How so diverse?

The role of chance in the evolution of self-incompatibility systems in flowering plants

Sylvain Billiard

EverEvol Workshop — 2022

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

Issues in $<$*origins of life*$>$ **research**

Some steps in the process could have been *chancy*

others could have been deterministic but highly *contingent*

still others could have been *the only way*

(adapted from C. Mariscal 2021)

Issues in *<insert your own biological obsession here>* **research**

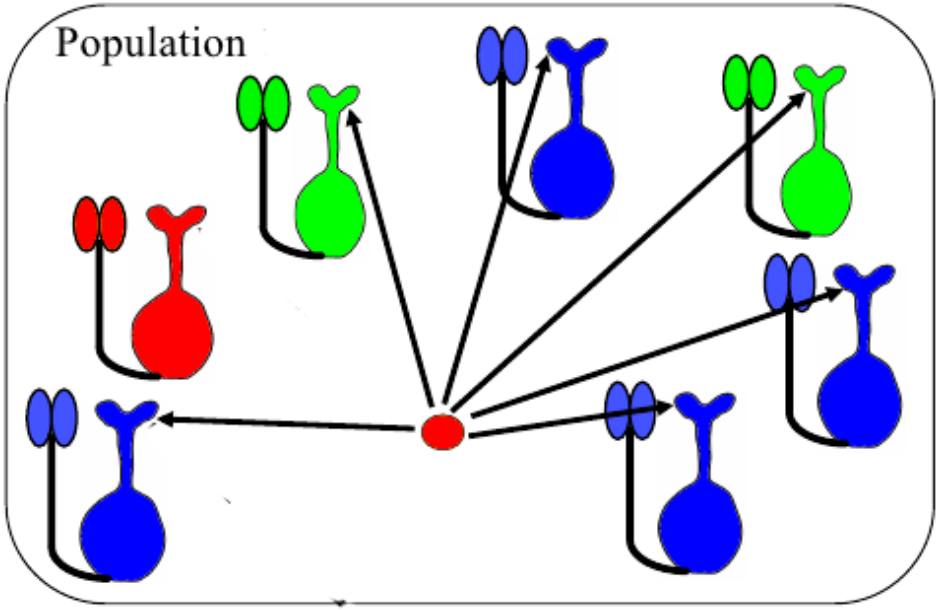
Some steps in the process could have been *chancy*

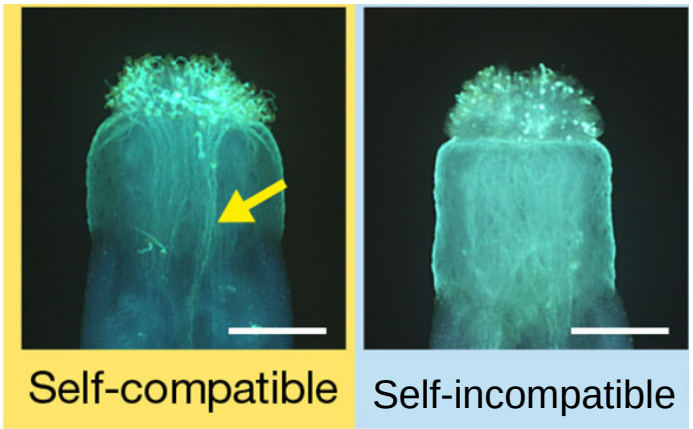
others could have been deterministic but highly *contingent*

still others could have been *the only way*

(adapted from C. Mariscal 2021)

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

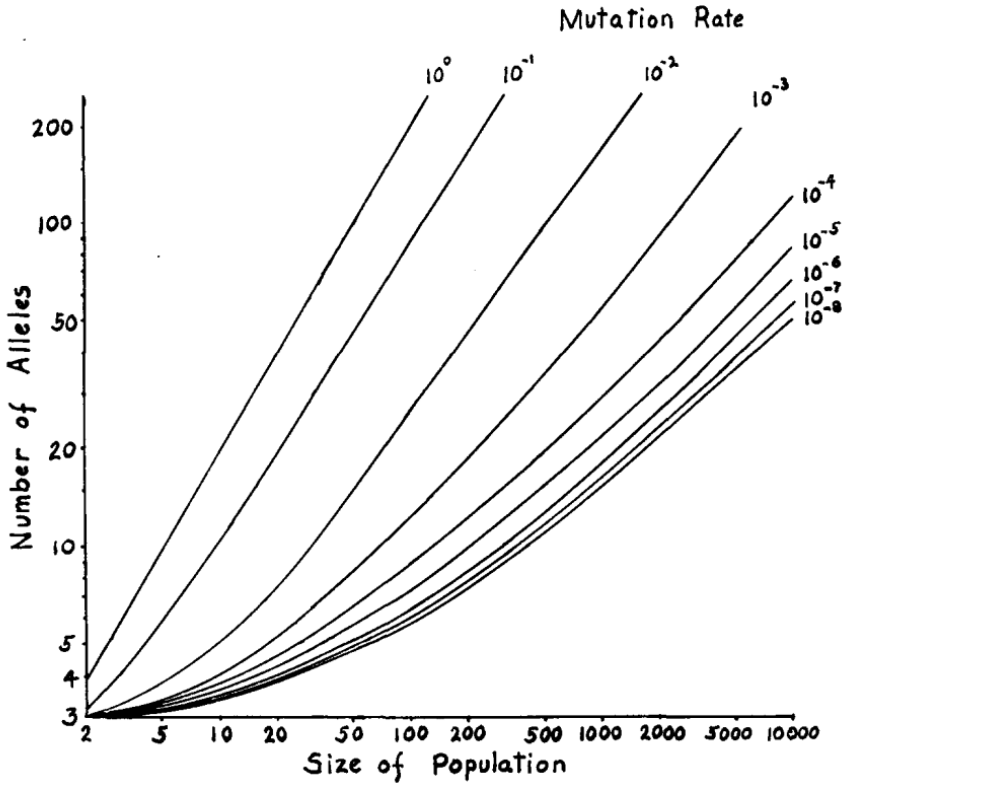
$$\Delta q = \frac{q(1-q)(1+R)}{2[R+q(1-R)]} - [1+u+m(1-q_t)]\,q + (v+mq_t)(1-q)$$

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

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

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

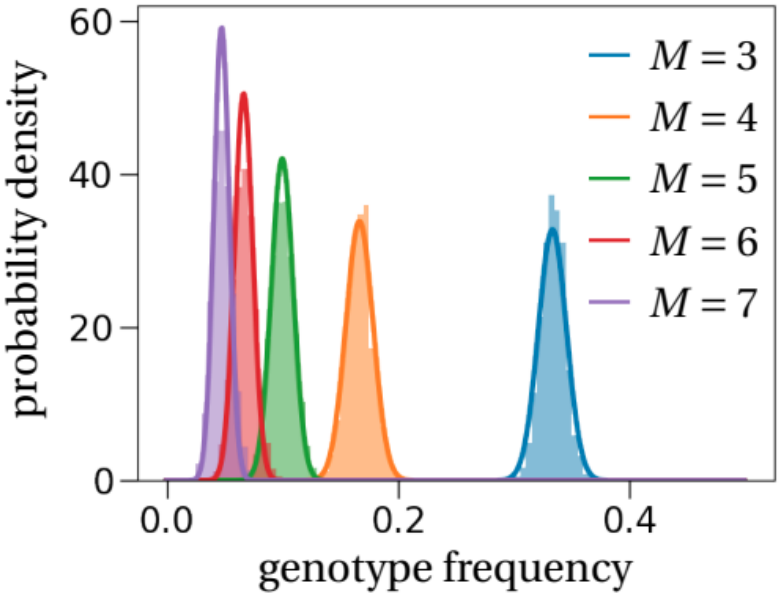
Genotypic frequencies

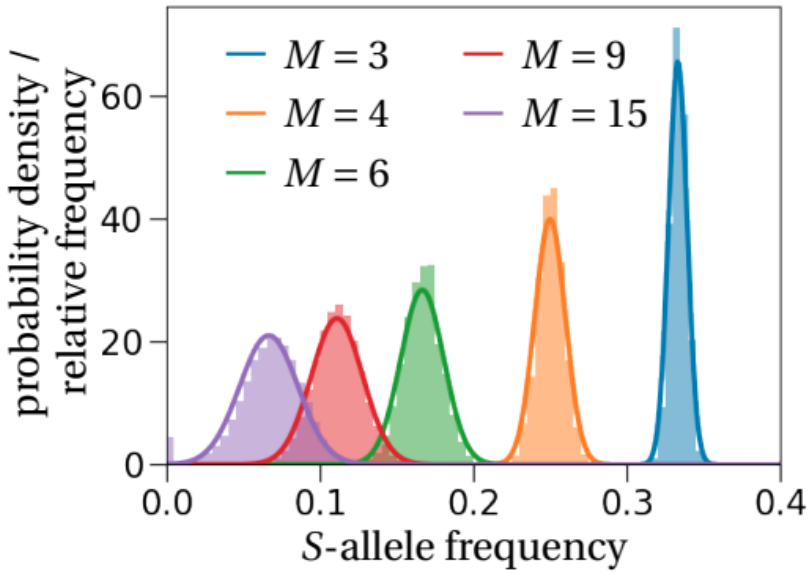
$$da_{ij}(t) = \mu_{ij}(t)dt + \sqrt{\frac{\sigma_{ij}^2(t)}{N}}dW_t^{ij}, \quad \rightarrow \quad \psi_{ij}^{*} \sim \begin{cases} \mathcal{N}\left(\frac{2}{M(M-1)}, \frac{4(M-2)(M+1)}{3NM^2(M-1)^2}\right), & M=3, \\ \mathcal{N}\left(\frac{2}{M(M-1)}, \frac{2(M-2)(M+1)}{NM^2(M-1)^2}\right), & M\geq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

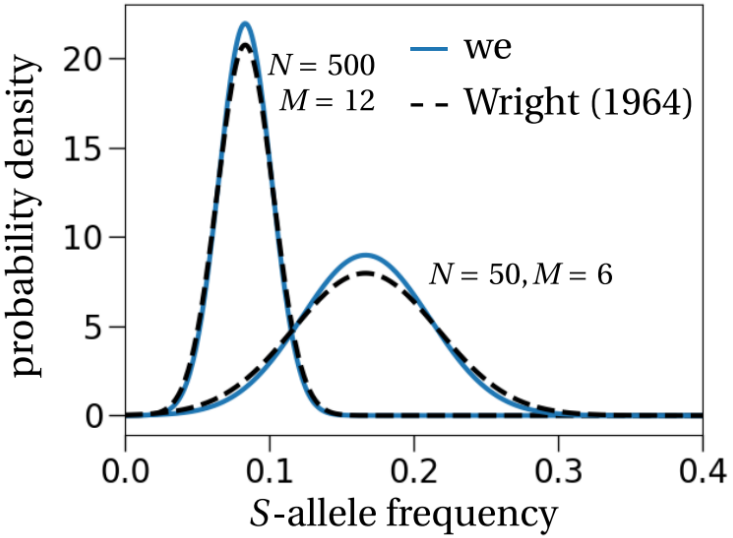
Allelic frequencies

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

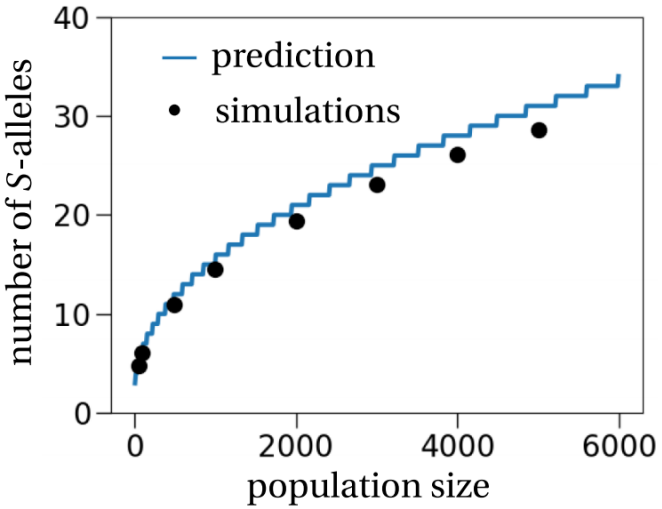
Stationary distributions



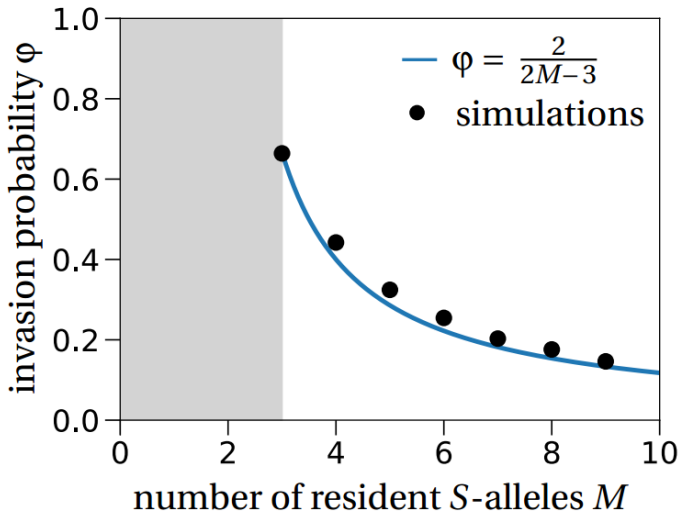




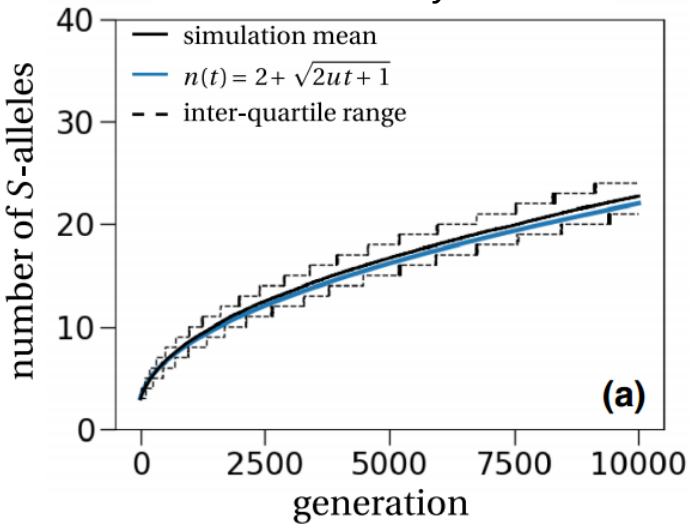
Expected diversity



Invasion probability



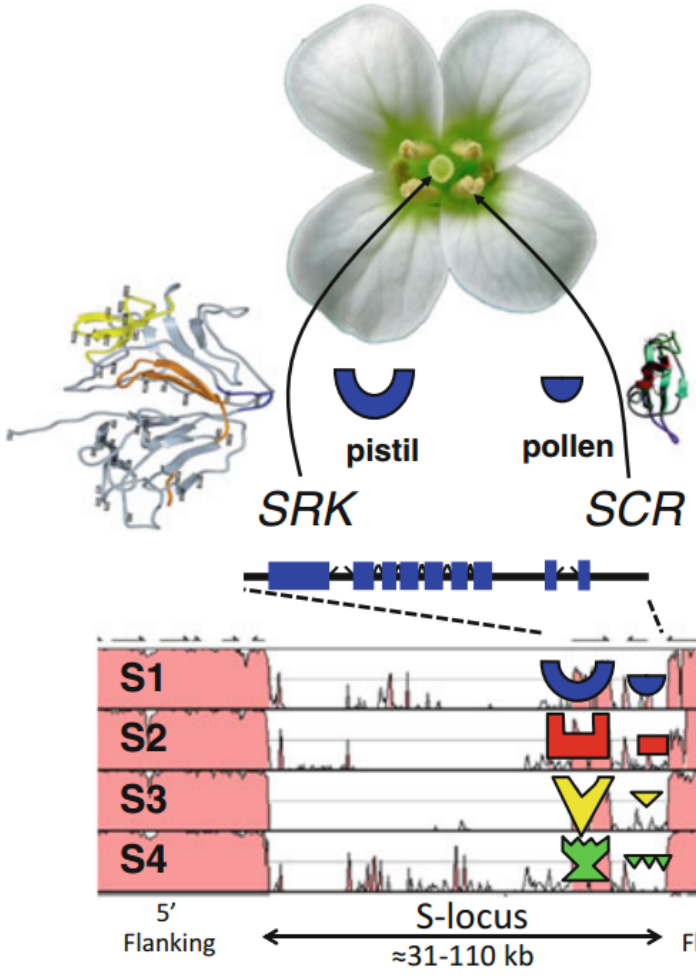
Diversification dynamics

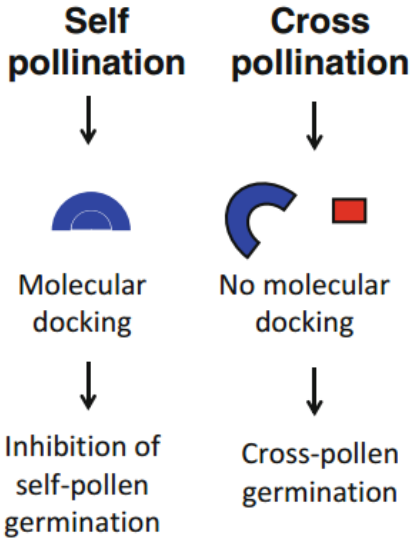


But...

- What's this mutation rate u?
- How new S-Alleles emerge from extant ones?
- Do genetic and genomic details matter?
- More generally: how do genetic and phenotypic novelties evolve?

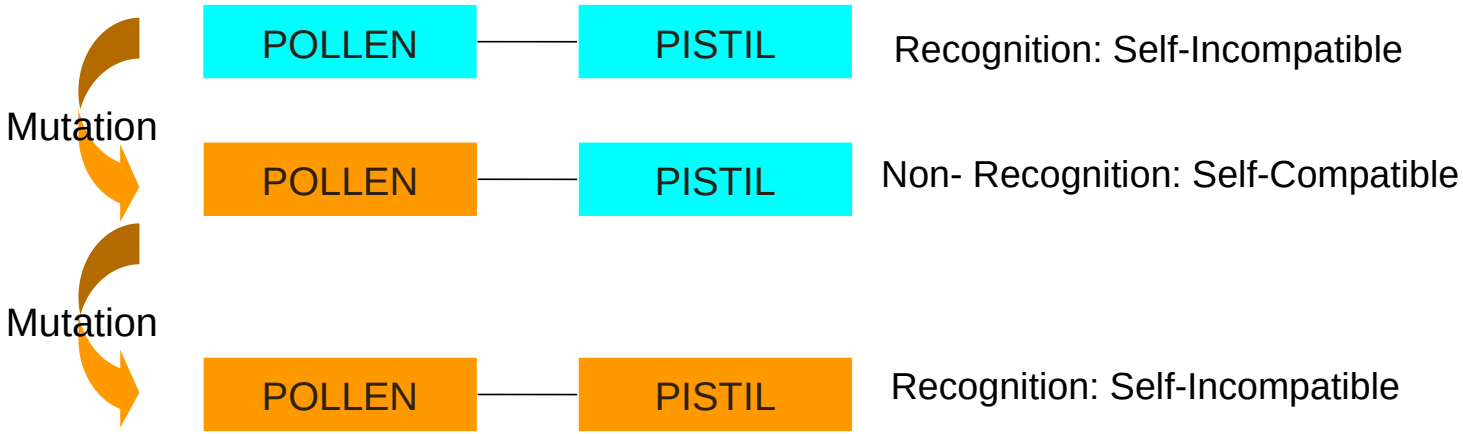
key-lock mechanism / two genes - one locus



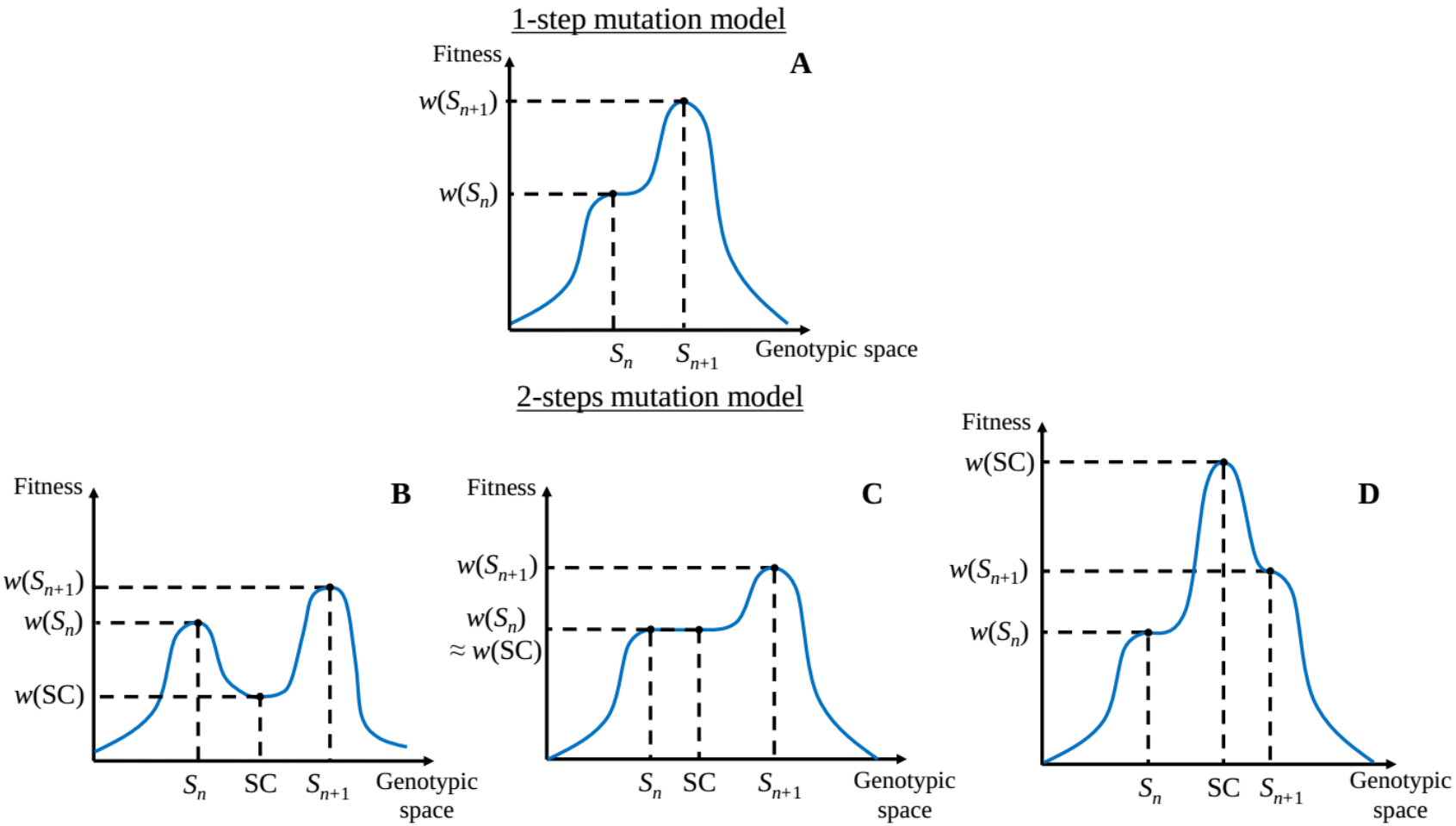


Papaveraceae, Prunus trees, Brassicaceae, at least.

Where's the problem?

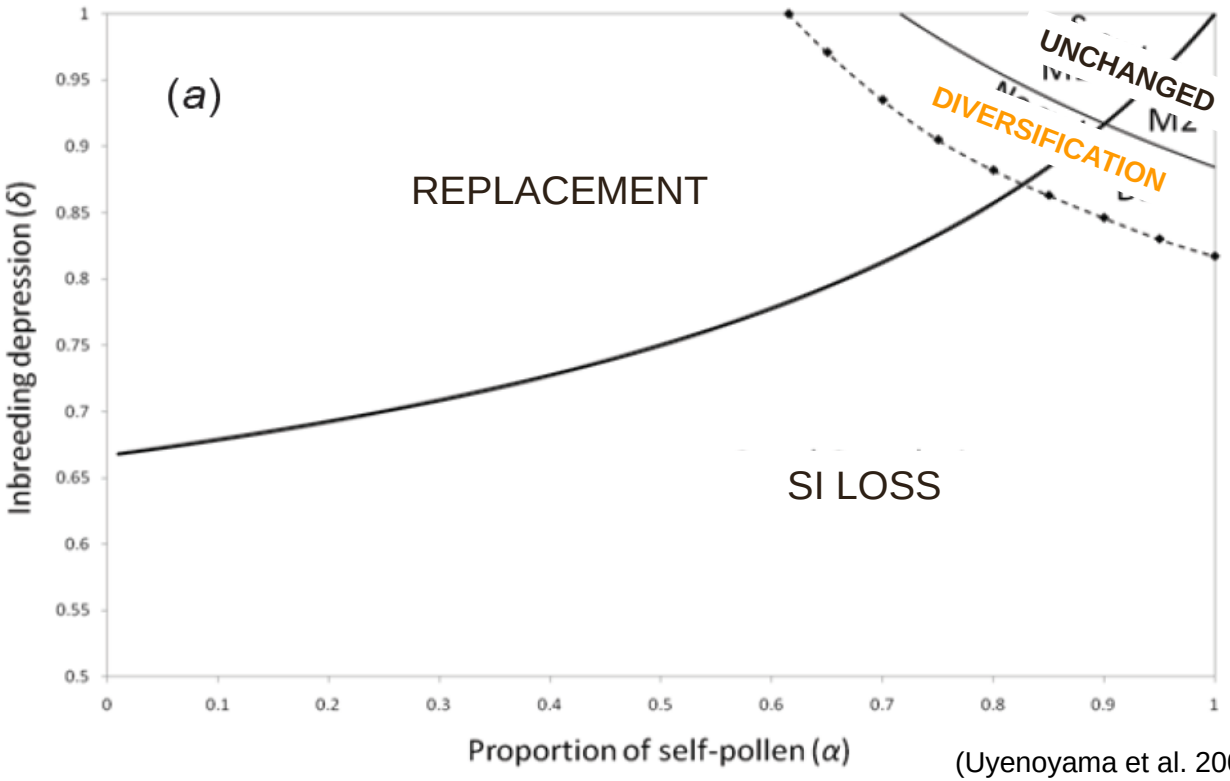


A crossing of fitness valley problem



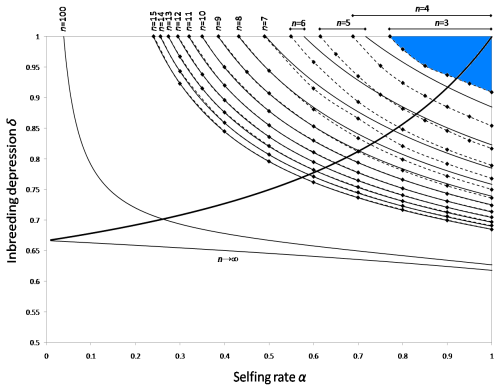
Diversification of S-alleles

$M = 5$

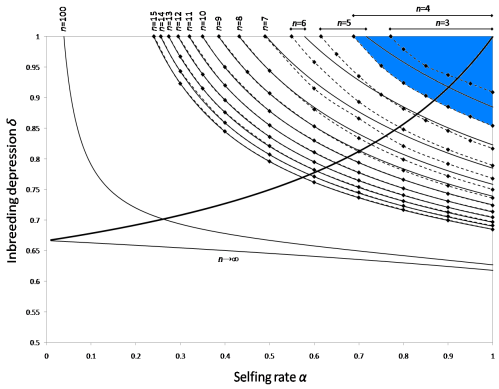


(Uyenoyama et al. 2001)
(Gervais et al. 2011)

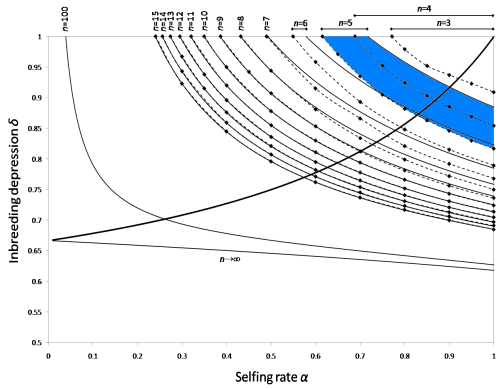
$M = 3$



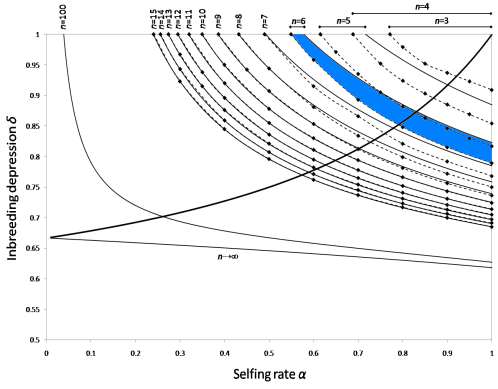
$M = 4$



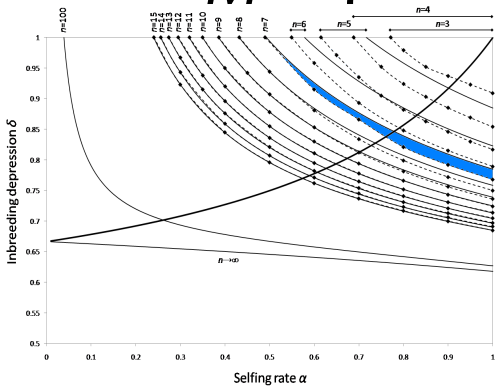
$M = 5$



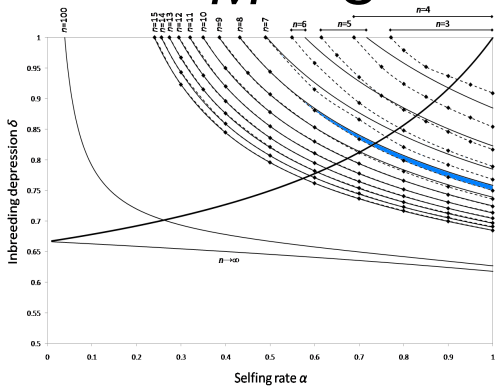
$M = 6$



$M = 7$



$M = 8$



(Gervais et al. 2011)

Stochastic tunneling (Weismann et al. 2009)

100 replicates

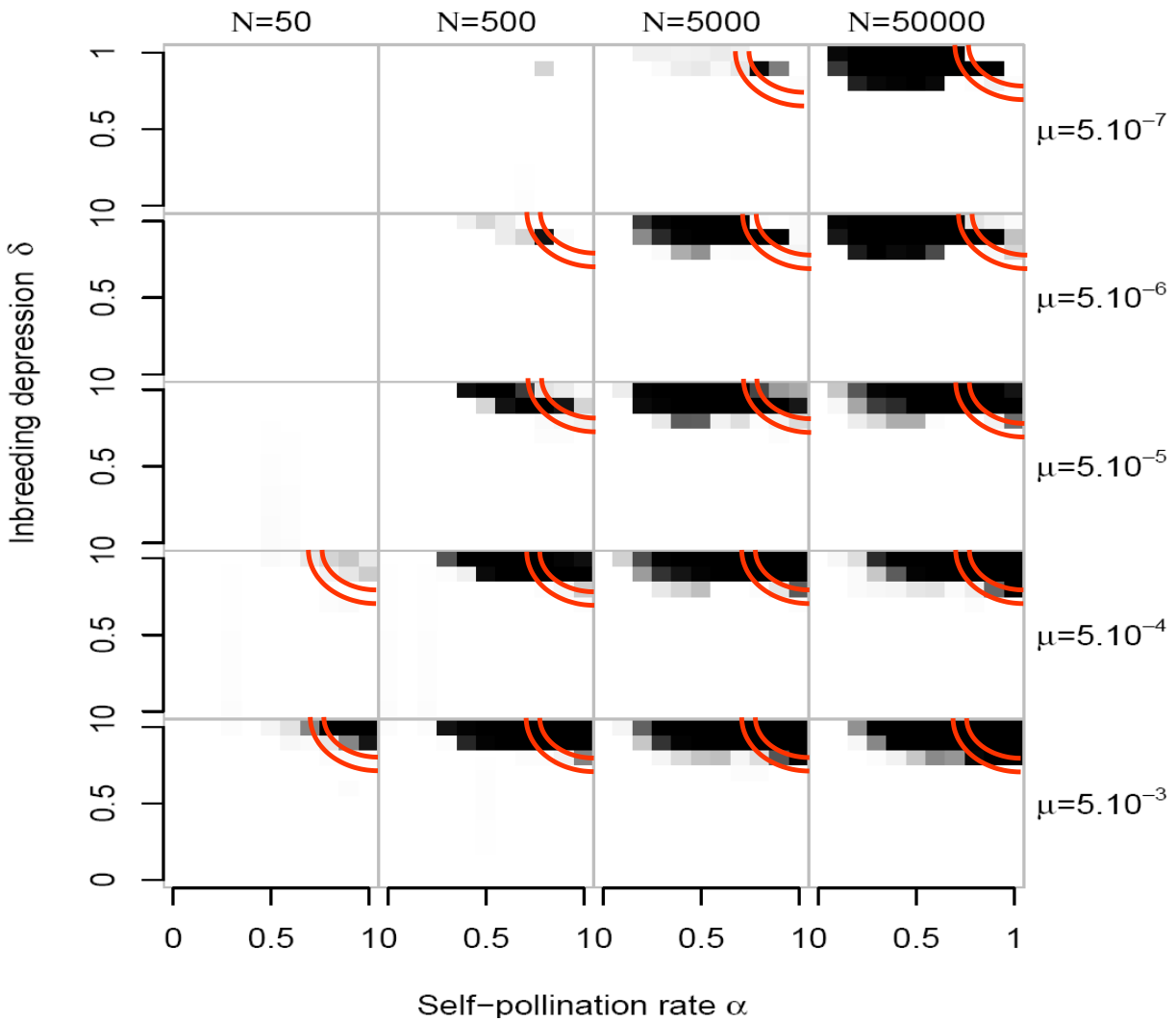
$n = 5$

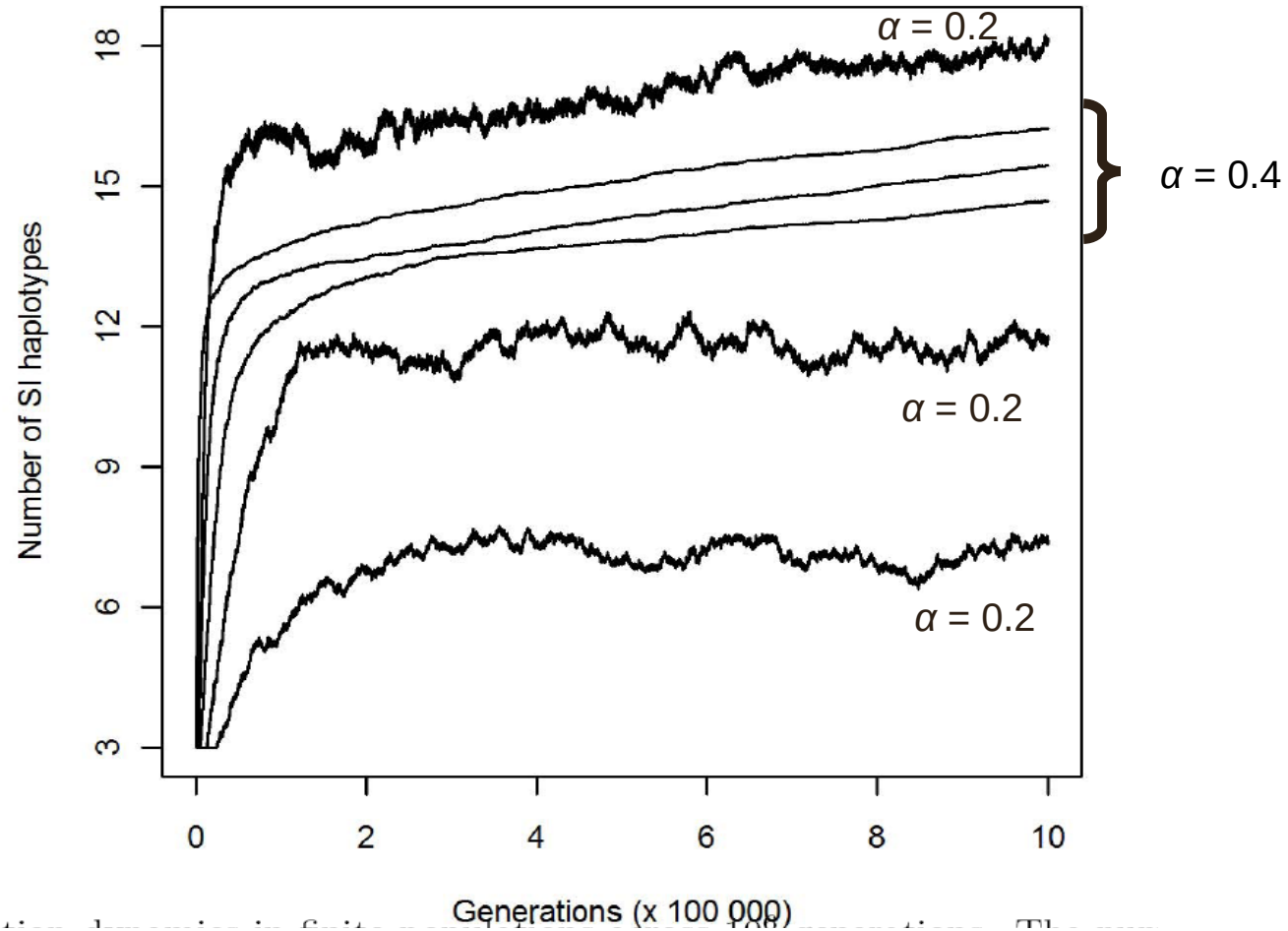
100% diversification

(n > 5)

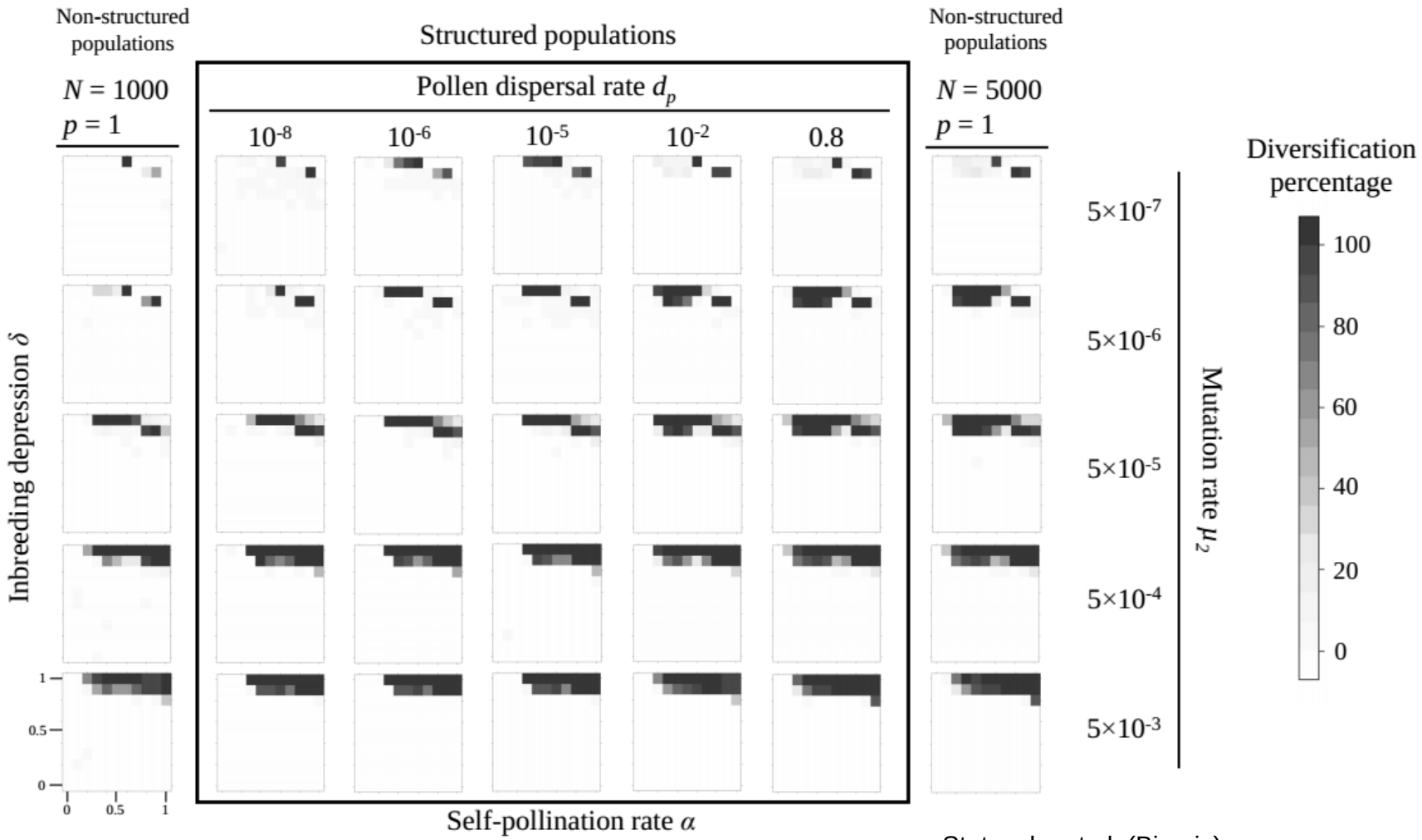
0% diversification

(0 < n < 5)

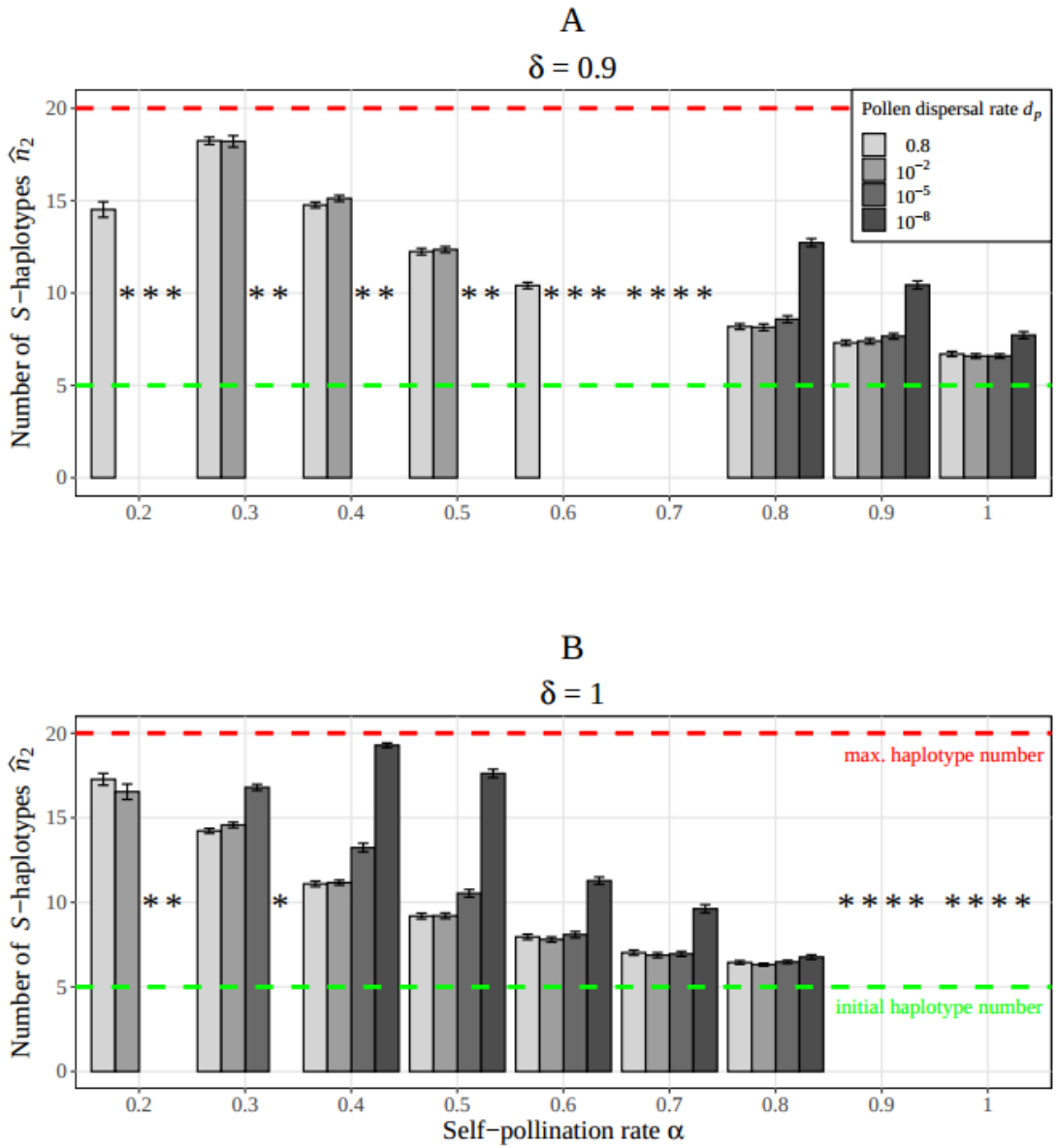



α = 0.2
α = 0.4
α = 0.2
α = 0.2
Number of SI haplotypes
3
6
9
12
15
18
0
2
4
6
8
10
Generations (x 100 000)

Let's consider a deme-subdivided population

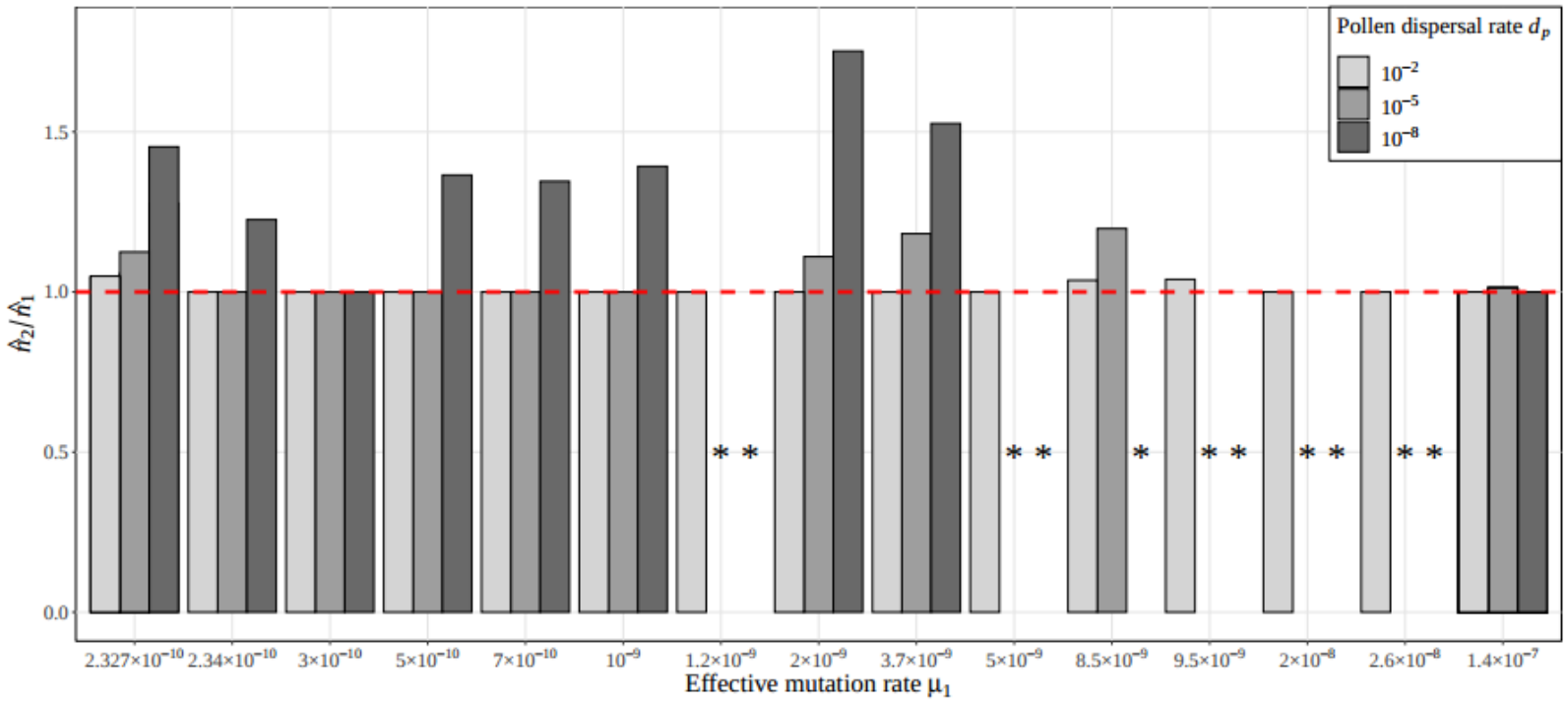


Stetsenko et al. (Biorxiv)



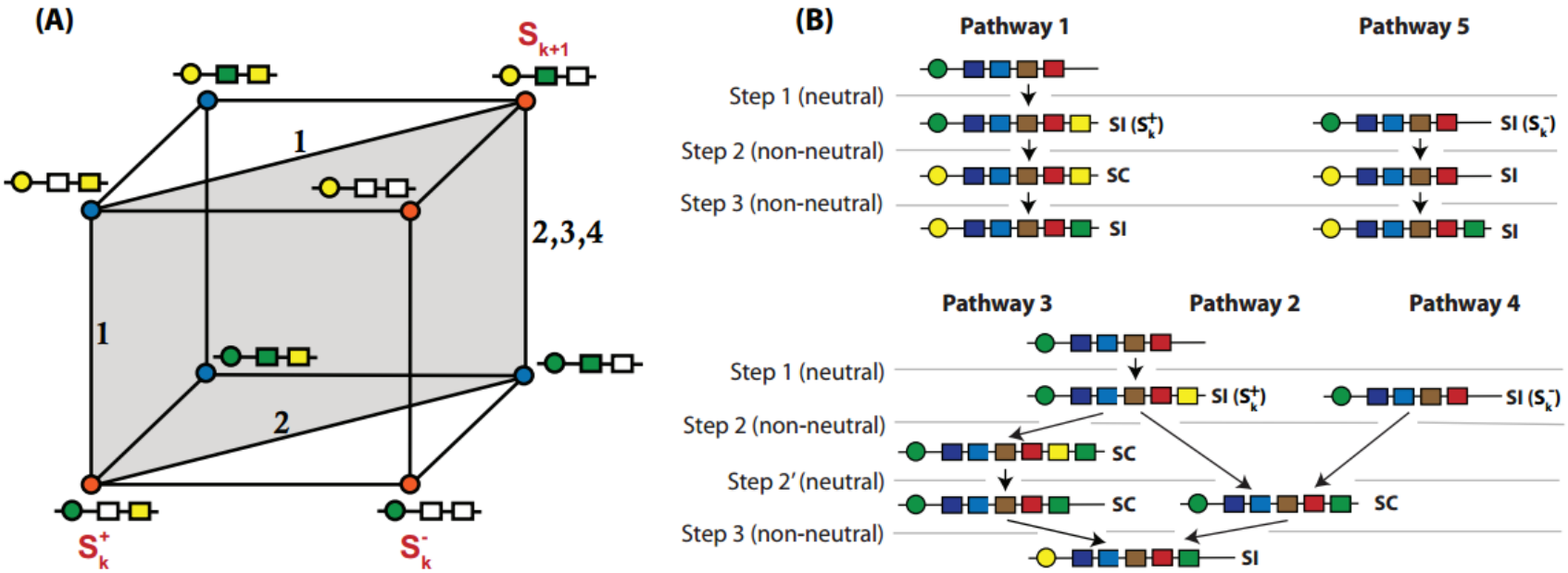
Stetsenko et al. (Biorxiv)

A puzzle


Pollen dispersal rate d_p
10^{-2}
10^{-5}
10^{-8}
$\hat{h}_2/\hat{h}_1$
1.5
1.0
0.5
0.0
2.327×10⁻¹⁰
2.34×10⁻¹⁰
3×10⁻¹⁰
5×10⁻¹⁰
7×10⁻¹⁰
10⁻⁹
1.2×10⁻⁹
2×10⁻⁹
3.7×10⁻⁹
5×10⁻⁹
8.5×10⁻⁹
9.5×10⁻⁹
2×10⁻⁸
2.6×10⁻⁸
1.4×10⁻⁷
Effective mutation rate μ_1

Alternative genetic mechanism (e.g. Solanaceae)

(for Muhittin)



SRNase/F-box non-self recognition self-incompatibility (Bodova et al. 2018)

Mostly contingency?

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)

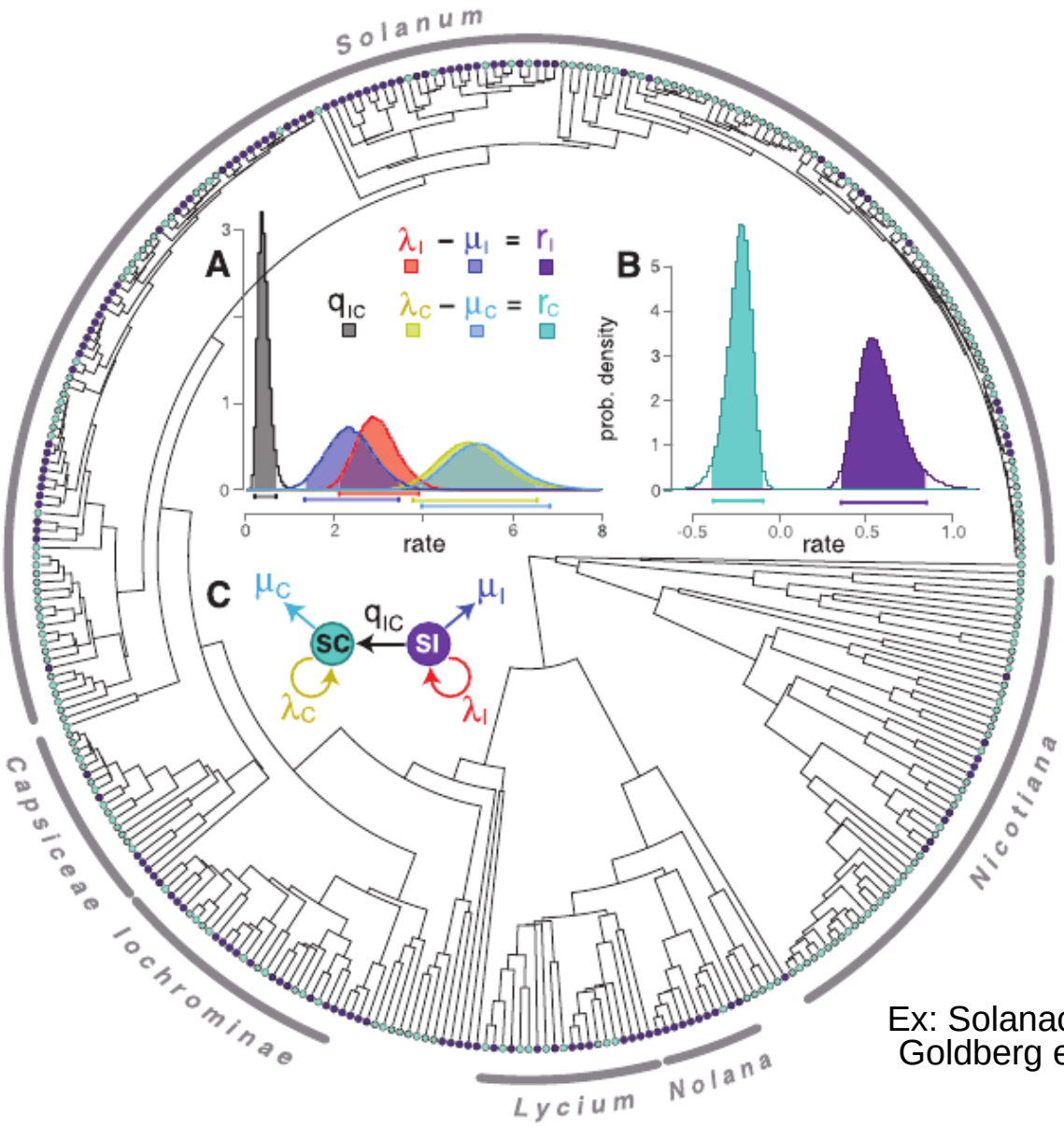
Arabidopsis halleri

66 S-alleles ~ 900 sampled individuals

(V. Castric, Pers. Comm.)

The role of chance and rare events: related puzzles

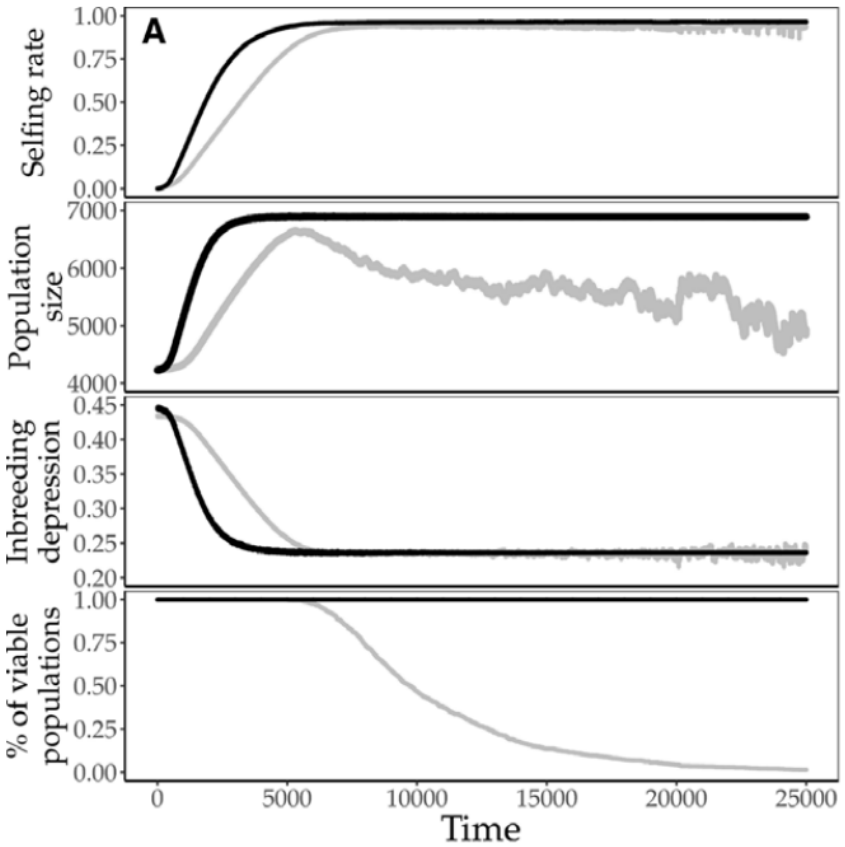
1. Self-incompatibility loss, extinction and speciation

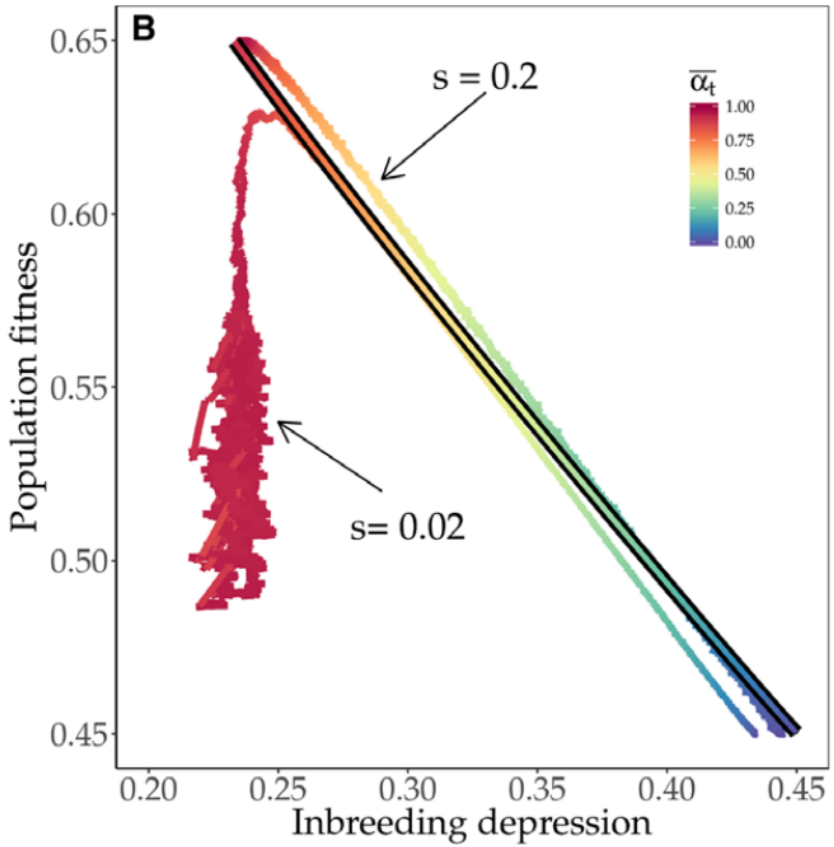


Ex: Solanaceae
Goldberg et al 2010

Let's add demography (eco-evolutionary model)

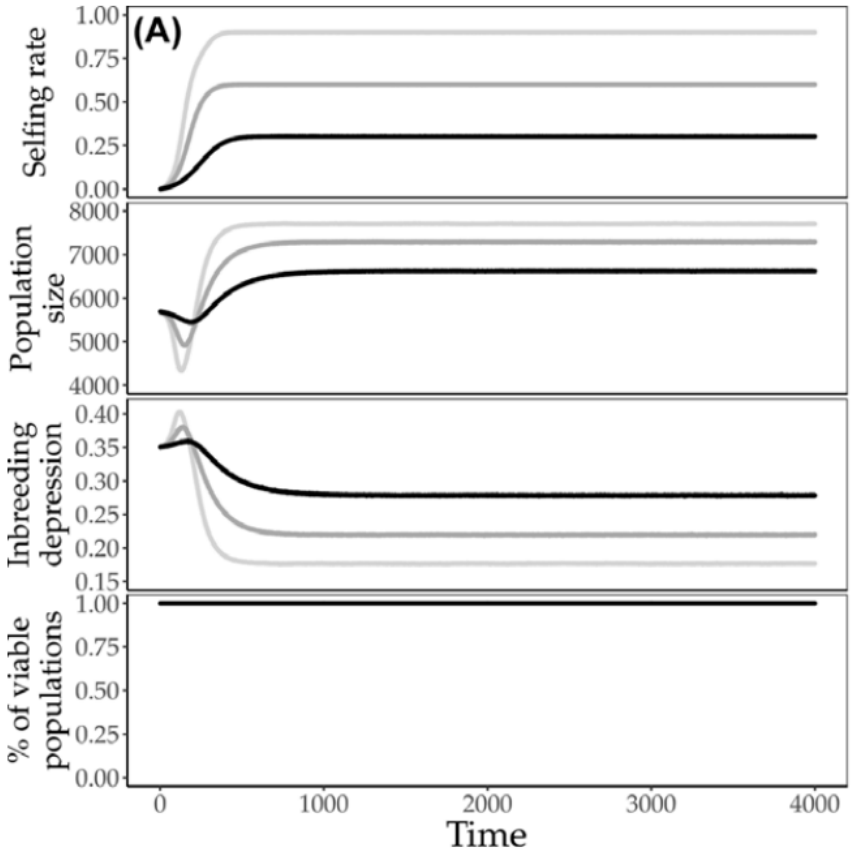
Small steps evolution of self-fertilization

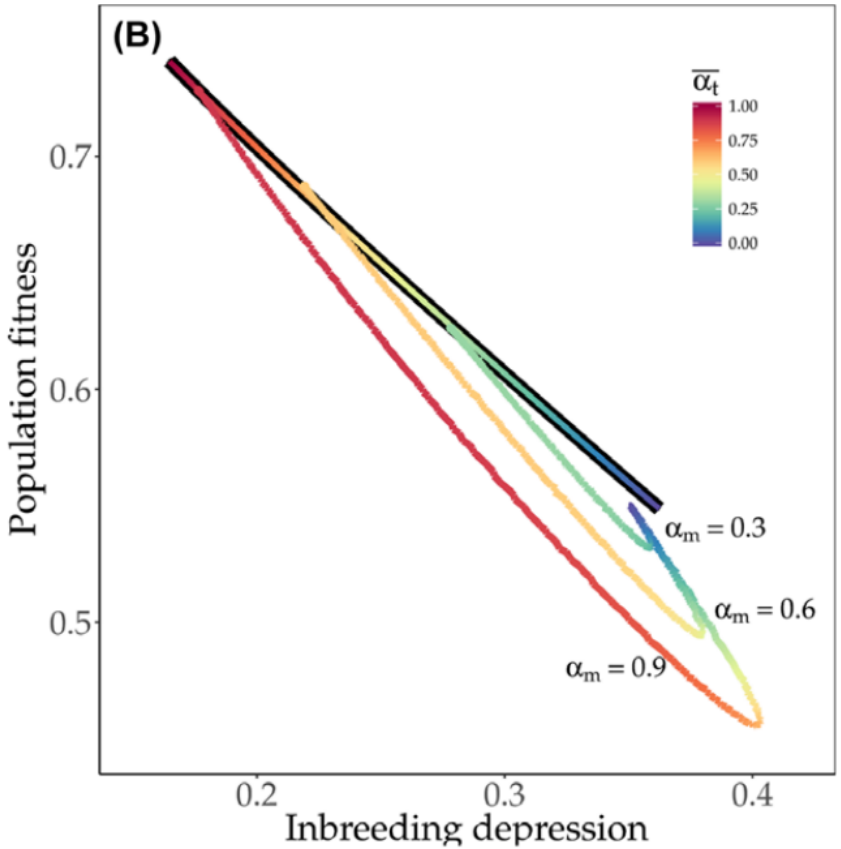




Abu Awad and Billiard (2017)

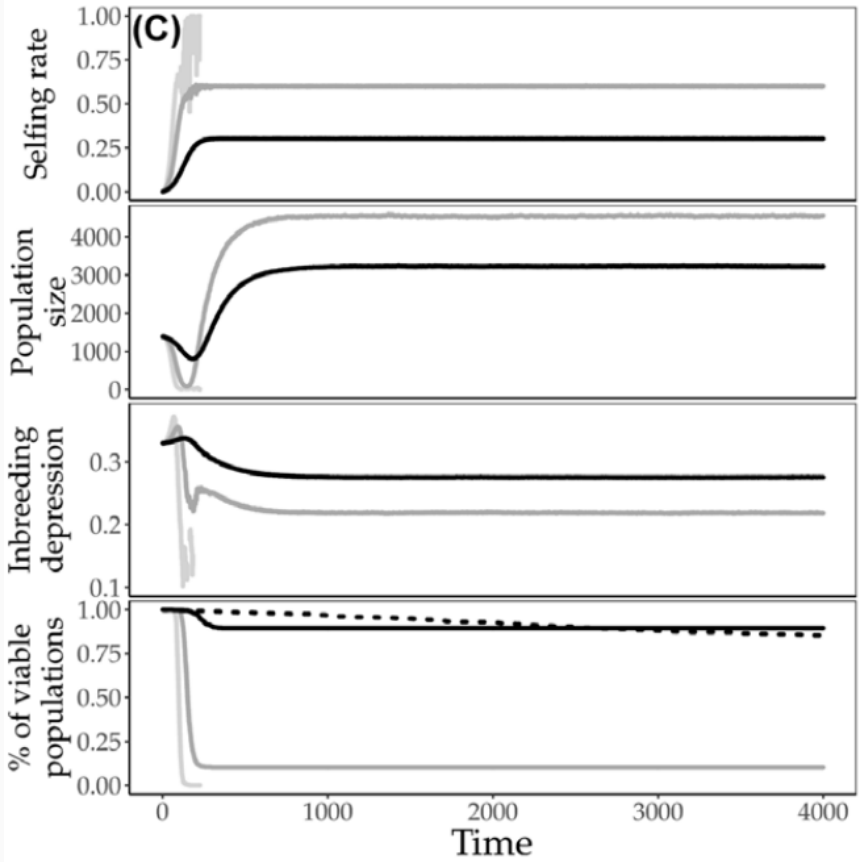
Large steps evolution of self-fertilization – Large reproductive rate

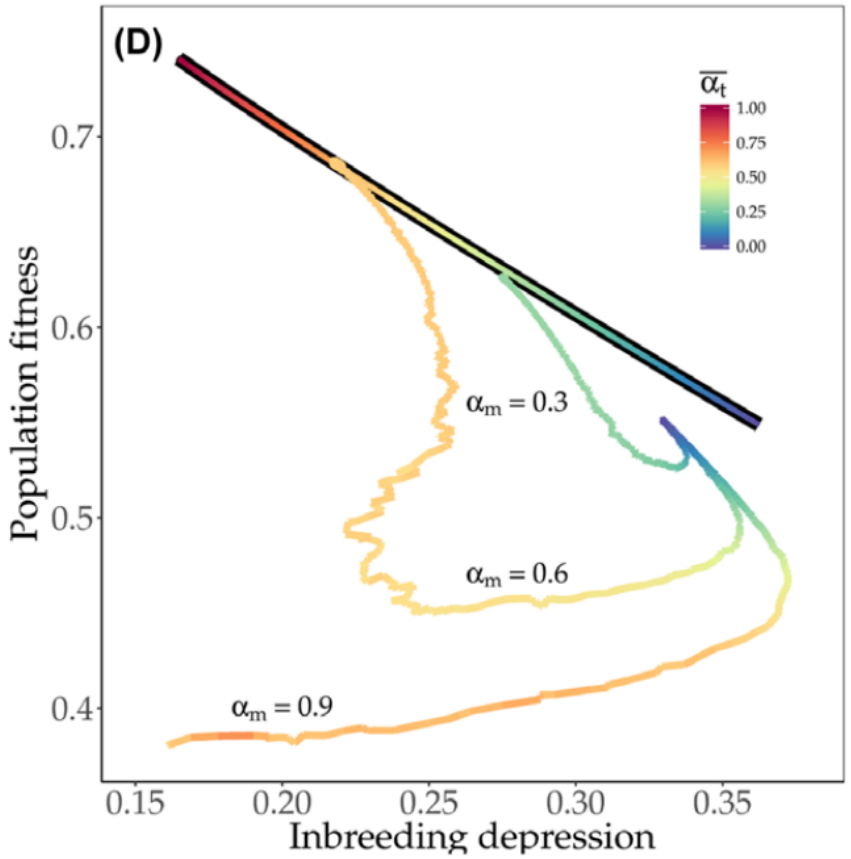




Abu Awad and Billiard (2017)

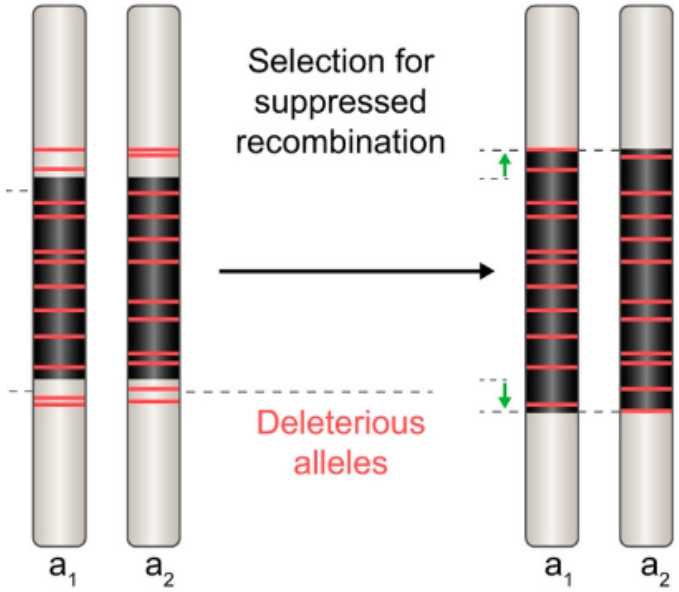
Large steps evolution of self-fertilization – Small reproductive rate



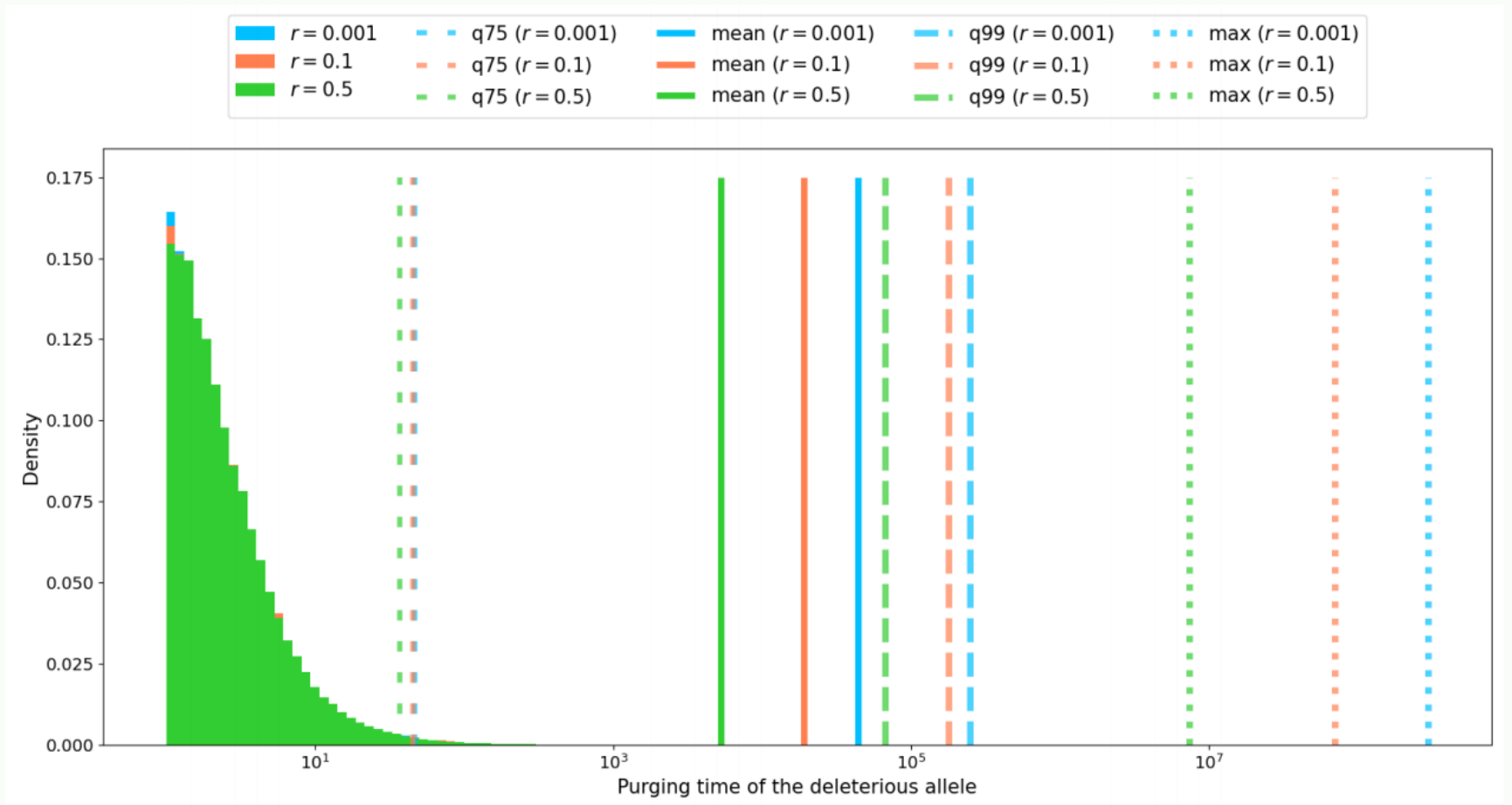


Abu Awad and Billiard (2017)

2. The evolution of close genomic region and sexual chrosomomes



Hartmann et al. 2020



Tezenas et al. (Biorxiv)

3.The rise of SI on its own

Issues in <*self-incompatibility*> **research** **(adapted from C. Mariscal 2021)**

Some steps in the process could have been *chancy*

others could have been deterministic but highly *contingent*

still others could have been *the only way*

What can we know?