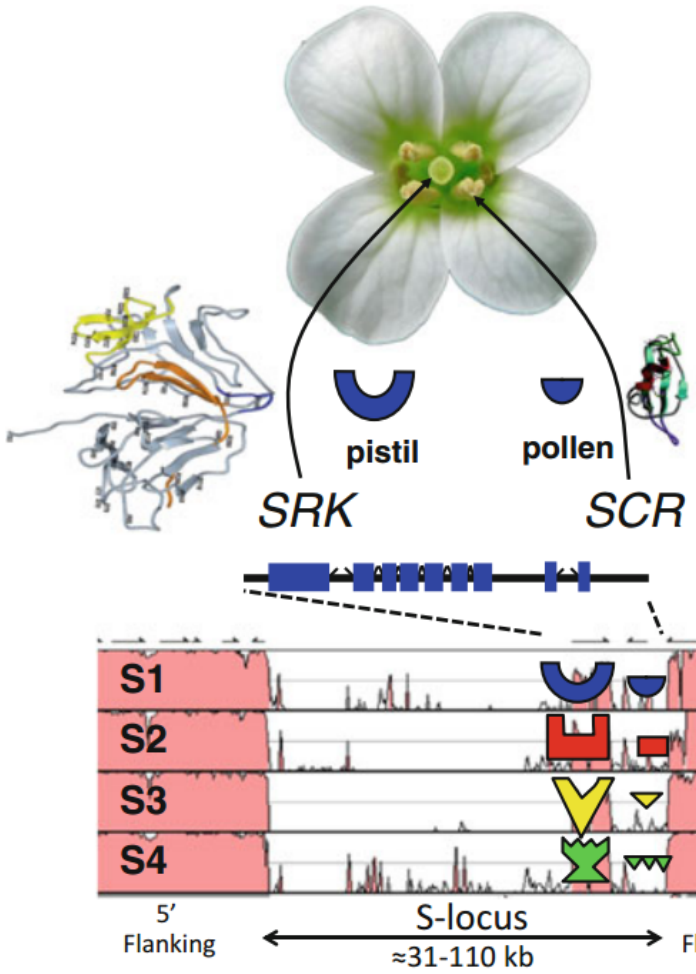
3. Estimator of M and N from a sample

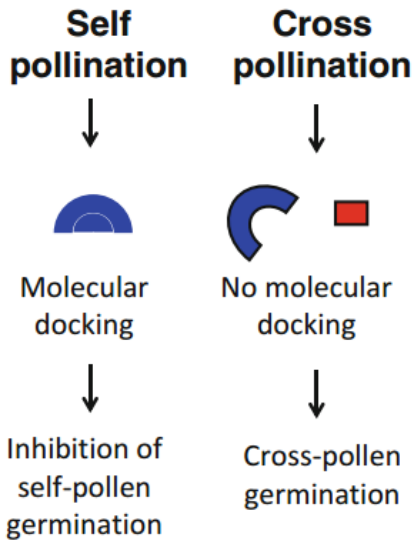
Prunus avium



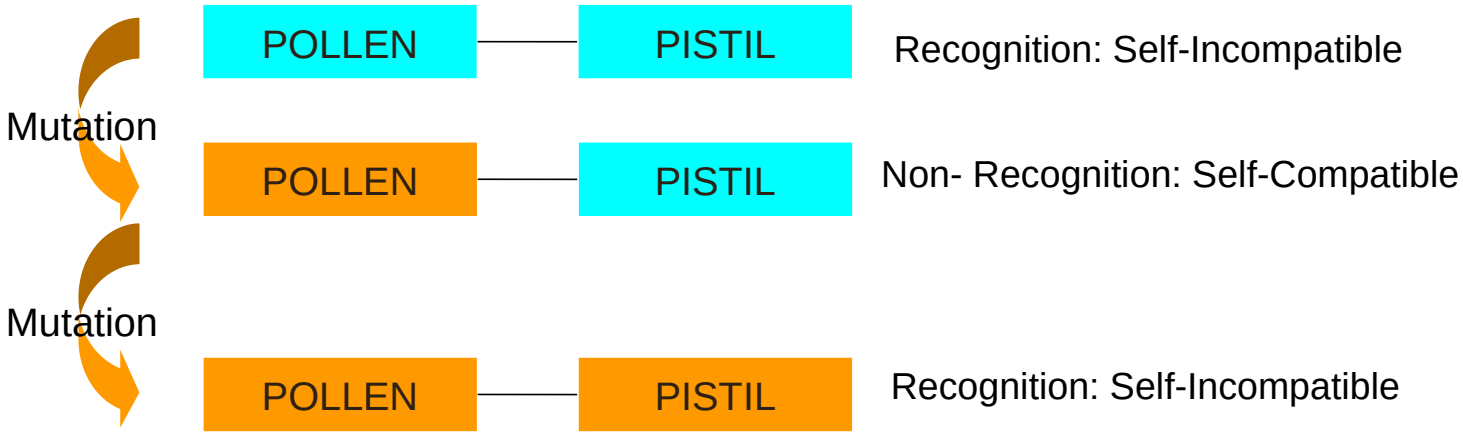
(Stoeckel et al. 2011)

4. Origin of the mutation rate u
key-lock mechanism / two genes - one locus

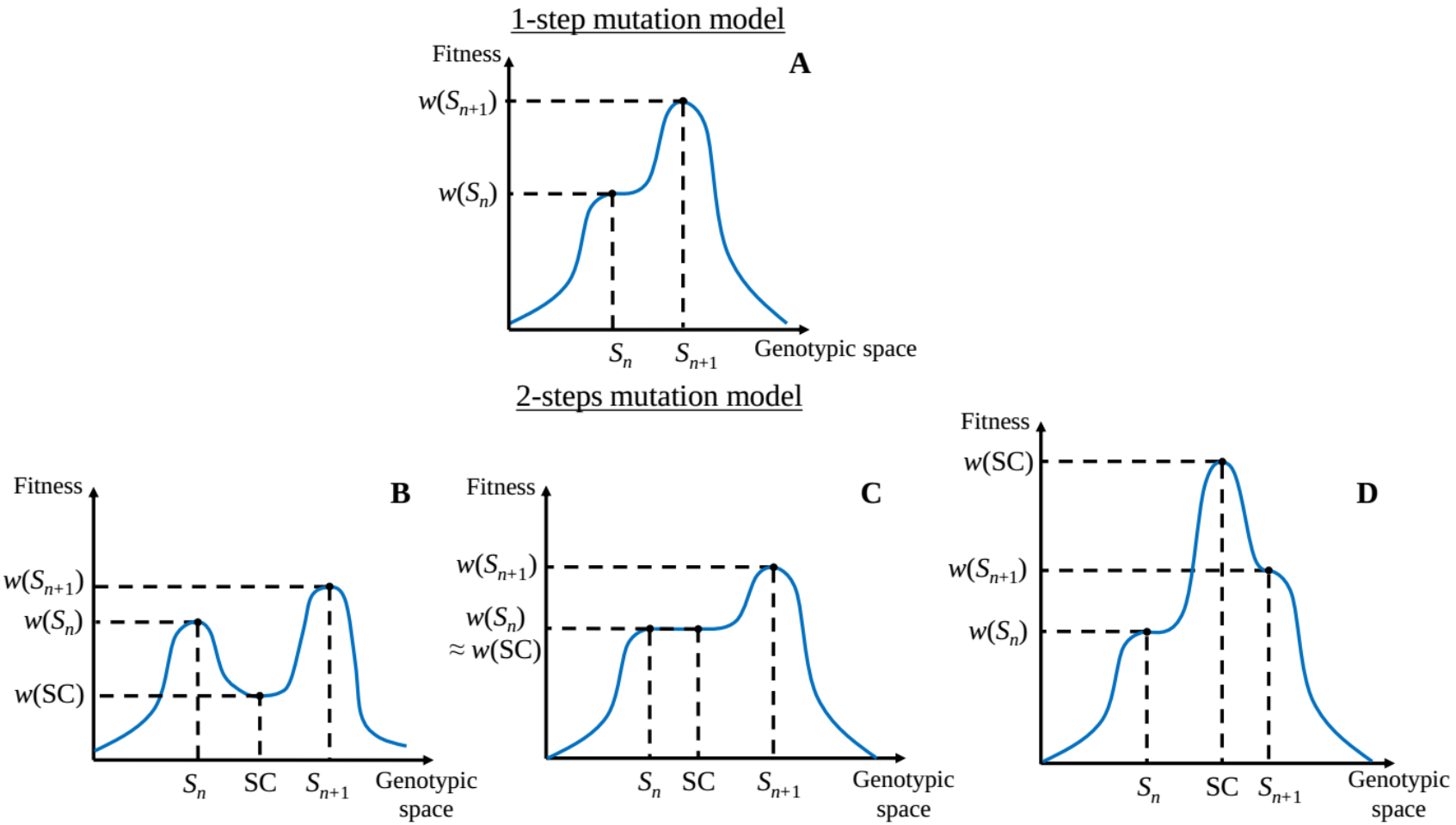




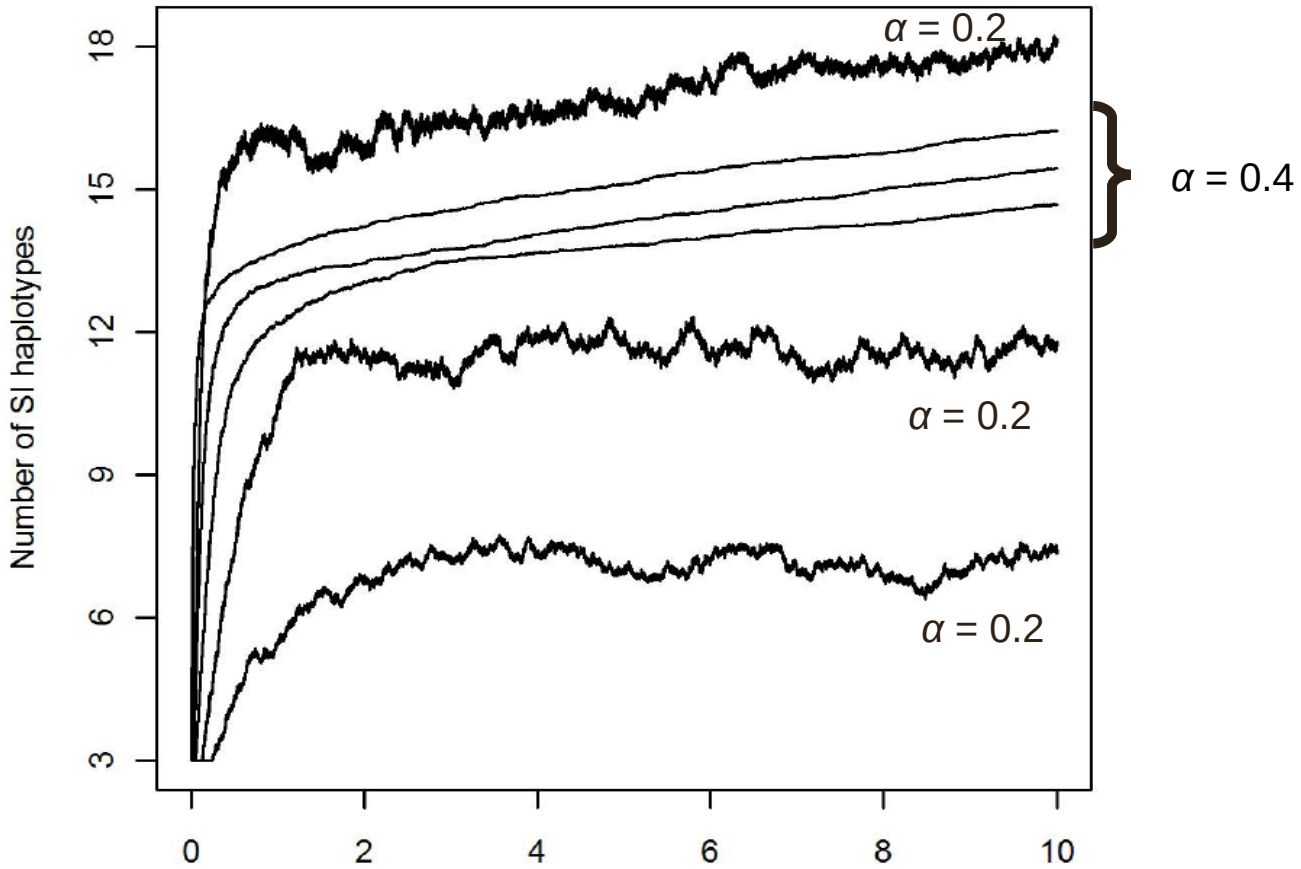
Where's the problem?



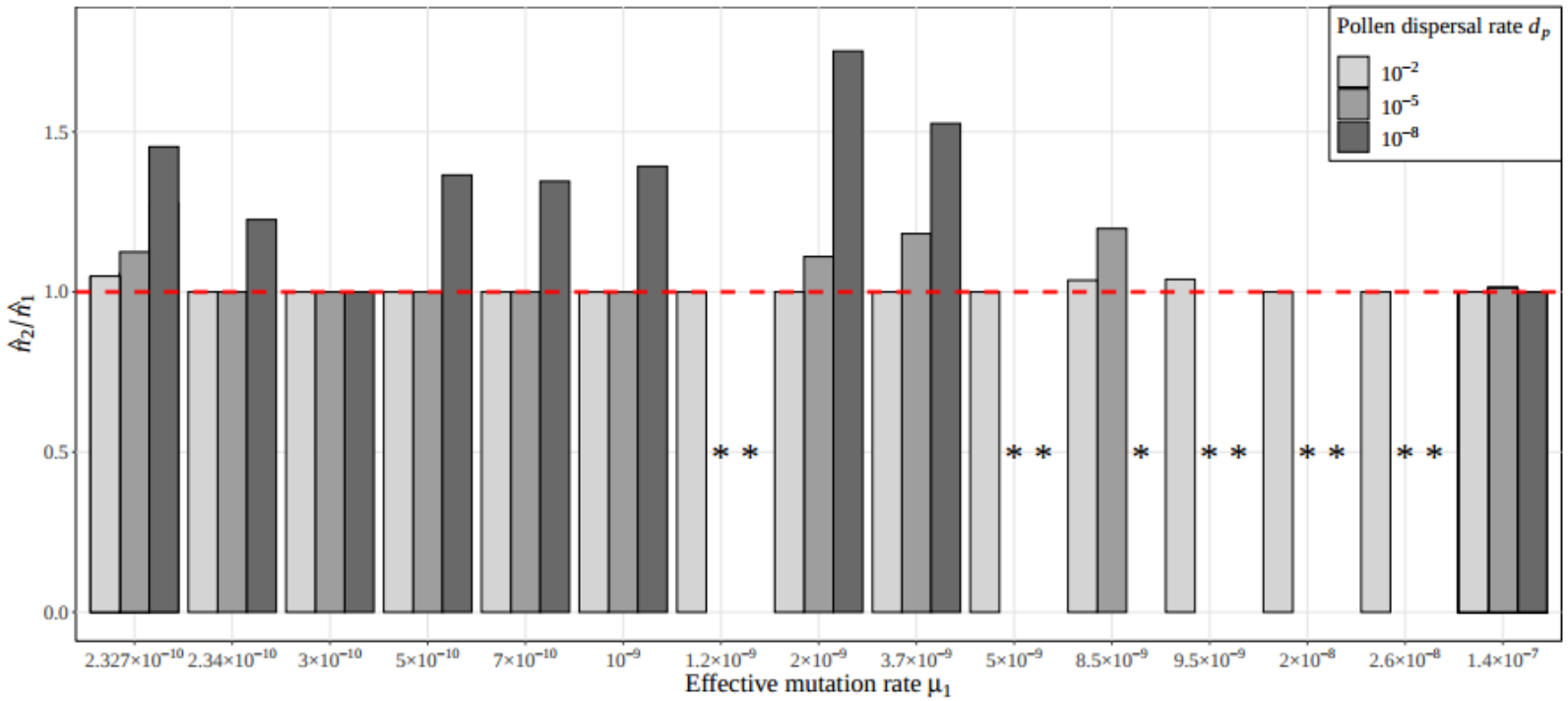
A crossing of fitness valley problem



Diversification of S-alleles: Individual-Based stochastic simulations



In a deme-subdivided population: A puzzle



Self-incompatibility as an archetype for stochastic models in population genetics: An update

PDEs, stochastic computer simulations, ...

- Fisher (1958), Wright (1960)
- Wright(1964, 1966)
- Ewens (1964, 1966), Kimura and Crown (1964), Mayo (1966), Moran (1962), Yokoyama and Nei (1979)

But...

- Moran (1962): lack of rigour
- Wright (1964): mathematic justification are not important as far as "*it works*"

A very recent micro $\rightarrow$ macro derivation

- Czuppon and Billiard (2022): old models and new results
- Still much to do! Keep on running Self-Incompatibility for population genetics!