Establishment of local adaptation in partly self-fertilizing populations

Bogi Trickovic[1,*] and Sylvain Glémin[2,3]

[1]Center for Mechanisms of Evolution, Arizona State University
[2]University of Rennes, CNRS, ECOBIO [(Ecosystèmes, biodiversité, évolution)] - UMR 6553, Rennes, France
[3]Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden
[*]Corresponding author: btrickov@asu.edu

Running head: local adaptation with selfing

Keywords: branching process, local adaptation, migration-selection balance, protected polymorphism, selfing

September 23, 2021

Abstract

Populations often inhabit multiple ecological patches and thus experience divergent selection, which can lead to local adaptation if migration is not strong enough to swamp locally adapted alleles. Conditions for the establishment of a locally advantageous allele have been studied in randomly mating populations. However, many species reproduce, at least partially, through self-fertilization, and how selfing affects local adaptation remains unclear and debated. Using a two-patch branching process formalism, we obtained a closed-form approximation under weak selection for the probability of establishment of a locally advantageous allele (P) for arbitrary selfing rate and dominance level, where selection is allowed to act on viability or fecundity, and migration can occur via seed or pollen dispersal. This solution is compared to diffusion approximation and used to investigate the consequences of a shift in a mating system on P, and the establishment of protected polymorphism. We find that selfing can either increase or decrease P, depending on the patterns of dominance in the two patches, and has conflicting effects on local adaptation. Globally, selfing favors local adaptation when locally advantageous alleles are (partially) recessive, when selection between patches is asymmetrical and when migration occurs through pollen rather than seed dispersal. These results establish a rigorous theoretical background to study heterogeneous selection and local adaptation in partially selfing species.

1 Introduction

Most advantageous alleles are lost from populations (Haldane, 1927), but those that escape extinction are ultimately destined to fix, providing that the fitness effect of an allele remains positive over space and time. However, fitness effects can vary across species' range, so that an allele that is advantageous in one ecological patch might be deleterious in another. Two mutually exclusive outcomes are then possible: Locally advantageous allele is maintained by divergent selection as polymorphism, or the allele with the best overall performance across the patches becomes fixed. Since ranges of essentially all species are spread across heterogeneous environments, it is of some interest to understand the conditions under which these outcomes are materialized. Although previous works investigated these conditions for randomly mating populations in both a deterministic (Bulmer, 1972, Felsenstein, 1976, Levene, 1953, Maynard Smith, 1970) and a stochastic setting (Tomasini and Peischl, 2018, Yeaman and Otto, 2011), conditions for maintenance of local adaptation in partially self-fertilizing populations remain unexplored. Partial selfing is widespread among eukaryotes and well-described in plants (Igic and Kohn, 2006), animals (Jarne and Auld, 2006), algae (Hanschen *et al.*, 2018), and fungi (Billiard *et al.*, 2012), and considering its effect on spatially heterogeneous selection is especially relevant for at least two reasons. First, partial selfing is more frequently found among sessile or less mobile organisms, for which landscape heterogeneity directly translates into local selective pressures because of the absence of the homogenizing effect of movements across the environment. Second, partial selfing likely influences the conditions for local adaptation by altering the effective population size and effective migration rate, hence the genetic structure of a population, but also by reducing the role of heterozygotes' fitness in the dynamics of allelic frequencies.

In a single population, selfing exerts two opposing effects on the fixation probability of an advantageous allele (Pollak (1987), Pollak and Sabran (1992),Charlesworth (1992), Caballero and Hill (1992)). Firstly, it increases the rate at which a rare advantageous allele spreads through the population. As the selfing rate increases, homozygotes appear more quickly so that the spread of an advantageous allele becomes increasingly decoupled from the fitness of the heterozygote. In other words, selfing increases the effective dominance coefficient: the rate of spread in a partially selfing population with dominance coefficient h, would be equal to the rate of spread in a randomly mating population with dominance coefficient $\tilde{h} = F + (1 - F)h$ (where F is the equilibrium population inbreeding coefficient). Secondly, since alleles making up the progeny are not independently sampled, selfing also reduces the efficacy of selection by reducing the effective population size, $\tilde{N}$, by a factor of $1 + F$ (Gale

(1990)). These two opposing effects cancel out when an advantageous allele is codominant ($h = 0.5$), implying that the fixation probability of a codominant advantageous allele is invariant across populations with different selfing rates. If an advantageous allele is partially recessive ($h < 0.5$), then the effect of increased efficacy of selection through increased effective dominance outweighs the reduction in $\tilde{N}$ yielding an increase in the establishment probability. Conversely, partially dominant advantageous alleles ($h > 0.5$) are less likely to fix because the reduced efficiency due to lower effective population size trumps increased efficacy of selection due to increased homozygosity. These predictions are altered for male and fecundity selection. Under selfing, no selection occurs on the male function, but the strength of selection on the female function is doubled because two gametes are transmitted for one selected parent (Damgaard, 2000).

The evolutionary dynamic becomes further elaborated in structured populations with divergent selection. The simplest scenario involves two patches with an invading allele being advantageous in one but deleterious in the other patch. The dynamic of local adaptation then further depends on the effect that selfing has on gene flow. Partial selfing reduces the effective haploid migration rate due to the reduction in the number of female gametes that can be sired by immigrating male gametes, for individuals in the selfing population will have already fertilized themselves to a large extent. This reduction in gene flow has been proposed to prevent the spread of maladaptive alleles from neighboring populations and thus to promote local adaptation (Linhart and Grant, 1996), but two meta-analyses of reciprocal transplant experiments reported the absence of correlation between local adaptation and the mating system (Hereford, 2010, Leimu and Fischer, 2008). However, it is not immediately clear how the interplay between the mating system, migration, and selection modes affects the propensity for local adaptation. When local reproduction is panmictic, maintenance of local polymorphism roughly requires that local selection is stronger than migration. On the one hand, selfing reduces effective migration. On the other hand, it may either increase or decrease the efficacy of local selection, depending on the balance between the effect of increasing effective dominance and increasing genetic drift. In addition, in diploids, conditions where alleles are (partially) dominant when locally beneficial and (partially) recessive when locally deleterious are especially favorable to the maintenance of polymorphism (Yeaman and Otto, 2011). By unmasking recessive alleles, selfing is expected to reduce the range of applicability of those conditions. Overall, we still lack correct theoretical predictions, which makes it difficult to interpret the empirical results such as in Leimu and Fischer (2008) and Hereford (2010).

In the present work, we extend the results of previous theoretical models of local adaptation to partially self-fertilizing populations. We first derive a closed-form approximation for the probability that a locally advantageous allele escapes extinction. We then use this result to examine how previously described effects of selfing affect (1) the probability that allele escapes extinction, (2) conditions required for the establishment of protected polymorphism, and (3) the dependence of these properties on the mode of selection and migration.

2 Model

We work with a hermaphroditic population that is divided into two patches connected by bidirectional migration. Migration can occur during both the haploid and the diploid phase. In the following, we will use the example of a plant life cycle, but results can readily be transposed to other life cycles. Thus, only male migration, pollen dispersal, occurs during the haploid phase, and diploid migration corresponds to seed dispersal. Fitness is controlled by a single biallelic locus, with allele A favored in the first patch and allele a favored in the other patch. These patches are referred to as favored and disfavored patch, respectively. The life cycle of the modeled organism is composed of the following sequential stages: (1) male and female meiosis leading to to the production of male gametophyte (pollen containing sperm cell) and female gametophyte (embryo sac, containing egg cell), (2) pollen dispersal, (3) mating (including selfing) and seed formation, (4) seed dispersal, and (5) seed establishment and development, which yields adults of the next generation. To make the model general, we allow selection to operate on viability and on male and female fecundity. More precisely, adults in patch i with genotype k (AA, Aa, or aa) produce $W_{i,k}^{\♀}$ female and $W_{i,k}^{\♂}$ male gametophytes. We do not consider haploid selection, but this could readily be included in the model setting. Finally, each offspring with genotype k survives to maturity with probability $V_{i,k}$. The population selfing rate, S, is defined as the proportion of individuals in the population that self-fertilize and is equal across patches. Following previous works (ex: Hartfield and Glémin, 2016, Hössjer and Tyvand 2016) we assume that the population reaches equilibrium for inbreeding coefficient F and genotypic composition on a much shorter time-scale than the change in allelic frequencies (separation of time scale argument). Within each patch, we can thus write genotype frequencies directly as a function of allelic frequencies and F. If selection is not too strong, we can also use the neutral expectation: $F = S/(2 - S)$. This assumption is relaxed in simulations. Mathematical symbols and their meaning are outlined in Table 1.

2.1 Deriving the difference equation for allele frequency

Using the preceding definitions, one can derive the difference equations for allele frequencies in each patch. Let X_i and Y_i be the frequencies of genotypes AA and Aa, and x_i be the frequency of allele A in i^{th} patch. Our system has two degrees of freedom, given that there are three genotype frequencies and they must add up to unity. Thus, we keep track of frequencies of genotypes AA and Aa, while the frequency of aa genotype is by definition $1 - X_i - Y_i$. Adult genotype frequencies in generation t in i^{th} patch are:

$$X_i = x_i^2(t) + F x_i(t)(1 - x_i(t)) \tag{1a}$$

$$Y_i = 2x_i(t)(1 - x_i(t))(1 - F) \tag{1b}$$

Adult genotypes may produce different number of male and female gametophytes. Let $W_{i,k}^{♂}$ and $W_{i,k}^{♀}$ be the number of male and female gametophytes that an adult with genotype k (AA, Aa and aa) produces in patch i. Thus, the frequency of female and male gametophyte A in i^{th} patch are, respectively:

$$x_i^{♀} = \left(W_{i,AA}^{♀} X_i + W_{i,Aa}^{♀} \frac{Y_i}{2} \right) \Big/ \overline{W}_i^{♀} \tag{2a}$$

$$x_i^{♂} = \left(W_{i,AA}^{♂} X_i + W_{i,Aa}^{♂} \frac{Y_i}{2} \right) \Big/ \overline{W}_i^{♂} \tag{2b}$$

where $\overline{W}_i^{♀} = X_i W_{i,AA}^{♀} + Y_i W_{i,Aa}^{♀} + (1 - X_i - Y_i) W_{i,aa}^{♀}$ and $\overline{W}_i^{♂} = X_i W_{i,AA}^{♂} + Y_i W_{i,Aa}^{♂} + (1 - X_i - Y_i) W_{i,aa}^{♂}$. After pollen migration, the frequency of allele A in pollen in i^{th} patch is:

$$g_i^{♂} = (1 - m_{ij}) x_i^{♂} + m_{ij} x_j^{♂} \tag{3}$$

where m_{ij} is the fraction of pollen in i^{th} patch that come from j^{th} patch. Pollen dispersal is followed by mating, which yields offspring genotypes:

$$X_i' = (1 - S) x_i^{♀} g_i^{♂} + S \left(\frac{W_{i,AA}^{♀}}{\overline{W}_i^{♀}} X_i + \frac{Y_i}{4} \frac{W_{i,Aa}^{♀}}{\overline{W}_i^{♀}} \right) \tag{4a}$$

$$Y_i' = (1 - S) \Big(x_i^{♀} (1 - g_i^{♂}) + (1 - x_i^{♀}) g_i^{♂} \Big) + S \frac{W_{i,Aa}^{♀}}{\overline{W}_i^{♀}} \frac{Y_i}{2} \tag{4b}$$

Seed migration changes the genotype frequency to:

$$X''_i = (1 - M_{ij})X'_i + M_{ij}X'_j \tag{5a}$$

$$Y''_i = (1 - M_{ij})Y'_i + M_{ij}Y'_j \tag{5b}$$

where M_{ij} is the fraction of seed in patch i originating from patch j. Genotype j in patch i has probability $V_{i,j}$ to survive to maturity, so the frequency of allele A in the generation $t+1$ is:

$$x_i(t+1) = \left(V_{i,AA}X''_i + V_{i,Aa}\frac{Y''_i}{2} \right) \Big/ \overline{V}_i \tag{6}$$

where $\overline{V}_i = X''_iV_{i,AA} + Y''_iV_{i,Aa} + (1 - X''_i - Y''_i)V_{i,aa}$.

Equation (6) is our difference equation that expresses the frequency of allele A in the next generation in terms of its frequency in the current generation.

2.2 Analysis of the deterministic model

We first analyze the deterministic model that gives the conditions for a protected polymorphism. This extends results of Bulmer (1972) to partial selfing and more general forms of selection. Given that investigation of the consequences of an arbitrary selection, regime is prohibitively complicated, the proceeding analysis is restricted to a few special cases. We are ultimately interested in three selection scenarios: (1) when selection operates on differential survival to maturity and when selection operates on (2) female and (3) male fecundity. For each of these scenarios, we first consider seed migration only, as in Bulmer (1972) initial model (also as in Yeaman and Otto, 2011). In addition to the comparison with previous models, it allows analyzing the effect of selfing on selection only, as selfing does not affect seed migration. Then, we consider pollen migration thus analyzing the joint effect of selfing on both selection and migration. In total, we examine six categories of scenarios.

We use the standard approach to obtain conditions for stable polymorphism by considering the conditions for which both monomorphic equilibria are unstable. Let both patches be fixed for allele a and allele A acts as an invader. Equivalent results can be obtained by considering patches fixed for allele A and a acting as the invader. The logic of derivation is the same in all cases, wherein we linearize equations (6) around $\vec{x} = [0, 0]^T$ to obtain the system of linear difference equations, where $\mathbf{J}$ is the associated Jacobian. While $\mathbf{J}$ captures the invasion

dynamics for an arbitrary mode of selection and migration, it is difficult to analyze, so we focus on special population genetic cases outlined above. Each of these scenarios will have an associated Jacobian $\mathbf{J}^{\{k\}}$:

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \mathbf{J}^{\{k\}} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad (J_{ij}^{\{k\}}) = \begin{cases} k = \circ, & M_{ij}\Delta V_{0,i} \\ k = ♀, & M_{ij}\Delta W_{0,j}^{♀} \\ k = ♂, & M_{ij}\Delta W_{0,j}^{♂} \end{cases} \tag{7}$$

Terms $\Delta V_{0,i}$, $\Delta W_{0,i}^{♀}$, and $\Delta W_{0,i}^{♂}$ are the intergenerational change in frequency of A in i^{th} patch due to selection on a particular fitness component, and M_{ij} accounts for migration between patches, with $M_{i,i} = 1 - M_{i,j}$, with $j \neq i$. In the case of viability selection, the frequency change of A in the patch 1 reflects either the contribution of alleles A that stay in patch 1 $(1 - M_{12})$ and survives to adulthood $(\Delta V_{0,1})$, or the contribution of alleles A that migrate from patch 2 (M_{12}) and then survive to adulthood $(\Delta V_{0,1})$. Similar logic holds for female and male fecundity selection but off-diagonal selection terms are exchanged because selection occurs prior to migration: note the j indices in ΔW terms instead of i in the ΔV term in equation (7). Selection terms for the three scenarios are obtained by keeping constant the two fitness components not involved in the scenarios in the general recursion equations. Thus, under seed migration the three selection terms are:

$$\Delta V_{0,i} = \frac{FV_{i,AA} + (1-F)V_{i,Aa}}{V_{i,aa}} \tag{8a}$$

$$\Delta W_{0,i}^{♀} = \frac{(1-F)(W_{i,aa}^{♀} + (1+F)W_{i,Aa}^{♀}) + F(3+F)W_{i,AA}^{♀}}{2(1+F)W_{i,aa}^{♀}} \tag{8b}$$

$$\Delta W_{0,i}^{♂} = \frac{(1-F)((1-F)W_{i,Aa}^{♀} + FW_{i,AA}^{♂}) + (1+3F)W_{i,aa}^{♂}}{2(1+F)W_{i,aa}^{♂}} \tag{8c}$$

These expressions can be made more intuitive by parameterizing fitnesses relative to genotype aa, where s_i°, $s_i^{♀}$, $s_i^{♂}$ are the reproductive advantage of AA homozygote relative to aa homozygote when selection acts on viability, female, and male fecundity, respectively. Parameters $\tilde{h}_i^{\circ}$, $\tilde{h}_i^{♀}$, $\tilde{h}_i^{♂}$ represent the effective dominance of the heterozygote in i^{th} patch for the three respective selection modes. These are composite parameters that depend on F and the actual dominance and are introduced to capture the fact that selfing increases the effective dominance of invading heterozygote due to contribution of mutant homozygotes to the invasion process. For the three cases of viability, female, and male fecundity effective

dominances are respectively:

$$\tilde{h}_i^{\circ} = (1-F)h_i^{\circ} + F \tag{9a}$$

$$\tilde{h}_i^{\text{♀}} = \frac{(1+3F)}{4(1+F)}(F + (1-F)h_i^{\text{♀}}) = \frac{1}{2}(1+S)(F + (1-F)h_i^{\text{♀}}) \tag{9b}$$

$$\tilde{h}_i^{\text{♂}} = \frac{(1-F)}{2(1+F)}(F + (1-F)h_i^{\text{♂}}) = \frac{1}{2}(1-S)(F + (1-F)h_i^{\text{♂}}) \tag{9c}$$

These forms capture the fact that selfing increases the effective dominance of invading mutants: the invader has dominance h_i when it invades in a heterozygous form (that is, $1-F$ fraction of time), and dominance of one when it invades in a homozygous form (that is, F fraction of time). Selfing also affects the global selection intensity on male and female fecundity. The 1/2 factor captures the fact that selection acts only on half of the gametes produced by each genotype, the $1-S$ factor that selection on male fecundity only operates under outcrossing and the $1+S$ factor that under selfing an individual contribute two alleles through seed production (female fitness). Taken together:

$$\Delta V_{0,i} = 1 + \tilde{h}_i^{\circ} s_i^{\circ} \tag{10a}$$

$$\Delta W_{0,i}^{\text{♀}} = 1 + \tilde{h}_i^{\text{♀}} s_i^{\text{♀}} \tag{10b}$$

$$\Delta W_{0,i}^{\text{♂}} = 1 + \tilde{h}_i^{\text{♂}} s_i^{\text{♂}} \tag{10c}$$

The fitness of the mutant always take the same form as in a haploid model with an effective selective advantage $\tilde{s} = \tilde{h}s$.

The preceding approach can be readily extended to the selective scenarios where migration occurs solely via pollen. One can still use the Jacobian in (7), albeit with the modified migration rate to account for the facts that (a) pollen is haploid and thus do not contribute in the same way as diploid seed to the gene pool of the next generation, and (b) that in all three cases migration operates prior to selection rather than after:

$$(J_{ij}^{\{k\}}) = \begin{cases} k = \circ, & \tilde{M}_{ij}^{\circ} \Delta V_{0,i} \\ k = \text{♀}, & \tilde{M}_{ij}^{\text{♀}} \Delta W_{0,i}^{\text{♀}} \\ k = \text{♂}, & \tilde{M}_{ij}^{\text{♂}} \Delta W_{0,i}^{\text{♂}} \end{cases} \tag{11}$$

Selection terms are still parameterized according to equations (10a–10c), but effective mi-

gration is given by:

$$\tilde{M}_{ij}^{\circ} = (1-S)\frac{m_{ij}}{2}\frac{1+h_i^{\circ}s_i^{\circ}}{1+\tilde{h}_i^{\circ}s_i^{\circ}} \tag{12a}$$

$$\tilde{M}_{ij}^{\text{♀}} = (1-S)\frac{m_{ij}}{2}\frac{1}{1+\tilde{h}_i^{\text{♀}}s_i^{\text{♀}}} \tag{12b}$$

$$\tilde{M}_{ij}^{\text{♂}} = (1-S)\frac{m_{ij}}{2}\frac{1+\frac{\tilde{h}_i^{\text{♂}}s_i^{\text{♂}}}{(1-S)/2}}{1+(1-m_{ij})\tilde{h}_j^{\text{♂}}s_j^{\text{♂}}+m_{ij}\tilde{h}_j^{\text{♂}}s_j^{\text{♂}}} \tag{12c}$$

Viability effective migration rate (12a) can be intuitively understood by noting that $1-S$ corresponds to the fraction of pollen that is exchanged between the patches, $m_{ij}/2$ corrects for the fact that haploid phase contributes two-fold less alleles to the gene pool of the next generation relative to the diploid progeny dispersal, and the relative fitness ratio $(1+h_i^{\circ}s_i^{\circ})/(1+\tilde{h}_i^{\circ}s_i^{\circ})$ accounts for the fold increase in invader frequency in the patch of origin due to the inhibitory effect of selfing on gene flow. Similar interpretation holds for (12b) and male fecundity selection (12c), except that fitness ratio now corrects for selection acting prior to migration. Note that for weak selection, the three effective migration rates reduce to $(1-S)\frac{m_{ij}}{2}$. The only complication is that when selection acts on the male fitness component, one has to modify the effective dominance along with the effective migration rate. More precisely, $\Delta W_{0,i}^{\text{♂}} = 1+\omega\tilde{h}_i^{\text{♂}}s_i^{\text{♂}}$, where ω:

$$\omega = 1 - m_{ij}\left(1-\frac{\tilde{h}_j^{\text{♂}}s_j^{\text{♂}}}{\tilde{h}_i^{\text{♂}}s_i^{\text{♂}}}\right) \tag{13}$$

Overall, a simple parameterization procedure allows the exploration of a wide breadth of population genetic scenarios.

2.3 Branching process approximation

By analyzing the stability of system 7, one can only examine conditions necessary for the invading allele to escape extinction. However, we also want to know the probability of the escape. We pursue this by representing the change in the number of invading allele as a multi-type branching process. Selection is thus assumed to be stronger than drift (i.e., $|N\tilde{h}_i^{\{k\}}s_i^{\{k\}}| \gg 1$). We work with two types, each tracking the number of mutant alleles in one of the two patches. The number of the mutant alleles is represented by the vector $\vec{N}(t) = [N_1(t), N_2(t)]^T$, where $N_1(t)$ and $N_2(t)$ are the number of mutants in patch 1 and 2 in generation t, respectively. Let f_i be the probability generating function (PGF) of offspring

number of mutant alleles of type i. The ultimate probability of extinction of type i $(1-P_i)$ is given by the smallest positive root of the following system of equations (Haccou *et al.* (2005)):

$$1 - P_1 = f_1(1 - P_1, 1 - P_2) \tag{14a}$$

$$1 - P_2 = f_2(1 - P_1, 1 - P_2) \tag{14b}$$

As we can assume that the number of offspring that types i and j leave are uncorrelated, then $f_k(1-P_1, 1-P_2) = f_{k,1}(1-P_1)f_{k,2}(1-P_2)$, with $k = \{1, 2\}$. Intuitively, the probability that mutant goes extinct conditioning on appearance in patch 1 is the probability that all of its offspring left in patch 1 ultimately go extinct $(f_{1,1}(1-P_1))$, and the probability that all offspring left in patch 2 via migration $(f_{1,2}(1-P_2))$ are ultimately lost. A similar explanation holds when mutant originates in patch 2. To solve the system, one needs to obtain the $f_{i,j}(z)$ PGFs.

In the Wright-Fisher model with random mating, the number of mutant offspring is approximately Poisson-distributed with mean $\Delta\mathcal{W}_{0,i}$, where $\mathcal{W}$ stands for the different form of selection as described above. Partial self-fertilization affects the mutant offspring distribution in two ways. First, it increases the mean number of offspring by inflating the effective dominance coefficient (see above). Second, selfing increases the variance in offspring number by the factor $1 + F$ relative to the randomly mating population (Caballero and Hill, 1992, Pollak, 1987). Therefore, offspring number should follow a distribution with mean $\Delta\mathcal{W}_{0,i}$ and variance $\Delta\mathcal{W}_{0,i}(1+F)$. To our knowledge, the full distribution has never been obtained, even in single populations, probably because only the two first moments are needed for branching process and diffusion approximations. In appendix, we derive the exact probability generating function (PGF) of the distribution in different cases. In a single population we have:

$$\begin{aligned} f_\phi(z) = (1-F)\, & \underbrace{e^{-(1-z)\left(\overbrace{(1-S)\frac{W^{\text{♀}}_{Aa}+W^{\text{♂}}_{Aa}}{2}V_{Aa}}^{\text{Through outcrossing}} + \overbrace{\frac{S}{2}W^{\text{♀}}_{Aa}V_{Aa}}^{\text{Through selfing}}\right)}}_{\text{Transmitted in } Aa} \underbrace{e^{-\frac{1}{2}(1-z^2)\frac{S}{2}W^{\text{♀}}_{Aa}V_{AA}}}_{\text{Transmitted in } AA \text{ (only selfing)}} \\ & + F\, \underbrace{e^{-(1-z)\left((1-S)\frac{W^{\text{♀}}_{AA}+W^{\text{♂}}_{AA}}{2}V_{Aa}\right)}}_{\text{Transmitted in } Aa \text{ (only outcrossing)}} \underbrace{e^{-\frac{1}{2}(1-z^2)SW^{\text{♀}}_{AA}V_{AA}}}_{\text{Transmitted in } AA \text{ (only selfing)}} \end{aligned} \tag{15}$$

The first term in equation (15) corresponds to a A parental allele coming from an Aa individ-

ual (probability $1-F$) and the second term to an AA individual (probability F). Offspring can be transmitted either in AA offspring, which occurs only through selfing (because we consider a rare mutant) or in Aa offspring, which can occur either through outcrossing or through selfing when the parent is Aa (so with a probability $S/2$). Note that fecundity selection depends on parent genotypes whereas viability selection depends on offspring genotype. Using the properties of PGFs we easily retrieve that the mean of the distribution is $\Delta\mathcal{W}_i$ obtained in the deterministic analysis and the variance is $\Delta\mathcal{W}_i(1+F)$. For $F=0$, equation (15) boils down to the PGF of a Poisson distribution. Hereafter, as for the deterministic analysis, we consider the different forms of selection separately by setting constant the other fitness components.

For the two-patch model we need to include migration in the $f_{i,j}(z)$ PGFs. For seed migration, we show that:

$$\begin{aligned} f_{i,j}(z) = (1-F)e^{-(1-z)\left((1-S)\frac{W^{\text{♀}}_{Aa,i}+W^{\text{♂}}_{Aa,i}}{2}M_{i,j}V_{Aa,j}+\frac{S}{2}W^{\text{♀}}_{Aa,i}M_{i,j}V_{Aa,j}\right)}e^{-\frac{1}{2}(1-z^2)\frac{S}{2}W^{\text{♀}}_{Aa,i}M_{i,j}V_{AA,j}} \\ + Fe^{-(1-z)\left((1-S)\frac{W^{\text{♀}}_{AA,i}+W^{\text{♂}}_{AA,i}}{2}M_{i,j}V_{Aa,j}\right)}e^{-\frac{1}{2}(1-z^2)SW^{\text{♀}}_{AA,i}M_{i,j}V_{AA,j}} \end{aligned} \tag{16}$$

recalling that $M_{i,i} = 1 - M_{i,j}$ with $j \neq i$. Note that fecundity selection occurs in patch i, before seed migration, whereas viability selection occurs in patch j, after seed migration. However, the different orderings of migration and selection yield the same results (see *Mathematica* notebook). We also retrieve that the mean of the four distributions are given by the Jacobian terms obtained above, $M_{i,j}\Delta\mathcal{W}_{i,j}$ and the variance is also inflated by $1+F$. For pollen migration the form is slightly different because migrants can only contribute to the next generation through outcrossing and male gametes (whereas selfed seeds can migrate). The form of the PGF is thus different for the resident and migrant contribution. We have:

$$\begin{aligned} f_{i,i}(z) = (1-F)e^{-(1-z)\left((1-S)\frac{W^{\text{♀}}_{Aa,i}+W^{\text{♂}}_{Aa,i}(1-m_{i,j})}{2}V_{Aa,i}+\frac{S}{2}W^{\text{♀}}_{Aa,i}V_{Aa,i}\right)}e^{-\frac{1}{2}(1-z^2)\frac{S}{2}W^{\text{♀}}_{Aa,i}V_{AA,i}} \\ + Fe^{-(1-z)\left((1-S)\frac{W^{\text{♀}}_{AA,i}+W^{\text{♂}}_{AA,i}(1-m_{i,j})}{2}V_{Aa,i}\right)}e^{-\frac{1}{2}(1-z^2)SW^{\text{♀}}_{AA,i}V_{AA,i}} \end{aligned} \tag{17a}$$

$$f_{i,j}(z) = (1-F)e^{-(1-z)(1-S)\frac{W^{\text{♂}}_{Aa,i}m_{i,j}}{2}V_{Aa,j}} + Fe^{-(1-z)(1-S)\frac{W^{\text{♂}}_{AA,i}m_{i,j}}{2}V_{Aa,j}} \quad \text{for } j \neq i \tag{17b}$$

As expected, the mean of the distributions are still given by the Jacobian terms with the

appropriate effective migration rates, $\tilde{M}_{i,j}\Delta\mathcal{W}_{i,j}$ (see above). However, the variances are not inflated by $1+F$. For migrant alleles, equation (17b) boils down to a Poisson distribution because the contribution is only through outcrossing. In contrast for resident allele, the variance is inflated by more than $1+F$ (see appendix for details). The reason is that for a given selfing rate, only alleles produced by outcrossing can be exported to the other patch, which increase the proportion of resident offspring contributed by selfing, hence the variance. Note, however, that the difference in PGFs between the two migration modes only affects high-order terms of selection and migration so has no effect on branching process approximation results, which can all be expressed with the same form using appropriate effective parameters as defined in previous section (see *Mathematica* notebook).

Previous section also showed that $\Delta\mathcal{W}_{0,i}$ always has the form of $1+\tilde{s}_i$, where $\tilde{s}_i$ is the effective advantage of an invading mutant in patch i, while M_{ij} represents the effective migration rate between patches i and j. Thus, one can derive P_i for this generic case and then parameterize migration and selection to reflect the specific population genetic case. Since we cannot obtain the exact solution for extinction probabilities, we approximate to weak selection using the approach of Tomasini and Peischl (2018), which is based on the approximation for slightly supercritical multitype branching processes (Haccou *et al.*, 2005). All model parameters are rescaled by $\tilde{s}_1$ such that $M_{ij} := \chi_{ij}\tilde{s}_1$ and $\tilde{s}_2 := \zeta\tilde{s}_1$. By linearizing around $\tilde{s}_1 = 0$ and neglecting higher-order terms, we obtain a closed-form solution for the probability of establishment conditioning on the patch in which locally advantageous allele appeared. All analytical solutions are compared to simulated data using the method described in Appendix D. Simulations were conducted with the help of GNU Parallel (Tange, 2011).

2.4 Diffusion process approximation

An alternative way of incorporating stochastic transmission of alleles across generations is by approximating frequency change to diffusion. This approach has been originally developed by Sakamoto and Innan (2019) in the case of wholly outcrossing population. Here, we recapitulate their derivation and extend it to the case of partial selfing. The establishment probability satisfies the Kolmogorov backward equation (KBE):

$$0 = \sum_{i=1}^{2} \mu_i \frac{\partial P(x_1, x_2)}{\partial x_i} + \sum_{i=1}^{2} \frac{\sigma_i}{2} \frac{\partial^2 P(x_1, x_2)}{\partial x_i^2} \tag{18}$$

where μ_i and σ_i are drift and diffusion coefficients for patch i, and are given by:

$$\mu_i = \tilde{s}_i x_i(1-x_i) + \tilde{M}_{ij}(x_j - x_i), \quad \sigma_i = \frac{x_i(1-x_i)}{2N_1/(1+F)} \tag{19}$$

Note that $\tilde{s}$ corresponds to effective selective advantage of mutant, and $\tilde{M}_{ij}$ is the effective migration rate; These were derived in section 2.2. Diffusion coefficient is inflated by the factor of $1+F$ as described previously. Unfortunately, we could not find neither the explicit nor approximate solution to this partial differential equation. Following treatment in Barton (1987), the problem is simplified by focusing on the first phase of fixation when mutant allele is rare. Thus, one can approximate coefficients as:

$$\sigma_i = \frac{x_i}{2N_1/(1+F)} + o(x_i^2), \quad \mu_i = \tilde{s}_i x_i + \tilde{M}_{ij}(x_j - x_i) + o(x_i^2) \tag{20}$$

which leads to the simplified KBE:

$$\begin{aligned} 0 =& \Big(\tilde{s}_1 x_1 + \tilde{M}_{12}(x_2 - x_1)\Big)\frac{\partial P(x_1,x_2)}{\partial x_1} + \Big(\tilde{s}_2 x_2 + \tilde{M}_{21}(x_1 - x_2)\Big)\frac{\partial P(x_1,x_2)}{\partial x_2} + \\ & \frac{x_1}{4N_1/(1+F)}\frac{\partial^2 P(x_1,x_2)}{\partial x_1^2} + \frac{x_2}{4N_2/(1+F)}\frac{\partial^2 P(x_1,x_2)}{\partial x_2^2} \end{aligned} \tag{21}$$

Because the invading allele is rare, the mutants in patch i go extinct independently of one another with probability e^{-l_i}, which means that the establishment probability takes the form:

$$P(x_1,x_2) = 1 - e^{-2\left(\frac{N_1}{1+F}x_1 l_1 + \frac{N_2}{1+F}x_2 l_2\right)} \tag{22}$$

Substituting eqn. 22 into 21 results in an algebraic equation which captures the establishment probability for an arbitrary number of initial mutants across two patches. We are interested in two special cases, when mutant either starts in favored ($x_1 = 1/(2N_1)$ and $x_2 = 0$) or disfavored patch ($x_1 = 0$ and $x_2 = 1/(2N_2)$). With this parameterization, we obtain the system of two non-linear equations:

$$0 = 2\tilde{M}_{21}N_2 l_2 + 2N_1 l_1 \tilde{s}_1 - N_1 l_1(2\tilde{M}_{12} + l_1) \tag{23}$$

$$0 = 2\tilde{M}_{12}N_1 l_1 + 2N_2 l_2 \tilde{s}_2 - N_2 l_2(2\tilde{M}_{12} + l_2) \tag{24}$$

By solving these for l_1 and l_2, taking the real root, and substituting into eqn. 22, one retrieves a closed-form solution for the establishment probability (see *Mathematica* notebook). With all of the simplifying assumptions above, the diffusion approximation works in the same

regime when $N_e s \gg 1$, as is the case with branching approximation. As will be shown later in the text, the comparison of two solutions reveals that diffusion approximation performs identically when the allele starts in the favored patch and slightly better when it originates in the disfavored one. Because branching process approximation is easier to work with (e.g., series expansion readily yields intuitive special cases), we use this solution throughout our analysis and note that the diffusion approach yields marginally better precision.

3 Results

3.1 Probability of establishment of a mutant

3.1.1 General equations

The probability of establishment of a new allele is non-zero if the branching process is super-critical, which means that the leading eigenvalue of the mean reproductive matrix must be larger than unity (Haccou *et al.*, 2005). For the invasion of allele A under viability selection, it corresponds to:

$$\frac{M_{12}}{\frac{((1-F)V_{1,Aa}+FV_{1,AA})-V_{1,aa}}{(1-F)V_{1,Aa}+FV_{1,AA}}} + \frac{M_{21}}{\frac{((1-F)V_{2,Aa}+FV_{2,AA})-V_{2,aa}}{(1-F)V_{2,Aa}+FV_{2,AA}}} < 1 \tag{25}$$

The denominator is the relative reproductive advantage of invading mutant against the resident allele. The invading population is a combination of mutant heterozygotes (when the two gene copies of an individual are not identical by descent, $1-F$ fraction of time) and homozygotes (when the two gene copies are identical by descent, F fraction of time). So the whole term $((1-F)V_{i,Aa} + FV_{i,AA})$ gives the fitness of the invading mutant weighted by the contribution of invading genotypes. In an outcrossing population ($F = 0$), condition (25) reduces to $M_{12}/(1 - V_{1,aa}/V_{1,Aa}) + M_{21}/(1 - V_{2,aa}/V_{2,Aa}) < 1$, which is identical to Bulmer's inequality. The same inequality with re-parameterized selection holds for fecundity and male sexual selection (see Appendix C). The full equations for the establishment probabilities are too long and not very informative and are thus reported in the appendix (see equations (A16) and (A17)). Without loss of generality, consider that allele A is advantageous in patch 1 but deleterious in patch 2 ($s_1 > 0 > s_2$); in the case of symmetrical migration ($M_{12} = M_{21} = M/2$) and migration prior to selection the general probability of establishment

of allele A in either patch 1 or 2 is given by:

$$P_1 = \max\left[\frac{1}{1+F}\frac{2\sigma(\sigma+\Delta)(\tilde{h}_1 s_1 + \tilde{h}_2 s_2 + \sigma - M)}{2\Delta(\sigma+\Delta) + M^2 + M(\sigma-\Delta)}, 0\right] \tag{26}$$

$$P_2 = \max\left[\frac{1}{1+F}\frac{2M\sigma(\tilde{h}_1 s_1 + \tilde{h}_2 s_2 + \sigma - M)}{2\Delta(\sigma+\Delta) + M^2 + M(\sigma-\Delta)}, 0\right] \tag{27}$$

where $\Delta = \tilde{h}_1 s_1 - \tilde{h}_2 s_2$ and $\sigma = \sqrt{\Delta^2 + M^2}$ with $\tilde{h}$ as defined above for viability, female and male fecundity selection. These resemble – but are not equal – to equations (4) and (5) in (Tomasini and Peischl, 2018), who obtained an inexact result because of a typo in their application of Haccou et al.'s theorem. Initially the typo came from Aeschbacher and Bürger (2014) and was independently detected by Pontz and Bürger (2021) (see their Appendix A). Interestingly, the correct solutions are more complicated and less elegant than the ones of Tomasini and Peischl (2018).

3.1.2 Limiting cases and comparison with previous results

Despite the complexity of the general equations, useful insights can be obtained from simple limit conditions. In the limit of no migration ($M_{12} = M_{21} = 0$), we get back to the single-patch scenario where selection favors allele A in patch 1: $P_1 = 2\frac{\tilde{h}_1 s_1}{1+F}$. This is equivalent to Charlesworth (1992) and Caballero and Hill (1992) for viability selection. For male and female fecundity selection, the result is similar to the one intuited in Damgaard (2000), except that we also accounted for the reduction in P due to increased variance of offspring number. Conversely, in the limit of full migration we obtain $P_1 = P_2 = \frac{\tilde{h}_1 s_1 + \tilde{h}_2 s_2}{1+F}$, which is equivalent to Nagylaki (1980), with the appropriate rescaling for dominance and partial selfing. The probability of establishment is simply the average over the two patches. Interestingly, the fact that the effects of selfing on viability selection cancel out for codominant alleles ($h_1 = h_2 = 1/2$) is no longer true with intermediate migration, even for seed migration that is not affected by selfing. In the limit of weak symmetrical migration, we obtain $P_1 = s_1 - \frac{M}{1+F}$ and $P_2 = \frac{M s_1}{(1+F)(s_1 - s_2)}$ and in the limit of strong migration (i.e. $M >> s_i$): $P_1 = \frac{s_1+s_2}{2} + (1+F)\frac{s_1(s_1-s_2)}{4M}$ and $P_2 = \frac{s_1+s_2}{2} - (1+F)\frac{s_2(s_1-s_2)}{4M}$. Thus, unlike in a single population a locally advantageous allele that is codominant in both patches is more likely to escape extinction in selfers than in outcrossers. The relation between F and P_2 is more complicated, and the allele can either be more likely, for high migration rates, or less likely, for low migration rates, to escape extinction. The effects of selfing under codominance only vanish either for no migration or full migration.

3.1.3 Comparison with simulations

In general, our approximation for viability shows good correspondence with simulations across different selfing rates (Figure 1). This also holds true across different dominance coefficients, and when selection or migration are asymmetrical (Figure 2). Correspondence is better in cases where the invading allele initially appears in the favored rather than in the disfavored patch. Diffusion approximation gives identical results when allele starts in the favored, and slightly better results when in the case of the disfavored patch. The improvement only occurs when the migration rate is low, and this pattern persists across dominances and modes of selection. A possible reason for the slightly worst accuracy of the branching process approach in the disadvantageous patch with low migration is that the rescaling of the model assumes that selection and migration parameters are of the order of $\tilde{s}_1$. For low migration, we can have $m_{21} << \tilde{s}_1$, so that migration could not be sufficient to introduce the allele in the favored patch before it goes extinct in the disfavored one.

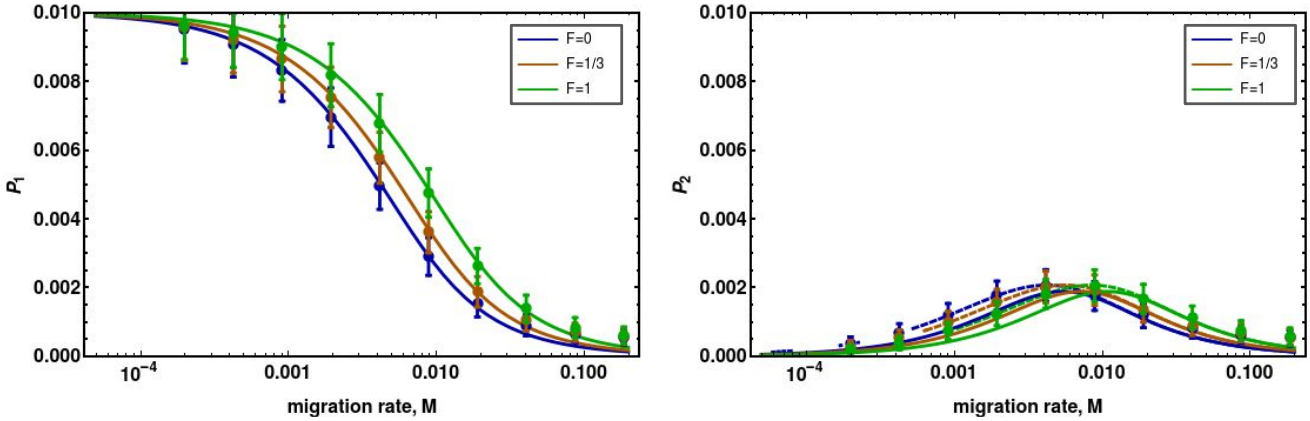


Figure 1: Comparison of analytical solution to simulated data across different selfing rates (corresponding to different equilibrium F). Dashed lines are diffusion approximation. Left panel: the probability of establishment when the allele emerges in the favored patch; Right panel: the probability of establishment when the allele emerges in the disfavored patch. Parameters: $h_1^\circ = h_2^\circ = 1/2$, $s_1^\circ = -s_2^\circ = 0.01$, $M_{12} = M_{21} = M$.

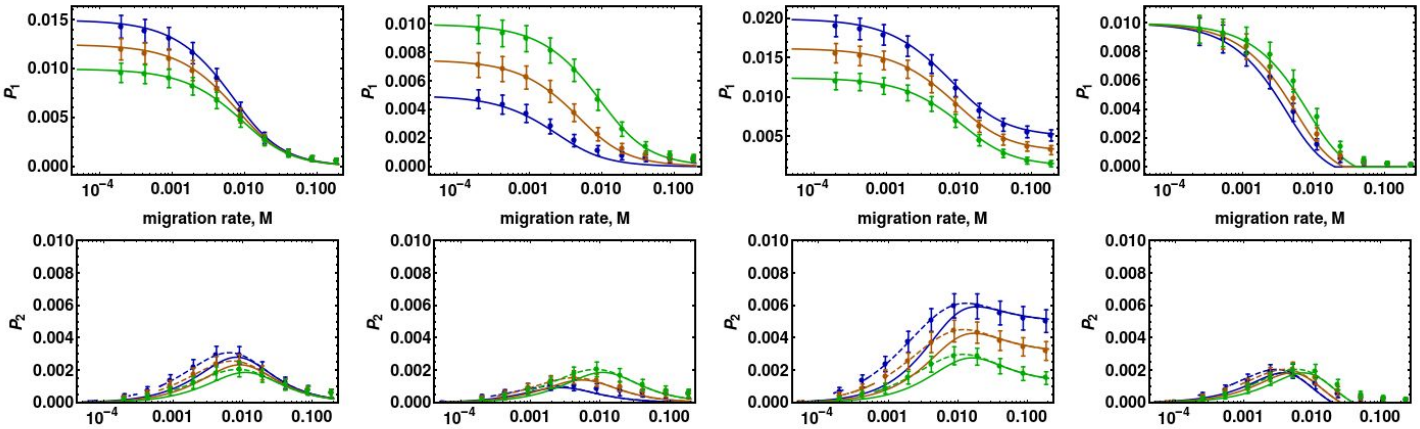


Figure 2: Simulations of different selection regimes. Going from left to right, the examined scenarios are as follows. Dashed lines denote diffusion approximation. First column: dominant case ($h_1^\circ = h_2^\circ = 3/4$, $s_1^\circ = -s_2^\circ = 0.01$, $M_{12} = M_{21} = M$); Second column: recessive case ($h_1^\circ = h_2^\circ = 1/4$, $s_1^\circ = -s_2^\circ = 0.01$, $M_{12} = M_{21} = M$); Third column: Asymmetric selection case ($s_1^\circ = 0.01, s_2^\circ = -0.0125$, $h_1^\circ = h_2^\circ = 1/2$, $M_{12} = M_{21} = M$); Fourth column: asymmetric migration ($s_1^\circ = -s_2^\circ = 0.01$, $h_1^\circ = h_2^\circ = 1/2$, $M_{12} = M$, and $M_{21} = 1.25M$).

P_1 is a monotonically decreasing function in M, because strong migration causes the spreading of the locally advantageous allele to become swamped by the resident from the disfavored patch. P_2 is non-monotonic function in M. Increasing migration initially increases the probability that an invading allele escapes from the disfavored patch before it goes extinct. However, at large migration rates, the spreading allele escapes the disfavored patch but is swamped due to a large influx of deleterious residents from the disfavored patch. For male fecundity selection, the analytical solutions for P_1 and P_2 are slightly worse against the simulations (Figure A11) than those provided for viability selection. Note, however that as the selfing rate tends towards 1, the effective selection coefficient on male function vanishes so that the strong selection condition for branching process approximation is not met anymore. Slight discrepancies are also observed when selection or migration are asymmetrical (Figures A12, and A13).

We also varied F continuously to obtain a fine-grained view of the analytic performance. Three insights emerge. First, there is a good agreement between simulated data and analytical solution, although deviations increase as the migration rate increases. Second, the branching process poorly captures the dynamics of the invasion when mutants are partially recessive. For example, in the limit of complete recesivity ($h = 0$), the mutant does not have any reproductive advantage (i.e., equations 10a–10c reduce to unity). We clearly see this in Figure 3, where the analytical solution underestimates P when the allele is partially recessive (first column). However, this discrepancy disappears as F increases, as homozy-

gotes increasingly contribute to the invading process. Third, diffusion approximation yields a better fit when mutant starts in the favored patch, but this improvement generally works only when M is low.

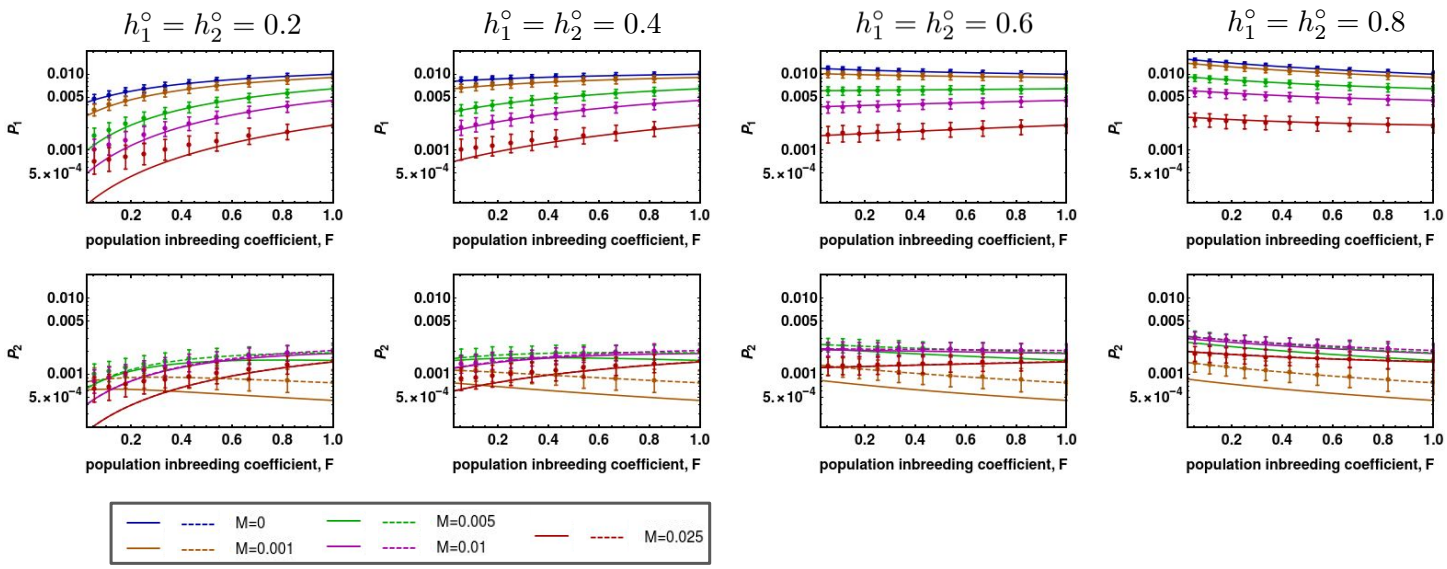


Figure 3: Simulations of different selection regimes. Going from left to right, the examined scenarios are as follows. Dashed lines denote diffusion approximation. Each column denotes the establishment probability for a set of dominance coefficients noted above the panel. Lines correspond to different migration rates (see legend box below). Other parameters: $s_1^\circ = -s_2^\circ = 0.01$, $M_{12} = M_{21} = M$). The axes are on logarithmic scale and error bars are excluded.

3.2 Decomposing the effects of selfing on the establishment probability

Selfing impedes the spread of an invading beneficial allele by reducing the effective population size and by increasing the effective dominance in the disfavored patch. On the other hand, selfing also facilitates the invasion process by increasing the effective dominance in the favored patch and reducing the gene flow when dispersal occurs through pollen. In this section, we ask: are the impeding effects generally more important than facilitating effects? When does the introduction of selfing increase P? To this end, we introduce an indicator $\beta_i(y)$ which denotes the effect that selfing has on P_i solely via parameter y. For example, $P_i^{\{\tilde{h}_1\}}$ corresponds to the establishment probability in patch i, if selfing only exerted its effect by

inflating favored dominance. More generally:

$$\beta_i(y) = \left(\frac{dP_i^{\{y\}}}{dF}\right)\bigg|_{F=0}, \qquad y \in \{\tilde{h}_i, \tilde{N}, \tilde{M}\}, \quad i \in \{1, 2\} \tag{28}$$

Consider decomposing the effects of selfing on the establishment of an allele affecting viability. For example, the indicator $\beta_1(\tilde{h}_1^\circ)$ tells us whether a shift to selfing increases P_1, if selfing solely acted through an increase in the effective dominance in the favored patch. This allows us to examine the effect of each factor separately and assess when one outweighs the other. The indicator is obtained by taking equation A16 and sequentially setting: (1) $F = 0$ which eliminates selfing's effect on $\tilde{N}$, then (2) parameterizing $\tilde{s}_1^\circ := s_1^\circ \tilde{h}_1^\circ$ which introduces the effect of selfing on favored dominance, and finally, (3) setting $\tilde{s}_2^\circ := s_2^\circ h_2^\circ$ which excludes the effect of selfing on the effective dominance in the disfavored patch. The outlined method is also applicable to other selection and migration scenarios, and the general procedure is given in Appendix F.

3.2.1 Emergence in the favored patch

We focus on the establishment conditioning on allele emerging in the favored patch. Consider viability selection first (see Figure 4). Given our interest in knowing whether selfing increases or decreases the establishment probability, we only focus on the region of parameter space where allele can become established in fully outcrossing population (right of the solid line in Figure 4). When the invading allele is partially recessive in favored and partially dominant in the disfavored patch (white region), the criterion for escaping extinction is not satisfied, and a shift to selfing can only have a promoting effect on the establishment in this region of parameter space.

Under seed dispersal, selfing only affects selection via effective dominance and the effective population size. One can recognize three regions of selfing effects (Figure 4). Firstly, when the invading mutant is partially recessive in the favored and partially dominant in the disfavored patch, selfing increases P_1 via the inflation of $\tilde{h}_1^\circ$ (blue region). More formally, this will be the outcome when $\beta_k(\tilde{h}_1^\circ) > -(\beta_k(\tilde{h}_2^\circ) + \beta_k(\tilde{N}))$ (dark blue region). This is because the mutant is already dominant in the disfavored patch so selfing does not significantly increase the rate of purging, but it does increase the rate of spread owing to the mutant being partially recessive in the favored patch. Thus, the net effect is increased P_1. Secondly, when the mutant is partially dominant in the favored and partially recessive in the disfavored patch, selfing decreases P_1 via inflation of $\tilde{h}_2^\circ$ (darker red region). This happens when

$\beta_k(\tilde{h}_1^\circ) < -\beta_k(\tilde{h}_1^\circ)$ (dark red region). The effect is opposite to the first regime: selfing has a larger effect on the increase in purging from the disfavored patch (since the dominance in the disfavored patch is low) than on the increase in spreading in the favored patch (because dominance in the favored patch is already high). The net effect is a decrease in P_1. The region of parameter space that lies between these two zones corresponds to cases when the facilitating effect of selfing (via dominance in the favored patch) roughly cancels out the impeding effect (via dominance in the disfavored patch) so that the net effect on P_1 is determined by the reduction in effective population size and the net effect of selfing is to decrease P_1.

Under pollen dispersal, we also found the same three regions but with different boundaries (Figure 4). In addition, selfing also affects pollen dispersal by reducing the effective number of migrants, so in the intermediate zone the effect of selfing on P_1 depends both on the reduction in effective population size and effective migration rate, which creates a fourth region – the region for which all parameter combinations where $\beta_k(\tilde{h}_1^\circ) + \beta_k(\tilde{M}) > -(\beta_k(\tilde{h}_2^\circ) + \beta_k(\tilde{N}))$ (light blue region). It corresponds to conditions where the reduction in effective migration rate overwhelms the impeding effects of selfing. Overall, selfing increases the establishment of new alleles under broader conditions under pollen than under seed dispersal. Note that selection coefficients are an order of magnitude lower in the right panel of Figure 4, and the region where selfing has a net-positive effect is comparable (if not larger) to the blue region in the left panel.

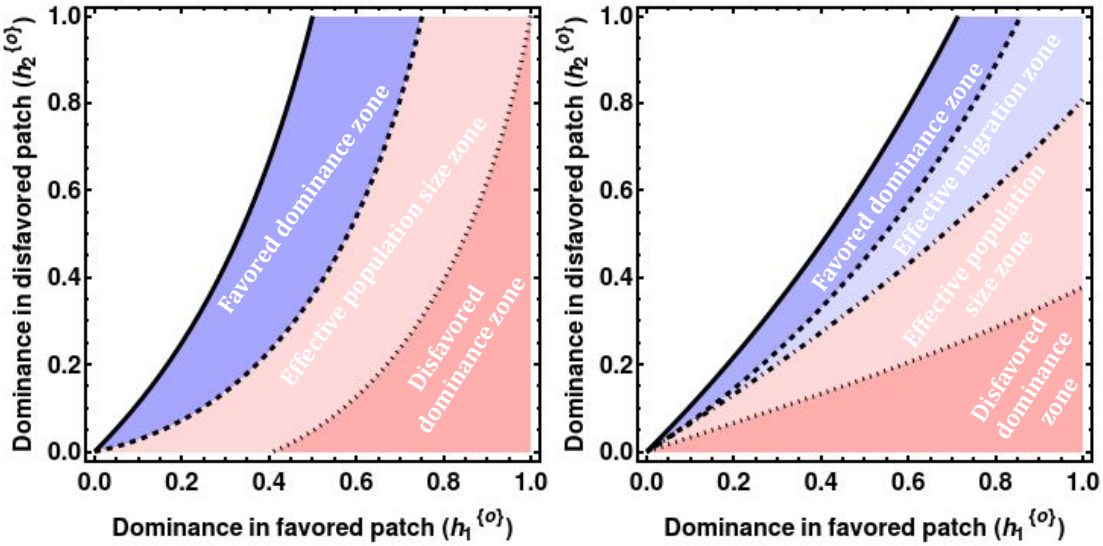


Figure 4: Four zones of selfing's effects on the establishment probability when mutant originates in the favored patch. White region denotes parameter space where the mutant cannot establish under outcrossing. Selection act only on viability, and migration occurs only via pollen (left), or only via seed (right). Blue and red regions correspond to parameter space where selfing facilitates and impedes P_1, respectively. Lines in the left panel are as follows. Solid black line: $P_1 = 0$; Dashed black line: $\beta_k(\tilde{h}_1^\circ) = -(\beta_k(\tilde{h}_2^\circ) + \beta_k(\tilde{N}))$; Dotted black line: $\beta_k(\tilde{h}_1^\circ) = -\beta_k(\tilde{h}_2^\circ)$. Lines in the right panel are as follows. Solid black line: $P_1 = 0$; Dashed black line: $\beta_k(\tilde{h}_1^\circ) = -\beta_k(\tilde{h}_2^\circ)$; Dashdotted line: $\beta_k(\tilde{h}_1^\circ) + \beta(\tilde{M}) = -(\beta_k(\tilde{h}_2^\circ) + \beta_k(\tilde{N}))$; Dotted line: $\beta_k(\tilde{h}_1^\circ) + \beta(\tilde{M}) = -\beta_k(\tilde{h}_2^\circ)$. Parameters for the left panel: $s_1^\circ = -s_2^\circ = 0.01$, $h_1^\circ = h_2^\circ = 1/2$, $M_{12} = M_{21} = 0.01$. Parameters for the right panel: $s_1^\circ = -s_2^\circ = 0.001$, $h_1^\circ = h_2^\circ = 1/2$, $M_{12} = M_{21} = 0.005$.

Although qualitatively similar results holds for two other modes of selection (Figure 5), one notices that the region where the selfing exerts its effect solely via effective dominance is significantly expanded. This is because an increase in the selfing rate diminishes the male component's relevance and increases the importance of the female component. The relative sizes of these regions vary based on parameter values, so more general quantitative statements are hard to make.

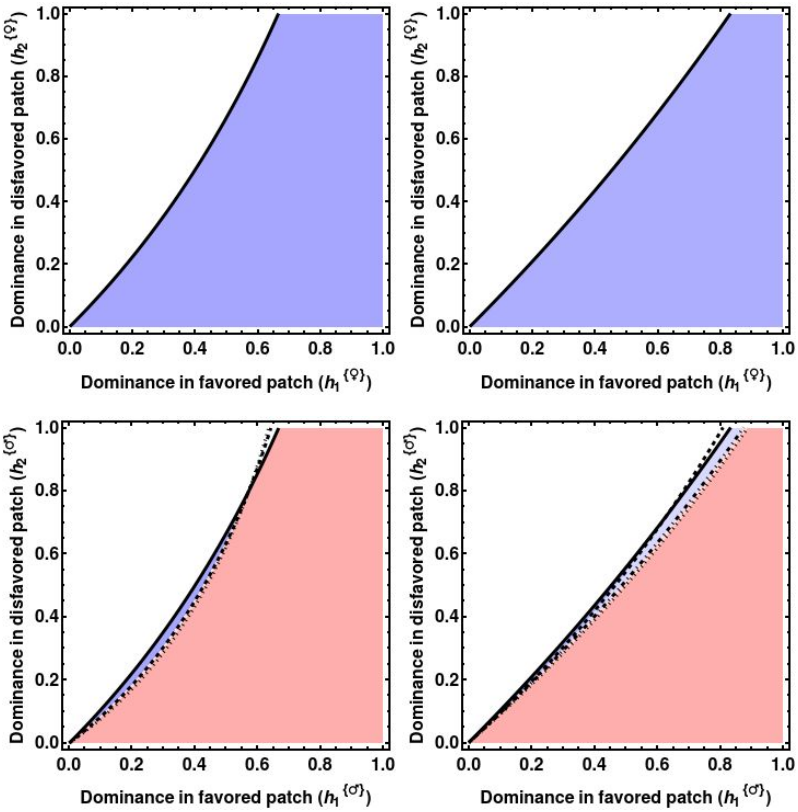


Figure 5: Zones of selfing's effect on P_1 under female fecundity (upper row), and on male fecundity (lower row). Color code as in Figure 4. Migration via seed is depicted in the left column, and via pollen in the right column. Black lines as in the previous figure. Depending on selection and migration scenario, parameters are either: $h_1^{♀} = h_2^{♀} = 1/2$ or $h_1^{♂} = h_1^{♂} = 1/2$, $s_1^{♀} = -s_2^{♀} = 0.01$ or $s_1^{♂} = -s_2^{♂} = 0.01$, and $M_{12} = M_{21} = 0.01$ or $m_{12} = m_{21} = 0.01$.

3.2.2 Emergence in the disfavored patch

Similar conclusions are reached if one considers emergence of mutant in the disfavored patch (Figure A16). However, under pollen dispersal, selfing reduces the effective migration rate, which can have either positive or negative effect on P_2 depending on the parameters of the model. This phenomenon arises due to P_2 being a non-monotonic function of m (see Figure 2). More specifically, if migration is much stronger than selection, then a mutant can escape

to favored patch where it can spread, but the influx of maladapted allele is also high, thus having a net impeding effect on invading allele. This means that introduction of selfing will reduce the effective migration rate, and thus increase P_2. Put differently, increasing F has the same kind of effect on P_2 as decreasing m. However, if the migration rate is already low enough, the introduction of selfing further depresses m and, thus, prevents invader from successfully escaping into the favored patch.

3.3 Consequence of selfing on the establishment of protected polymorphism

Once the allele A escapes extinction, it can either fix across both patches or segregate at intermediate frequencies for a finite but large number of generations due to divergent selection. Because both patches are of finite size, one of the two alleles will ultimately fix but this quasi-stationary behavior corresponds to protected polymorphism in deterministic models. We wish to delineate the effects of selfing on these two outcomes. Given that an invading allele can appear in any of the two patches where the favored patch accounts for z fraction of the total population across both patches, the global probability of alleles A and a becoming established is:

$$P^{(A)} = zP_1^{(A)} + (1-z)P_2^{(A)} \tag{29}$$

$$P^{(a)} = zP_1^{(a)} + (1-z)P_2^{(a)} \tag{30}$$

where $P_i^{(A)}$ and $P_i^{(a)}$ are probabilities that A and a become established, conditioning on initially appearing in patch i. For the sake of simplicity, we will further assume that patches are of equal size ($z = 1/2$), and migration is symmetrical. Probability that allele A is established is computed by parameterizing equations (26) and (27) where selection coefficients were determined from Jacobian associated with boundary $\vec{x} = [0, 0]^T$, while the establishment probability of a was calculated with elements of Jacobian associated with boundary $\vec{x} = [1, 1]^T$. For example, if selective advantage of invading heterozygote Aa relative to homozygote aa in patch i is $s_i h_i$, then the advantage of invading Aa relative to homozygote AA is $-s_i(1-h_i)/(1+s_i)$.

3.3.1 Conditions for protected polymorphism and critical migration rates

Polymorphism is protected if both $P^{(A)} > 0$ and $P^{(a)} > 0$, that is for the conditions under which the branching processes for the invasion of A and for the invasion of a are supercritical. This boils down to the extension of Bulmer's inequality given by equation (25) directly

obtained from the Jacobian matrix, which must also be satisfied with genotypes AA and aa shifted (corresponding to the supercriticality for the invasion of allele a). From this, we can define critical migration rates above which polymorphism is lost and the allele with the highest marginal fitness invades as in Yeaman and Otto (2011). Without loss of generality we can assume that $s_1 > -s_2 > 0$. For seed dispersal and viability selection M_c is then given by:

$$M_c = \begin{cases} 1, & \text{if } \frac{(1+(1-F)h_1 s_1)(1+(1-F)h_2 s_2)}{(1+s_1)(1+s_2)} > 1 \\ -\frac{(1-(1-F)h_1)(1-(1-F)h_2)s_1 s_2}{(1-(1-F)h_1)s_1+(1-(1-F)h_2)s_2+(1-F)(h_1+h_2-2(1-F)h_1 h_2)s_1 s_2}, & \text{otherwise} \end{cases} \tag{31}$$

For $F = 0$, (31) reduces to equation (5) in Yeaman and Otto (2011) and for $F = 1$ to $M_c = -\frac{s_1 s_2}{s_1+s_2}$. The first condition corresponds to a form of dominance reversal where the allele is dominant when advantageous and recessive when deleterious. Under such a condition, polymorphism can be maintained even under full migration.

Depending on dominance, selfing thus has two opposite effects. On the one hand, by unmasking recessive alleles, selfing makes selection stronger relative to migration and so enlarges the conditions for the maintenance of protected polymorphism. On the other hand, selfing makes dominance reversal conditions less likely as selection mainly operates on homozygotes, and so restricts the conditions for protected polymorphism. Figure 6 shows the dominance conditions under which selfing increases or decreases the critical migration rate. We determined these conditions in two ways: either when M_c is higher under full selfing than under full outcrossing (i.e., $M_c|_{F=0} < M_c|_{F=1}$) or whether introduction of selfing in an otherwise outcrossing population increases M_c, that is $\left.\frac{\partial M_c}{\partial F}\right|_{F=0} > 0$. For symmetrical selection ($s_2 = -s_1$), selfing increases migration rate if $h_2 > h_1$. Under the opposite condition, the critical migration is 1 and local adaptation is always maintained under outcrossing. As selection becomes more asymmetrical, selfing favors local adaptation for a broader range of dominance coefficients. Note that the effect of selfing on the migration rate is not monotonic for a narrow range of conditions (light blue region in Figure 6), where little selfing disfavors local adaption ($\left.\frac{\partial M_c}{\partial F}\right|_{F=0} < 0$) but high selfing increases it ($M_c|_{F=0} < M_c|_{F=1}$). Finally, it is worth noting that when selfing favors local adaptation, the effect is rather weak (blue curves on Figure 6). On the contrary, under dominance reversal conditions, above a given threshold, selfing dramatically reduces the migration rate from 1 to $-\frac{s_1 s_2}{s_1+s_2}$ under full selfing. Similar results are obtained for fecundity selection but with selfing favoring local adaption under broader conditions for female fecundity and for narrower conditions for male fecundity (see Appendix).

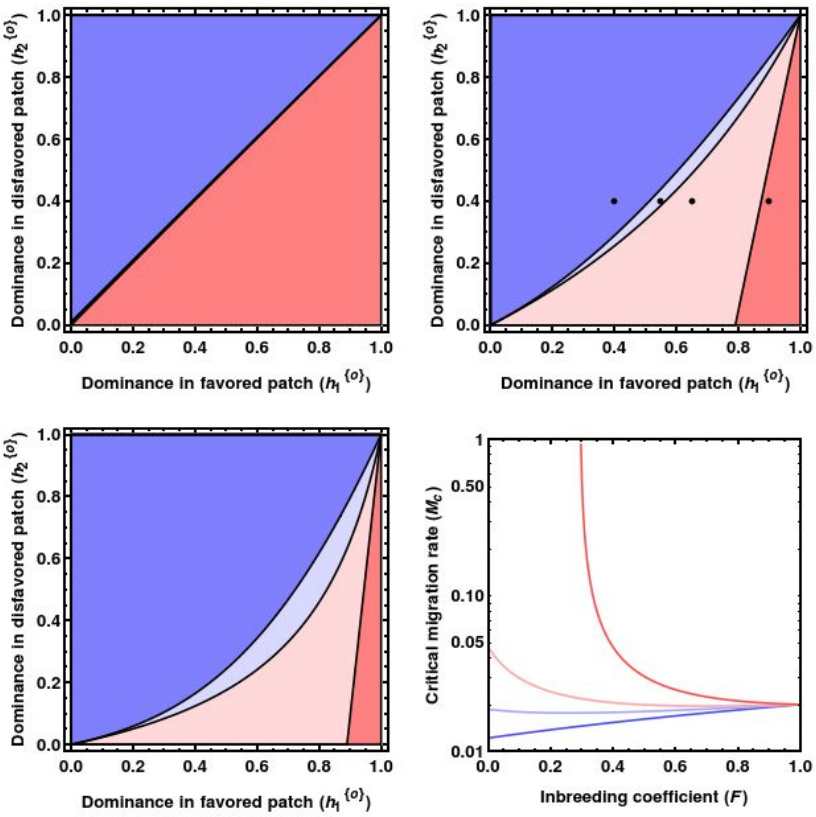


Figure 6: Effect of selfing on the critical migration rate. The three first panels show dominance conditions under which selfing increases (blue regions) or decreases (red regions) the critical migration rate, M_c. Selection is symmetrical in the first panel ($s_1 = -s_2 = 0.01$), and asymmetrical in the second ($s_1 = 0.02$ and $s_2 = -0.01$) and third panels ($s_1 = 0.05$ and $s_2 = -0.01$). In the light blue region, M_c is not monotonic with F: introduction of selfing in an outcrossing population decreases M_c ($\frac{\partial M_c}{\partial F}|_{F=0} < 0$) but above a given selfing rate, M_c becomes higher than in an outcrossing population and we always have $M_c|_{F=0} < M_c|_{F=1}$. In the dark blue region, selfing always increases M_c ($M_c|_{F=0} < M_c|_{F=1}$). In the dark red region, $M_c = 1$ in outcrossing populations and $M_c < 1$ in the light red region. The last panel illustrates how M_c varies with F in the four regions, corresponding to the black dots of the second panel. For all curves $s_1 = 0.02$ and $s_2 = -0.01$. Dark blue: $h_1 = 0.4$ and $h_2 = 0.4$; light blue: $h_1 = 0.55$ and $h_2 = 0.4$; light red: $h_1 = 0.65$ and $h_2 = 0.4$; Dark red: $h_1 = 0.9$ and $h_2 = 0.4$. Note that M_c is in log-scale.

The picture is different under pollen dispersal. Under full selfing, effective migration is zero so selection can act independently in each patch and polymorphism is always maintained. As a consequence, there is always a threshold selfing rate above which the critical migration rate is higher than in an outcrossing population. Here, we thus concentrate on the effect of the introduction of selfing in an outcrossing population, so the conditions for which $\frac{\partial M_c}{\partial F}\Big|_{F=0} > 0$. Under symmetrical selection, the conditions are the same as for seed migration. However, for asymmetrical selection, selfing increases M_c for a much broader range of parameters.

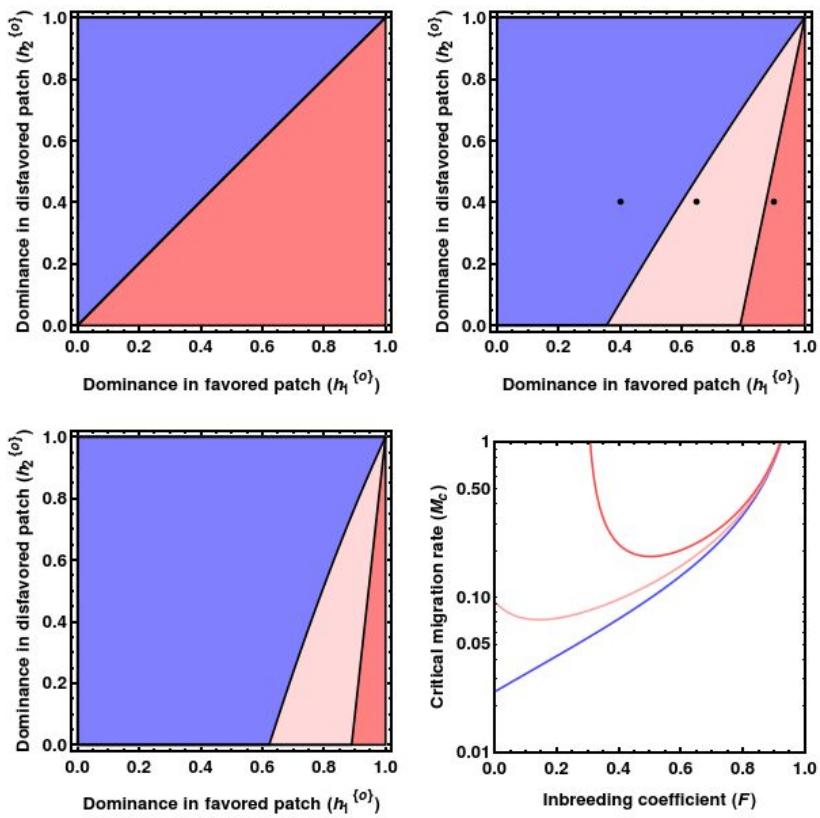


Figure 7: Same legend and same parameters as in Figure 6 but with pollen dispersal instead of seed dispersal.

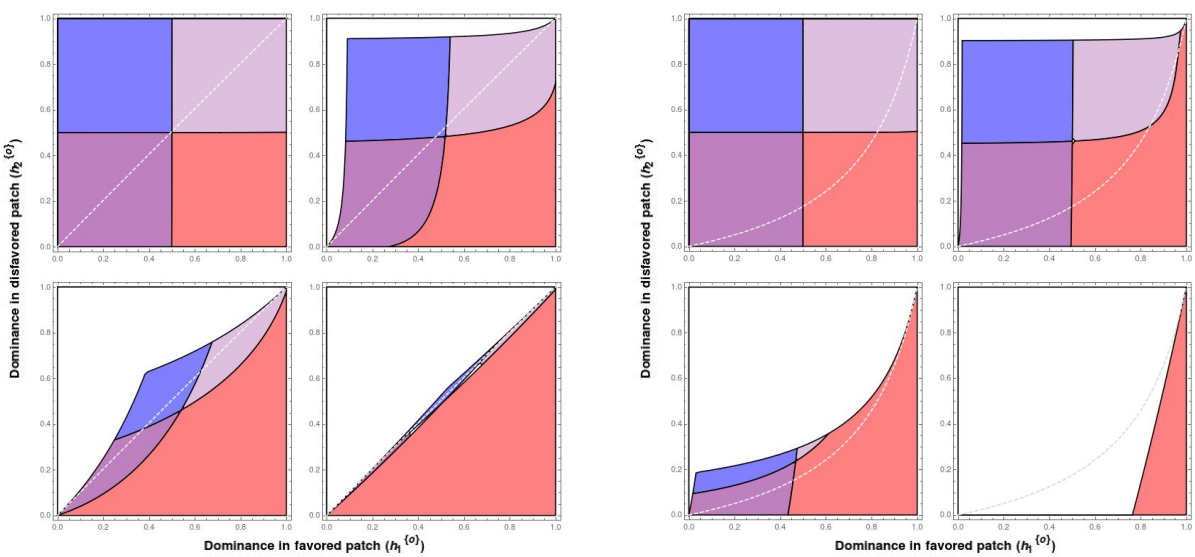
3.3.2 Establishment and maintenance of protected polymorphism

The above analysis showed how selfing affects the critical migration rate, hence the conditions for local adaptation. However, when local adaptation is possible, selfing also affects the probability of establishment and maintenance of polymorphism. This can be analyzed by

quantifying how selfing affects $P^{(A)}$ and $P^{(a)}$, which is given by:

$$b_k = \left(\frac{dP^{(k)}}{dF} \right) \bigg|_{F=0} > 0 \quad \text{where } k = \{A, a\} \tag{32}$$

Four outcomes are possible upon the introduction of selfing. First, $P^{(A)}$ increases, and $P^{(a)}$ decreases, meaning that allele A is more likely to fix across both patches in selfers than in outcrossers. Second and conversely, $P^{(A)}$ decreases, but $P^{(a)}$ increases, implying that selfers are less likely to fix A compared to an outcrossing population. Third, both $P^{(A)}$ and $P^{(a)}$ increase upon shift to selfing, meaning that selfing increases the probability that protected polymorphism is established. Fourth and final, both P_A and P_a decrease, making the protected polymorphism less likely to become established in the selfing population. Some examples are given for different migration rates with symmetrical or asymmetrical selection (Figure 8).



(a) Symmetrical selection

(b) Asymmetrical selection

Figure 8: Effect of selfing on the establishment of protected polymorphism under symmetrical (a) and asymmetrical selection (b). Five regions can be distinguished: blue: $b_A > 0$ and $b_a > 0$, red: $b_A < 0$ and $b_a < 0$, dark purple $b_A > 0$ and $b_a < 0$, light purple $b_A < 0$ and $b_a > 0$, and white: polymorphism cannot be maintained under outcrossing. The dotted lines correspond to the limiting conditions for which selfing increases the critical migration rate (above the line) as in Figure 6. (a) $s_1 = 0.01$ and $s_2 = -0.01$; $m = 0.0001, 0.001, 0.01, 0.1$ from top left to bottom right. (b) $s_1 = 0.05$ and $s_2 = -0.01$; $m =$ as in (a).

One can intuitively understand these outcomes by noting that dominance reverses when the alternative allele invades. So, if invader A is dominant, then invader a will be recessive.

The dominant allele A is likely to escape extinction when invading but unlikely to result in protected polymorphism because allele a is recessive and thus likely to go extinct. This appears clearly in the limit of low migration, where probabilities of emergence tend toward single population predictions, with four quadrants corresponding to the four dominance conditions (Figure 8). Importantly, these four regions vary with the amount of migration, and they do not directly align with the conditions given by the critical migration rates. The reason is that the sign of $P^{(A)}$ and $P^{(a)}$ only depends on the effect of selfing through the effective dominance, whereas the absolute value also depends on the reduction in effective size. Here, we only consider the two-fold reduction in N_e caused by selfing, but the linked selection can further reduce the effective size of the local population under selfing (Roze, 2016). This will not affect the critical migration rate but can strongly reduce probabilities of emergence hence the establishment of local adaptation. Following this rationale, Yeaman and Otto (2011) considered a critical migration rate above which the probability of emergence of both alleles is higher than $1/(4N_e)$. High local drift under selfing can thus strongly reduce the conditions for the establishment of local adaptation.

4 Discussion

We investigated the effect of selfing on the establishment of an allele in a population subdivided into two patches, where only one locus determines the fitness of an organism. By representing the spread of an invading allele as a branching process, we obtained a closed-form analytical solution for the probability of establishment of the locally advantageous allele for an arbitrary population selfing rate. This extends the work of Tomasini and Peischl, 2018 to include diploidy, dominance, and partial selfing. We also corrected for a typo in the derivation of their equations but which surprisingly lead to less elegant results. Below, we discuss the implication for adaptation in partially selfing species.

A well-known result is that selfing favors the establishment (and fixation) of recessive alleles but disfavors the establishment of dominant ones, with no effect under exact codominance ($h = 1/2$) (Caballero and Hill, 1992, Charlesworth, 1992). The pattern is more complex in a subdivided population with heterogeneous selection. First, the codominant allele has a higher probability of becoming established in the selfing population than in the outcrossing population if it emerges in the favored patch and generally smaller probability if it emerges in the disfavored patch. Therefore, the probability of establishment of a codominant allele is not independent of the mating system, as was the case for a single partly selfing population (Charlesworth (1992)). Second, assuming that migration and selection are symmetrical, self-

ing will increase the establishment probability when the invading mutant is recessive in the favored patch (resulting in a maximal increase in $\tilde{h}_1$) and dominant in the disfavored patch (yielding a minimal increase in $\tilde{h}_2$). If the two dominances are roughly equal, the opposed effects of selfing on effective dominances cancel each other out, and the net effect is set by the interplay between the positive effect of reduced effective migration and the negative effect of the reduction in N_e.

Once the invading allele escapes extinction, it can either fix in both populations, thus one population is fixed for a maladaptive allele, or segregate at intermediate frequencies, thus contributing to protected polymorphism, and both populations can be considered as locally adapted. It is usually thought that selfers are better locally adapted because of the reduced gene flow between patches with different selection demands (Linhart and Grant (1996)). However, selfing affects in a complex way three key parameters determining local adaptation - selection, drift, and gene flow. By separately considering seed and pollen migration, we were able to distinguish the straightforward effect of reducing gene flow from the more subtle effects of altering genotypic frequencies and drift induced by selfing.

Under seed dispersal only, selfing does not affect gene flow but still alters the propensity for local adaptation. The conditions under which selfing favors local adaptation strongly depend on the dominance of alleles in the two patches. Under outcrossing, the most favorable condition for local adaptation is under dominance reversal, that is, when the allele is dominant when locally beneficial and recessive when locally deleterious. If so, protected polymorphism can be maintained even under full migration ($M_c = 1$), which corresponds to conditions for polymorphism in Levene (1953)'s model. Under this condition, selfing destabilizes polymorphism, which cannot be maintained under full dispersal: above a given selfing rate, the critical migration rate strongly decreases (Figure 6, and see also Glémin, 2021, for the Levene's model with selfing). Conversely, under most other conditions, especially when selection is asymmetrical, selfing tends to increase the critical migration rate, hence favors local adaptation. What dominance patterns across heterogeneous habitat are the most frequent in natural populations is unknown. However, dominance reversal is maybe not as unlikely as it may seem because it can naturally arise in simple fitness landscape models (Connallon and Chenoweth, 2019). If we now consider pollen dispersal, selfing enhances local adaptation by reducing gene flow. Under dominance reversal conditions, this leads to non-monotonic behaviors where local adaptation can be the most easily maintained, either under outcrossing or under high selfing (Figure 7). Characterizing the pollination and seed dispersal modes and their quantitative impacts on gene flow appears thus crucial to make

proper predictions on the effect of selfing. Finally, if we also consider the effect of linked selection that reduces local effective size beyond the two-fold level in selfing populations (Roze, 2016), selfing reduces the possibility of local adaptation, which requires stronger selection as exemplified by Hodgins and Yeaman (2019) who simulated local adaptation in selfing populations with and without background selection. Overall, the complex and contradictory effects of selfing on local adaptation may explain why no general empirical pattern has emerged so far (Hereford, 2010, Leimu and Fischer, 2008).

Despite the large uncertainties about the global effect of selfing on local adaptation, the analysis of the model yields some predictions about the genetic architecture of local adaptation in selfing versus outcrossing species. In outcrossing species, the global adaptation from de novo mutations is predicted to be biased towards dominant mutations, the so-called Hadane's sieve (Haldane, 1927, Ronfort and Glémin, 2013), and local adaptation to beneficial dominant/deleterious recessive mutations (Yeaman and Otto, 2011). On the contrary, no sieve related to dominance is expected under high selfing. As for other forms of selection, we also predict that local adaptation should be more prominent on female traits because selfing reinforces selection on female fecundity components at the cost of male fecundity components. A similar conclusion was reached by Olito *et al.* (2018) from a model including male/female antagonism and heterogeneous habitat with full dispersal. Finally, beyond the current work, simulations showed that local adaptation genes tend to aggregate into clusters in outcrossers (Yeaman and Whitlock, 2011) whereas the genetic architecture tends to be more diffuse in selfers (Hodgins and Yeaman, 2019, Le Thierry d'Ennequin *et al.*, 1999). Thus, a natural extension of the model would be to consider local adaptation at two loci to dissect the additional effect of selfing on genetic linkage. However, using the same formalism would require following a multi-type branching process in a higher dimension (at least six), which remains highly challenging.

5 Data availability

Supplemental files are available at `https://github.com/BogiTrick/local_adaptation_single_locus` and at Figshare (`https://figshare.com/s/683d607a35f45a4978b1`). These entries contain: *Mathematica* notebook required for reproduction of all reported theoretical results and figures; C++ source code of the custom simulator used for generation of simulated data for comparison with analytics; Shell scripts for the bulk run of the simulator; and R scripts used to process the raw simulated data for plotting in the above-mentioned notebook.

6 Acknowledgements

We thank Matthew Hartfield for the critical reading of the manuscript. B. T. was supported in the initial phase of this project by Erasmus Mundus Joint Masters Degree Scholarship awarded by Erasmus+ programme. Author contributions: B.T. and S.G. developed model. B.T. wrote the code and run the simulations. B.T. and S.G. wrote the manuscript.

References

Aeschbacher, S. and R. Bürger, 2014 The effect of linkage on establishment and survival of locally beneficial mutations. Genetics **197**: 317–336.

Barton, N., 1987 The probability of establishment of an advantageous mutant in a subdivided population. Genet. Res., Camb. **50(1)**: 35–40.

Billiard, S., M. Lopez-Villavicencio, M. E. Hood, and T. Giraud, 2012 Sex, outcrossing and mating types: unsolved questions in fungi and beyond. Journal of Evolutionary Biology **25**: 1020–38.

Bulmer, M., 1972 Multiple niche polymorphism. The American Naturalist **106**: 254–257.

Caballero, A. and W. G. Hill, 1992 Effects of partial inbreeding on fixation rates of mutant genes. Genetics **131**: 493–507.

Charlesworth, B., 1992 Evolutionary rates in partially self-fertilizing species. The American Naturalist **140**: 126–148.

Connallon, T. and S. F. Chenoweth, 2019 Dominance reversals and the maintenance of genetic variation for fitness. PLoS Biol **17**: e3000118.

Damgaard, C., 2000 Fixation of advantageous alleles in partially self-fertilizing populations: the effect of different selection modes. Genetics **154(2)**: 813–821.

Felsenstein, J., 1976 The theoretical population genetics of variable selection and migration. Annual Review of Genetics **10**: 253–280.

Gale, J. S., 1990 *Theoretical population genetics*. Springer Science & Business Media.

Glémin, S., 2021 Balancing selection in self-fertilizing populations. Evolution **75(5)**: 1011–1029.

Haccou, P., P. Jagers, and V. A. Vatutin, 2005 *Branching processes: variation, growth, and extinction of populations*. Cambridge University Press.

Haldane, J., 1927 The mathematical theory of natural and artificial selection. Proc. Camb. Philos. Soc. **23**: 838—-844.

Hanschen, E. R., M. D. Herron, J. J. Wiens, H. Nozaki, and R. E. Michod, 2018 Repeated evolution and reversibility of self-fertilization in the volvocine green algae. Evolution **72**: 386–398.

Hereford, J., 2010 Does selfing or outcrossing promote local adaptation? American Journal of Botany **97**: 298–302.

Hodgins, K. A. and S. Yeaman, 2019 Mating system impacts the genetic architecture of adaptation to heterogeneous environments. New Phytol. **224**: 1201–1214.

Igic, B. and J. R. Kohn, 2006 The distribution of plant mating systems: study bias against obligately outcrossing species. Evolution Int J Org Evolution **60**: 1098–103.

Jarne, P. and J. R. Auld, 2006 Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. Evolution **60**: 1816–24.

Le Thierry d'Ennequin, M., B. Toupance, T. Robert, B. Godelle, and P. H. Gouyon, 1999 Plant domestication: a model for studying the selection of linkage. Journal of Evolutionary Biology **12**: 1138–1147.

Leimu, R. and M. Fischer, 2008 A meta-analysis of local adaptation in plants. PLOS ONE **3**: 1–8.

Levene, H., 1953 Genetic equilibrium when more than one ecological niche is available. The American Naturalist **87**: 331–333.

Linhart, Y. B. and M. Grant, 1996 Evolutionary significance of local differentiation in plants. Annu. Rev. Ecol. Syst. **27**: 237–277.

Maynard Smith, J., 1970 Genetic polymorphism in a varied environment. The American Naturalist **104**: 487–490.

Nagylaki, T., 1980 The strong-migration limit in geographically structured populations. Journal of Mathematical Biology **9**: 101—-114.

Olito, C., J. K. Abbott, and J. C. Y., 2018 The interaction between sex-specific selection and local adaptation in species without separate sexes. Phil. Trans. R. Soc. B **373**: 20170426.

Pollak, E., 1987 On the theory of partially inbreeding finite populations. i. partial selfing. Genetics **117**: 353–360.

Pollak, E. and M. Sabran, 1992 On the theory of partially inbreeding finite populations. iii. fixation probabilities under partial selfing when heterozygotes are intermediate in viability. Genetics **131**: 979–985.

Pontz, M. and R. Bürger, 2021 The effects of epistasis and linkage on the invasion of locally beneficial mutations and the evolution of genomic islands. bioRxiv .

Ronfort, J. and S. Glémin, 2013 Mating system, Haldane's sieve, and the domestication process. Evolution **67**: 1518–26.

Roze, D., 2016 Background selection in partially selfing populations. Genetics **203**: 937–57.

Sakamoto, T. and H. Innan, 2019 The evolutionary dynamics of a genetic barrier to gene flow: From the establishment to the emergence of a peak of divergence. Genetics **212(4)**: 1383–1398.

Tange, O., 2011 GNU Parallel – The Command-Line Power Tool. The USENIX Magazine **36**: 42–47.

Tomasini, M. and S. Peischl, 2018 Establishment of locally adapted mutations under divergent selection. Genetics **209**: 885–895.

Yeaman, S. and S. P. Otto, 2011 Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution **65**: 2123–2129.

Yeaman, S. and M. C. Whitlock, 2011 The genetic architecture of adaptation under migration-selection balance. Evolution **65**: 1897–911.

Appendices

A Deriving the PGF of the distribution of allele offspring number

A.1 Single population

We first derive the full probability generating function (PGF) of the number of mutant alleles under partial selfing in a single population. When the A mutant is rare, it can be found either in a Aa individual with probability $1-F$ and in a AA individual with probability F. The PGF can thus be written as $(1-F)f_{Aa}(z)+Ff_{AA}(z)$ where f_{Aa} and f_{AA} are the PGF for the parent allele being in each genotype, respectively. We can then decompose the total number of alleles, G_i, contributed by one individual of genotype i as the sum of alleles transmitted through outcrossed ovule, $N_{i,o}$, through exported pollen, $N_{i,p}$, and through selfed ovule, $N_{i,s}$. Because the mutant is rare, it can only be transmitted in heterozygote after outcrossing, so in a single copy. After selfing, two copies are always transmitted if the parent is AA and either two or one copies are transmitted if the parent is Aa. In this last case we note $N_{i,s}^{AA}$ and $N_{i,s}^{Aa}$ are the number of homozygote and heterozygote seeds produced under selfing. The total number of offspring alleles transmitted at the next generation is thus:

$$G_{AA} = N_{AA,o} + N_{AA,p} + 2N_{AA,s} \tag{A1a}$$

$$G_{Aa} = N_{Aa,o} + N_{Aa,p} + 2N_{Aa,s}^{AA} + N_{Aa,s}^{Aa} \tag{A1b}$$

We assume that the number of exported pollen and the different numbers of seeds follow independent Poisson distributions. The different means depends on the fecundity of the parent, the viability of the seed produced and the proportion of each category:

$$\begin{aligned}
\mathbb{E}[N_{Aa,o}] &= \frac{1}{2}(1-S)W_{Aa}^{♀}V_{Aa} \\
\mathbb{E}[N_{Aa,p}] &= \frac{1}{2}(1-S)W_{Aa}^{♂}V_{Aa} \\
\mathbb{E}[N_{Aa,s}^{AA}] &= \frac{1}{2}\frac{S}{2}W_{Aa}^{♀}V_{AA} \\
\mathbb{E}[N_{Aa,s}^{Aa}] &= \frac{1}{2}\frac{S}{2}W_{Aa}^{♀}V_{Aa}
\end{aligned}
\qquad
\begin{aligned}
\mathbb{E}[N_{AA,o}] &= \frac{1}{2}(1-S)W_{AA}^{♀}V_{Aa} \\
\mathbb{E}[N_{AA,p}] &= \frac{1}{2}(1-S)W_{AA}^{♂}V_{Aa} \\
\mathbb{E}[N_{AA,s}] &= \frac{1}{2}SW_{AA}^{♀}V_{AA}
\end{aligned} \tag{A2}$$

The factor 1/2 corresponds to the fact that the focal A is chosen with probability 1/2. Put another way, on average each individual transmits two alleles so the contribution for a single

allele must be halved. Finally, we use the property that the PGF of $2X$ is $f(z^2)$ where $f(z)$ is the PGF of X and that the PGF of the sum of independent variables is the product of the PGFs. Noting $g(\lambda, z) = e^{-(1-z)\lambda}$, the PGF of a Poisson distribution with mean λ, we have:

$$\begin{aligned} f_\phi(z) = (1-F)g(\mathbb{E}[N_{Aa,o}], z)g(\mathbb{E}[N_{Aa,p}], z)g(\mathbb{E}[N^{Aa}_{Aa,s}], z)g(\mathbb{E}[N^{AA}_{Aa,s}], z^2) \\ + Fg(\mathbb{E}[N_{AA,o}], z)g(\mathbb{E}[N_{AA,p}], z)g(\mathbb{E}[N_{AA,s}], z^2) \end{aligned} \tag{A3}$$

which yields equation (15) of the main text.

From the properties of PGFs we can easily obtain the moments of the distribution for the different forms of selection (see *Mathematica* notebook):

$$\begin{aligned} \mathbb{E}_\phi &= f'_\phi(1) \\ &= 1 + \tilde{h}s \end{aligned} \tag{A4a}$$

$$\begin{aligned} \mathbb{V}_\phi &= f''_\phi(1) + f'_\phi(1) - f'_\phi(1)^2 \\ &= (1+F)(1+\tilde{h}s) + o(s) \end{aligned} \tag{A4b}$$

So we retrieved the well known result that selfing increases variance in allele offspring number by $1+F$. Interestingly, the full distribution presents a peculiar non-monotonic behavior for high selfing, with an excess (resp. a deficit) in even (resp. odd) numbers. Note also that the full distributions are not exactly the same under the different modes of selection.

A.2 Two-patch model

We need to derive the distribution of the number of mutant alleles issued from one patch and staying in the same patch, with PGF $f_{i,i}(z)$, and the distribution of those establishing in the other patch, with PGF $f_{i,j}(z)$. Assuming that the distribution of resident and migrant alleles are independent, the total number of alleles produced by patch i has PGF: $f_i(z) = f_{i,i}(z)f_{i,j}(z)$. The $f_{i,j}(z)$ can be obtained using the same equations as for a single population by paying attention to the order of events to correctly set indices: male fecundity selection, pollen migration, female fecundity selection, reproduction, seed migration and viability selection. Under seed migration, fecundity selection occurs in patch i but viability

selection in patch j. The means given in equation (A2) become:

$$\begin{aligned}
\mathbb{E}[N_{Aa,o,ij}] &= \frac{1}{2}(1-S)W^{\text{♀}}_{Aa,i}M_{i,j}V_{Aa,j} \\
\mathbb{E}[N_{Aa,p,ij}] &= \frac{1}{2}(1-S)W^{\text{♂}}_{Aa,i}M_{i,j}V_{Aa,j} \\
\mathbb{E}[N^{AA}_{Aa,s,ij}] &= \frac{1}{2}\frac{S}{2}W^{\text{♀}}_{Aa,i}M_{i,j}V_{AA,j} \\
\mathbb{E}[N^{Aa}_{Aa,s,ij}] &= \frac{1}{2}\frac{S}{2}W^{\text{♀}}_{Aa,i}M_{i,j}V_{Aa,j}
\end{aligned}
\qquad
\begin{aligned}
\mathbb{E}[N_{AA,o,ij}] &= \frac{1}{2}(1-S)W^{\text{♀}}_{AA,i}M_{i,j}V_{Aa,j} \\
\mathbb{E}[N_{AA,p,ij}] &= \frac{1}{2}(1-S)W^{\text{♂}}_{AA,i}M_{i,j}V_{Aa,j} \\
\mathbb{E}[N_{AA,s,ij}] &= \frac{1}{2}SW^{\text{♀}}_{AA,i}M_{i,j}V_{AA,j}
\end{aligned}
\tag{A5}$$

with $M_{i,j} = 1 - M_{i,i}$. Plugging (A5) into (A3) yields equation (16) in the main text. From the corresponding PGFs we then retrieve the same mean as obtained by the deterministic analysis (see below), and variances inflated by $1 + F$ (see *Mathematica* notebook):

$$\mathbb{E}_{\phi,ij} = M_{i,j}(1 + \tilde{h}_i s_i) \quad \text{for fecundity selection} \tag{A6a}$$

$$= M_{i,j}(1 + \tilde{h}_j s_j) \quad \text{for viability selection} \tag{A6b}$$

$$\mathbb{V}_{\phi,ij} = (1+F)\mathbb{E}_{\phi,ij} + o(s_i, s_j) \tag{A6c}$$

Note that in subsequent analyses the order of migration and selection terms yields the same results. Under pollen migration, the PGF for resident and migrant contribution have different forms because an allele can contribute offspring to the other patch only through outcrossing and through the male pathway. For offspring contributing to the resident patch we have:

$$\begin{aligned}
\mathbb{E}[N_{Aa,o,ii}] &= \frac{1}{2}(1-S)W^{\text{♀}}_{Aa,i}V_{Aa,i} \\
\mathbb{E}[N_{Aa,p,ii}] &= \frac{1}{2}(1-S)W^{\text{♂}}_{Aa,i}(1-m_{i,j})V_{Aa,i} \\
\mathbb{E}[N^{AA}_{Aa,s,ii}] &= \frac{1}{2}\frac{S}{2}W^{\text{♀}}_{Aa,i}V_{AA,i} \\
\mathbb{E}[N^{Aa}_{Aa,s,ii}] &= \frac{1}{2}\frac{S}{2}W^{\text{♀}}_{Aa,i}V_{Aa,i}
\end{aligned}
\qquad
\begin{aligned}
\mathbb{E}[N_{AA,o,ii}] &= \frac{1}{2}(1-S)W^{\text{♀}}_{AA,i}V_{Aa,i} \\
\mathbb{E}[N_{AA,p,ii}] &= \frac{1}{2}(1-S)W^{\text{♂}}_{AA,i}(1-m_{i,i})V_{Aa,i} \\
\mathbb{E}[N_{AA,s,ii}] &= \frac{1}{2}SW^{\text{♀}}_{AA,i}V_{AA,i}
\end{aligned}
\tag{A7}$$

whereas for offspring contributing to the other patch:

$$\begin{aligned}
\mathbb{E}[N_{Aa,o,ii}] &= 0 \\
\mathbb{E}[N_{Aa,p,ii}] &= \frac{1}{2}(1-S)W^{\text{♂}}_{Aa,i}m_{i,j}V_{Aa,j} \\
\mathbb{E}[N^{AA}_{Aa,s,ii}] &= 0 \\
\mathbb{E}[N^{Aa}_{Aa,s,ii}] &= 0
\end{aligned}
\qquad
\begin{aligned}
\mathbb{E}[N_{AA,o,ii}] &= 0 \\
\mathbb{E}[N_{AA,p,ii}] &= \frac{1}{2}(1-S)W^{\text{♂}}_{AA,i}m_{i,i}V_{Aa,j} \\
\mathbb{E}[N_{AA,s,ii}] &= 0
\end{aligned}
\tag{A8}$$

Plugging (A7) and (A8) into (A3) yields equations (17a) and (17b) in the main text. The mean and variance for resident offspring are:

$$\mathbb{E}_{\phi,ii} = (1-\tilde{M}_{i,j})(1+\tilde{h}_i s_i) \quad \text{for viability and female fecundity selection} \tag{A9a}$$
$$= (1-\tilde{M}_{i,j})(1+\omega\tilde{h}_i s_i) \quad \text{for male fecundity selection} \tag{A9b}$$
$$\mathbb{V}_{\phi,ii} = \mathbb{E}_{\phi,ij}\left(1+\frac{F}{1-\tilde{M}_{ij}}\right) + o(s_i, s_j) \tag{A9c}$$

and for migrant offspring:

$$\mathbb{E}_{\phi,ij} = \tilde{M}_{i,j}(1+\tilde{h}_j s_j) \quad \text{for viability fecundity selection} \tag{A10a}$$
$$= 0 \quad \text{for female fecundity selection} \tag{A10b}$$
$$= \tilde{M}_{i,j}(1+\omega\tilde{h}_i s_i) \quad \text{for male fecundity selection} \tag{A10c}$$
$$\mathbb{V}_{\phi,ij} = \mathbb{E}_{\phi,ij} + o(s_i, s_j) \tag{A10d}$$

where $\tilde{M}_{i,j} \approx m_{i,j}(1-S)/2$ are effective migration rates as defined in the main text (equations 12a to 12c) and ω is a correcting factor defined in equation (13). Compared to seed migration, the variance is not uniformly increased by $1+F$. As migrant offspring can only be produced through outcrossing, the distribution is simply Poisson and the variance equal to the mean. On the contrary, because the proportion of outcrossed offspring contributing to the resident patch is reduced due to pollen migration, the variance is inflated by more than $1+F$. However, the difference between the two modes of migration does not affect the following approximations for weak selection.

B The establishment probability approximated to weak selection

Following the procedure outlined in Haccou *et al.* (2005) (Section 5.6.2) and previously adapted for a similar problem by Tomasini and Peischl (2018), we seek to approximate P by working with slightly supercritical process. Let ρ be the leading eigenvalue and $\mathbf{u} = [u_1, u_2]^T$ and $\mathbf{v} = [v_1, v_2]^T$ normed left and right eigenvectors of of the mean reproductive matrix $\mathbf{M}$.

Eigenvectors are normalized such that:

$$u_i > 0, \quad v_i > 0, \quad i \in \{1, 2\}$$
$$\sum_i u_i v_i = \sum_i u_i = 1$$

We choose a parameter ϵ in the model, such that the process is slightly supercritical when ϵ is small. All other parameters in the model are rescaled by ϵ. More formally:

$$\rho(\epsilon) \to 1 \quad \text{as} \quad \epsilon \to 0$$

Eigenvalues and eigenvectors of $\mathbf{M}$ are dependent on ϵ. Probability of establishment starting from a single copy of allele in patch i is given by:

$$P_i = 2(\rho(\epsilon) - 1)\frac{v_i(\epsilon)}{B(\epsilon)} + o(\epsilon) \tag{A11}$$

$$B(\epsilon) = \sum_i u_i(\epsilon)\mathrm{Var}\left(\phi_{ij}\sum_j v_j(\epsilon)\right) + (\rho(\epsilon)(\rho(\epsilon) - 1))\sum_j u_j(\epsilon)v_j(\epsilon)^2 \tag{A12}$$

As $\epsilon \to 0$, ρ approaches unity and the second term in B (equation (A12)) can be neglected. As per equation 5.85 in Haccou *et al.* (2005), we take that $B(\epsilon) \to B(0)$ and $v_i(\epsilon) \to v_i(0)$:

$$P_k \approx 2(\rho(\epsilon) - 1)\frac{v_i(0)}{B(0)} \tag{A13}$$

$$B(0) \approx \sum_k u_i(0)\mathrm{Var}\left(\phi_{ij}\sum_j v_j(0)\right) \tag{A14}$$

Recall that $\tilde{s}_i$ is the advantage of an allele in patch i accounting for the effect of self-fertilization. Assuming weak selective advantage, $\tilde{s}_1$ is taken as ϵ. All other parameters are expressed in terms of $\tilde{s}_1$:

$$\tilde{M}_{ij} = \tilde{s}_1\chi_{ij}, \quad \tilde{s}_2 = \zeta\tilde{s}_1$$

The leading eigenvalue of $\mathbf{M}$ can be written as $\rho = 1 + \tilde{c}\tilde{s}_1 + o(\tilde{s}_1^2)$. Dropping the higher-order terms in $\tilde{s}_1$, we retrieve the expression for $\tilde{c}$:

$$\tilde{c} = \frac{1}{2}\left(\sqrt{(\zeta + \chi_{12} - 1)^2 + 2\chi_{21}(-\zeta + \chi_{12} + 1) + \chi_{21}^2} + \zeta - \chi_{12} - \chi_{21} + 1\right)$$

By taking only the zeroth term of v_i and B in Taylor expansion about $\tilde{s}_1 = 0$, we obtain the

approximation of the probability of establishment conditioning on a single mutant appearing in patch i:

$$P_k \approx 2\tilde{c}\frac{v_k(0)}{B(0)}\tilde{s}_1 \tag{A15}$$

Transforming back to original variables $\tilde{M}_{ij}$ and $\tilde{s}_i$ and letting

$$\psi = \sqrt{2\tilde{M}_{12}\left(\tilde{M}_{21} - \tilde{s}_1 + \tilde{s}_2\right) + \left(\tilde{M}_{21} + \tilde{s}_1 - \tilde{s}_2\right)^2 + \tilde{M}_{12}^2}$$

we have:

$$\begin{aligned} P_1 = & \qquad \text{(A16)} \\ & 2\left(-\tilde{M}_{12}\left(-2\tilde{M}_{21} + 2\tilde{s}_1 - 2\tilde{s}_2 + \psi\right) + \left(\tilde{M}_{21} + \tilde{s}_1 - \tilde{s}_2\right)\left(\tilde{M}_{21} + \tilde{s}_1 - \tilde{s}_2 + \psi\right) + \tilde{M}_{12}^2\right) \\ & \left(-\tilde{M}_{12} + \tilde{M}_{21} + \tilde{s}_1 - \tilde{s}_2 + \psi\right)\left(-\tilde{M}_{12} - M_{21} + \tilde{s}_1 + \tilde{s}_2 + \psi\right) \\ & \left((1+F)\left(\left(-\tilde{M}_{12} + \tilde{M}_{21} + \tilde{s}_1 - \tilde{s}_2 + \psi\right)^3 + 8\tilde{M}_{12}^2\tilde{M}_{21}\right)\right)^{-1} \\ P_2 = & \qquad \text{(A17)} \\ & 4\tilde{M}_{12}\left(-\tilde{M}_{12}\left(-2\tilde{M}_{21} + 2\tilde{s}_1 - 2\tilde{s}_2 + \psi\right) + \left(\tilde{M}_{21} + \tilde{s}_2 - \tilde{s}_2\right)\left(\tilde{M}_{21} + \tilde{s}_1 - \tilde{s}_2 + \psi\right) + \tilde{M}_{12}^2\right) \\ & \left(-\tilde{M}_{12} - \tilde{M}_{21} + \tilde{s}_1 + \tilde{s}_2 + \psi\right) \\ & \left((1+F)\left(\left(-\tilde{M}_{12} + \tilde{M}_{21} + \tilde{s}_1 - \tilde{s}_1 + \psi\right)^3 + 8\tilde{M}_{12}^2\tilde{M}_{21}\right)\right)^{-1} \end{aligned}$$

When migration is symmetrical ($\tilde{M}_{12} = \tilde{M}_{21} = \tilde{M}$), ψ reduces to $\sqrt{4\tilde{M}^2 + (\tilde{s}_1 - \tilde{s}_2)}$, previously defined as the scaled measure of the heterogeneity in selection and migration. Setting $F = 0$, the equations (A16) and (A17) behave similarly to those derived previously (Tomasini and Peischl (2018)). Comparing the analytical solution against simulations, one can see that the solution reported here fits simulations slightly better than the Tomasini-Peischl result (Figure A9). The discrepancy is probably caused by the latter's use of B term from Aeschbacher and Bürger (2014) (see their equation S22), which neglects to square elements of the eigenvectors after factoring them out of the variance. We, on the other hand, computed B directly from Haccou *et al.* (2005), which does not suffer from this error. Curiously, the Tomasini-Peischl approximation is much more elegant as the denominator can be interpreted as the measure of the heterogeneity of migration and selection. Relative to

the result reported in Sakamoto and Innan (2019), our solution had identical performance in the favored patch, and worse performance in the disfavored patch (pink dashed lines). This discrepancy occurs only when the migration rate is low.

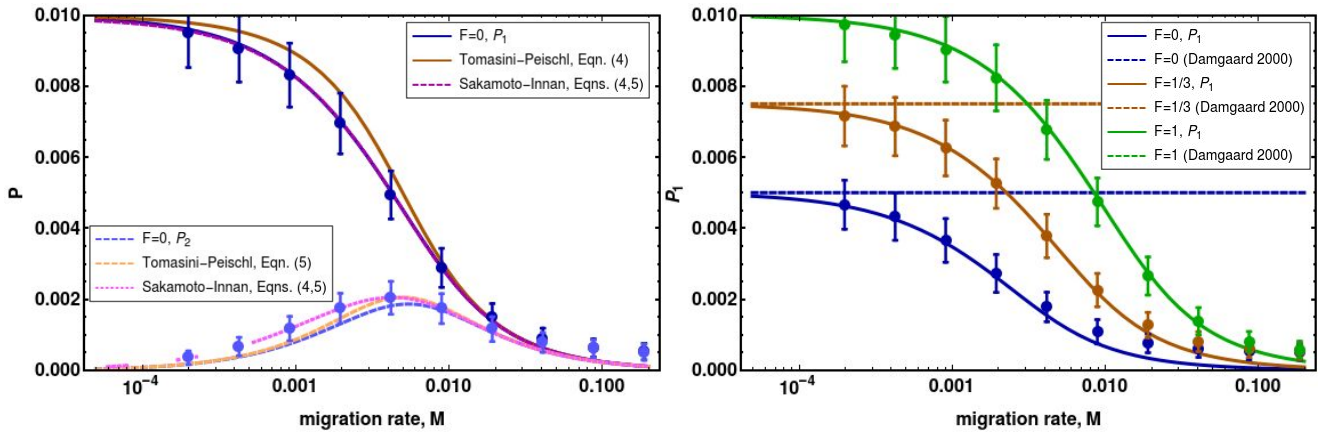


Figure A9: Comparison of analytical solution to simulated data and previous results. Comparison to approximation Tomasini and Peischl (2018) (left), and comparison to the single-patch heuristic under fecundity selection Damgaard (2000) (right). All equations parameterized with $h_1^{\circ} = h_2^{\circ} = h_1^{\mathrm{♀}} = -h_2^{\mathrm{♀}} = 1/2$, $s_1^{\circ} = -s_2^{\circ} = s_1^{\mathrm{♀}} = -s_2^{\mathrm{♀}} = 0.01$, $M_{12} = M_{21} = M$.

C The criterion for escaping extinction

C.1 Seed dispersal

If the selective disadvantage in the disfavored patch is too large or migration is too strong, the spreading locally advantageous allele can be swamped by its deleterious counterpart. The range of parameters that are necessary but not sufficient for a successful invasion are obtained by linearizing the system about $\vec{x} = [0, 0]^T$ and investigating the conditions required for this equilibrium to be locally unstable. If migration occurs prior to selection – which is the case when seed disperses and selection acts on viability, then Jacobian $\mathbf{J}$ is:

$$\mathbf{J} = \begin{bmatrix} (1 - M_{12})\Delta\mathcal{W}_{0,1} & M_{12}\Delta\mathcal{W}_{0,1} \\ M_{21}\Delta\mathcal{W}_{0,2} & (1 - M_{21})\Delta\mathcal{W}_{0,2} \end{bmatrix} \tag{A18}$$

$$\Delta W_{0,i} = (F\mathcal{W}_{i,AA} + (1 - F)\mathcal{W}_{i,Aa})/\mathcal{W}_{i,aa} \tag{A19}$$

Symbol $\mathcal{W}$ is a place-holder for any of the three selection modes (V, $W^{\mathrm{♀}}$, or $W^{\mathrm{♂}}$). If selection occurs prior to migration – which happens when seed disperses and selection acts on sexual components– then Jacobian is:

$$\mathbf{J} = \begin{bmatrix} (1-M_{12})\Delta\mathcal{W}_{0,1} & M_{12}\Delta\mathcal{W}_{0,2} \\ M_{21}\Delta\mathcal{W}_{0,1} & (1-M_{21})\Delta\mathcal{W}_{0,2} \end{bmatrix} \tag{A20}$$

However, the eigenvalues of these two matrices are identical, so we here use (A18). The equilibrium is locally unstable whenever the leading eigenvalue of $\mathbf{J}$ is greater than unity:

$$\rho = \frac{1}{2}\Bigg((1-M_{12})\Delta\mathcal{W}_{0,1} + (1-M_{21})\Delta\mathcal{W}_{0,2} + \sqrt{4(M_{12}+M_{21}-1)\Delta\mathcal{W}_{0,1}\Delta\mathcal{W}_{0,2} + \Big((M_{21}-1)\Delta\mathcal{W}_{0,1} + (M_{21}-1)\Delta\mathcal{W}_{0,2}\Big)^2} \Bigg) \tag{A21}$$

Term $\Delta\mathcal{W}_{0,i}$ can be thought of as the rate of spread of A allele in i^{th} patch. Then $\Delta\mathcal{W}_{0,1} > 1$ (because A is advantageous in the first patch), and $0 < \Delta\mathcal{W}_{0,2} < 1$ (because A is deleterious in the second patch). It is important to note that these inequalities hold regardless of the mode of selection. More formally:

$$\Delta V_{0,1} > 1, \quad 1 > \Delta V_{0,2} > 0 \tag{A22}$$

$$\Delta W_{0,1}^{\mathrm{♀}} > 1, \quad 1 > W_{0,2}^{\mathrm{♀}} > 0 \tag{A23}$$

$$\Delta W_{0,1}^{\mathrm{♂}} > 1, \quad 1 > W_{0,2}^{\mathrm{♂}} > 0 \tag{A24}$$

Substituting (A19) in (A21), and rearranging (see *Mathematica* notebook for details), we find that $\rho > 1$ when:

$$\frac{M_{12}}{\frac{((1-F)\mathcal{W}_{1,Aa}+F\mathcal{W}_{1,AA})-\mathcal{W}_{1,aa}}{(1-F)\mathcal{W}_{1,Aa}+F\mathcal{W}_{1,AA}}} + \frac{M_{21}}{\frac{((1-F)\mathcal{W}_{2,Aa}+F\mathcal{W}_{2,AA})-\mathcal{W}_{2,aa}}{(1-F)\mathcal{W}_{2,Aa}+F\mathcal{W}_{2,AA}}} < 1 \tag{A25}$$

Terms in denominators of inequality A25 represent the relative fitness of the invading allele A. Letting $F = 0$ and parameterizing such that $\mathcal{W}_{i,AA} = 1-s_1^\circ$, $\mathcal{W}_{i,Aa} = 1$, and $\mathcal{W}_{i,aa} = 1-t_1^\circ$, we retrieve Bulmer's inequality $M_{12}/t_1^\circ + M_{21}/t_2^\circ$. Parameterizing according to our fitness scheme ($\mathcal{W}_{i,AA} = 1 + s_i^\circ$, $\mathcal{W}_{i,Aa} = 1 + s_i^\circ\tilde{h}_i^\circ$, and $\mathcal{W}_{i,aa} = 1$) yields:

$$\frac{\left(1+s_1^\circ\tilde{h}_1^\circ\right)M_{12}}{s_1^\circ\tilde{h}_1^\circ} + \frac{\left(1+s_2^\circ\tilde{h}_2^\circ\right)M_{21}}{s_2^\circ\tilde{h}_2^\circ} \tag{A26}$$

Conversely to the previous case, $0 < \Delta\mathcal{W}_{1,1} < 1$ (as invading allele a is deleterious in the first patch) and $\Delta\mathcal{W}_{1,2} > 1$ (given that a is beneficial in the second patch). Once the allele A

has escaped extinction, it can either fix in both patches or be maintained for finite number of generations by divergent selection. Thus, the criterion for protected polymorphism is that both allele A and a can escape extinction. By linearizing the system of replicator equations about $\vec{x} = [1, 1]^T$ we obtain Jacobian:

$$\mathbf{J} = \begin{bmatrix} (1 - M_{12})\Delta\mathcal{W}_{1,1} & M_{12}\Delta\mathcal{W}_{1,1} \\ M_{21}\Delta\mathcal{W}_{1,2} & (1 - M_{21})\Delta\mathcal{W}_{1,2} \end{bmatrix} \tag{A27}$$

$$\Delta\mathcal{W}_{1,i} = (F\mathcal{W}_{i,aa} + (1 - F)\mathcal{W}_{i,Aa})/\mathcal{W}_{i,AA} \tag{A28}$$

Therefore, allele a is allowed to invade whenever:

$$\frac{M_{12}}{\frac{((1-F)\mathcal{W}_{1,Aa}+F\mathcal{W}_{1,aa})-\mathcal{W}_{1,AA}}{(1-F)\mathcal{W}_{1,Aa}+F\mathcal{W}_{1,aa}}} + \frac{M_{21}}{\frac{((1-F)\mathcal{W}_{2,Aa}+F\mathcal{W}_{2,aa})-\mathcal{W}_{2,AA}}{(1-F)\mathcal{W}_{2,Aa}+F\mathcal{W}_{2,aa}}} < 1 \tag{A29}$$

which reduces to $M_{12}/s_1^\circ + M_{21}/s_2^\circ$ with fitness notation of Bulmer.

C.2 Pollen dispersal

The analysis is more complicated for the three selection scenarios when pollen disperses because one has to re-parameterize migration rate in addition to dominance coefficients. We use Jacobian of the form:

$$\mathbf{J} = \begin{bmatrix} (1 - \tilde{M}_{12})\Delta\mathcal{W}_{0,1} & \tilde{M}_{12}\Delta\mathcal{W}_{0,1} \\ \tilde{M}_{21}\Delta\mathcal{W}_{0,2} & (1 - \tilde{M}_{21})\Delta\mathcal{W}_{0,2} \end{bmatrix} \tag{A30}$$

The leading eigenvalue of (A30) is:

$$\rho = \frac{1}{2}\Bigg((1 - \tilde{M}_{12})\Delta\mathcal{W}_{0,1} + (1 - \tilde{M}_{21})\Delta\mathcal{W}_{0,2} + \sqrt{4(\tilde{M}_{12} + \tilde{M}_{21} - 1)\Delta\mathcal{W}_{0,1}\Delta\mathcal{W}_{0,2} + \Big((\tilde{M}_{21} - 1)\Delta\mathcal{W}_{0,1} + (\tilde{M}_{21} - 1)\Delta\mathcal{W}_{0,2}\Big)^2}\Bigg), \tag{A31}$$

where $\Delta\mathcal{W}$ and $\tilde{M}$ are parameterized as outlined in Section 2.2. Given that $1/2 \geq \tilde{M}_{ij} \geq 0$ for all i and j, and

$$\Delta V_{0,1} > 1, \quad 1 > \Delta V_{0,2} > 0 \tag{A32}$$

$$\Delta W_{0,1}^{\mathrm{♀}} > 1, \quad 1 > W_{0,2}^{\mathrm{♀}} > 0, \tag{A33}$$

the same inequalities hold as in the case of seed dispersal. However, under male fecundity selection migration rates have to be:

$$m_{ij} < \frac{W^{♂}_{j,aa}\Big(W^{♂}_{i,aa} - (1-F)W^{♂}_{i,Aa} - FW^{♂}_{i,AA}\Big)}{W^{♂}_{i,aa}\Big((1-F)W^{♂}_{j,Aa} + FW^{♂}_{j,AA}\Big) - W^{♂}_{j,aa}\Big((1-F)W^{♂}_{i,Aa} + FW^{♂}_{i,AA}\Big)}, \tag{A34}$$

for both m_{12} and m_{21}, or:

$$m_{ij} > \frac{W^{♂}_{j,aa}\Big(W^{♂}_{i,aa} - (1-F)W^{♂}_{i,Aa} - FW^{♂}_{i,AA}\Big)}{W^{♂}_{i,aa}\Big((1-F)W^{♂}_{j,Aa} + FW^{♂}_{j,AA}\Big) - W^{♂}_{j,aa}\Big((1-F)W^{♂}_{i,Aa} + FW^{♂}_{i,AA}\Big)}, \tag{A35}$$

for both m_{12} and m_{21}. The conditions (A34) and (A35) are identical to saying that $\omega_i > 0$ for all i, or $\omega_i < 0$ for all i. A possible intuitive explanation for these conditions is as follows. If migration from favored to disfavored patch is too high, the mutant alleles are transferred to a disfavored patch where they are purged, thus causing the mutant to go extinct. If, on the other hand, migration from disfavored to favored patch is high, then the spreading mutant is swamped by the influx of deleterious residents from the opposite patch.

D Simulation method

Selection and migration are assumed to alter genotype frequencies deterministically. Genetic drift is implemented by randomly drawing genotypes from multinomial distribution right after reproduction. We relax the assumption that the population has to reach equilibrium in the inbreeding coefficient F. Each simulation run terminates in a successful or failed invasion and is composed of the following four steps:

1. Inject a single heterozygote containing allele A in a population that is fixed for a allele. When interested in P_i, the mutant is injected in i^{th} patch.

2. Update genotype frequencies by applying equations (2a)–(4b); This emulates selection on sexual components, pollen dispersal, and reproduction (including selfing).

3. Sample the genotypes from multinomial distribution to determine the genotype frequencies after the reproduction step: $\mathbf{MN}([X_i, Y_i, Z_i], N_i^*)$, where N_i^* is the size of ith patch.

4. Update genotype frequencies due to seed dispersal by implementing (5a) and (5b).

5. Compute the number of each genotype after viability selection as $X_i^* := X''_i V_{i,AA}$, $Y_i^* := Y'''_i V_{i,Aa}$, and $Z_i^* := 1 - X^* - Y^*$.

6. Is the number of invading mutant alleles larger than 1000? If not, convert genotype numbers to genotype frequencies by dividing drawn genotype numbers with the total patch size and begin the new generation by going to step 2. If yes, count it as a successful invasion and terminate the simulation run.

If the mutant allele does not invade in 10,000 generations, we count the simulation run as a failed invasion. The establishment probability is obtained by running 10,000 simulations and computing the fraction of runs that ended in the successful invasion. All error bars in the figures correspond to the standard deviation of this metric. The chosen number for invasion threshold is well over $1/s_1$, given that we use $s_1 = 0.01$. Both patches contain 10,000 individuals, and populations are always in $|Ns| \gg 1$ regime.

E Simulations under various selection and migration modes

Our approximation gives a good fit for simulated data under female and male fecundity, although the fit is not excellent as in the case of viability selection (Figures A10–A13). In the case of male fecundity selection under complete selfing ($S = 1$), the male component does not contribute to the fitness, and the allele is expected to behave neutrally. One can see that this occurs because the probability that the invading allele reaches the threshold that we use to determine whether the invasion is successful is inversely proportional to the size of the threshold, $1/N_{\text{thres}}$, (grey dashed line in figures below). The figures below compare the analytics to simulated data, with upper row denoting the case when allele originates in the favored patch, and bottom row depicts the same dynamic but when allele appears in the disfavored patch. From left to right, columns show codominant, partially dominant, and partially recessive case, respectively.

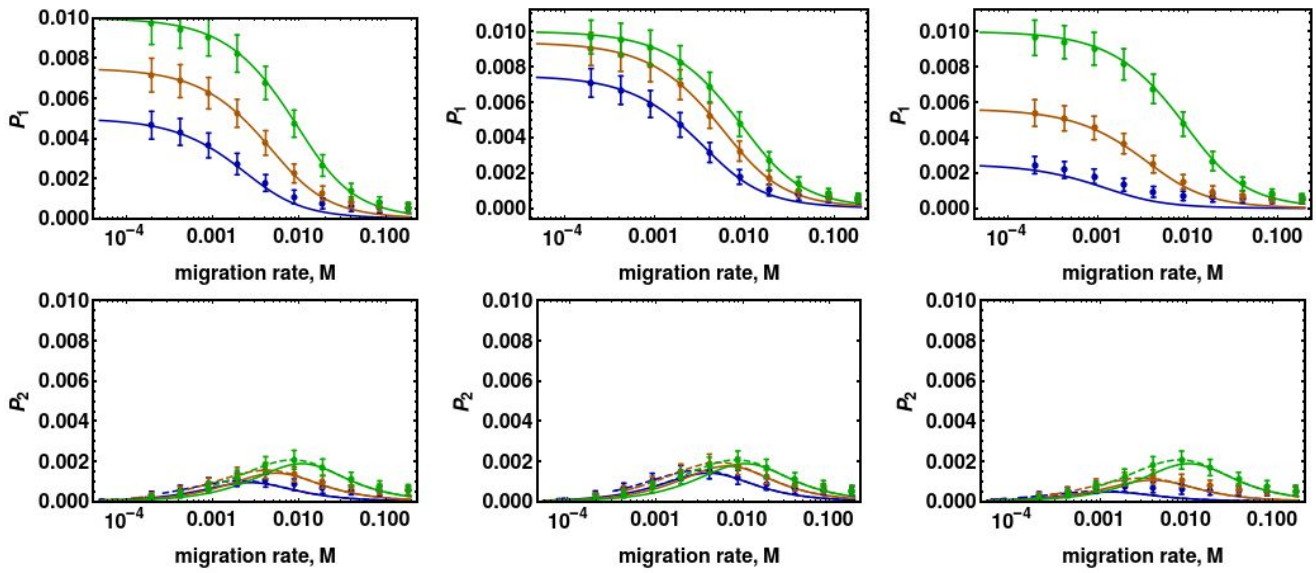


Figure A10: Comparison of analytical solution to simulations when selection acts on the female fecundity and only seed disperses. Left column: codominant case ($h_1^{♀} = h_2^{♀} = 1/2$); Middle column: dominant case ($h_1^{♀} = h_2^{♀} = 3/4$); Right column: recessive case ($h_1^{♀} = h_2^{♀} = 1/4$). Upper and lower panels depict the establishment probability conditioning on allele emerging in favored and disfavored patch, respecitvely. Parameters: $s_1^{♀} = -s_2^{♀} = 0.01$, $M_{12} = M_{21} = M$.

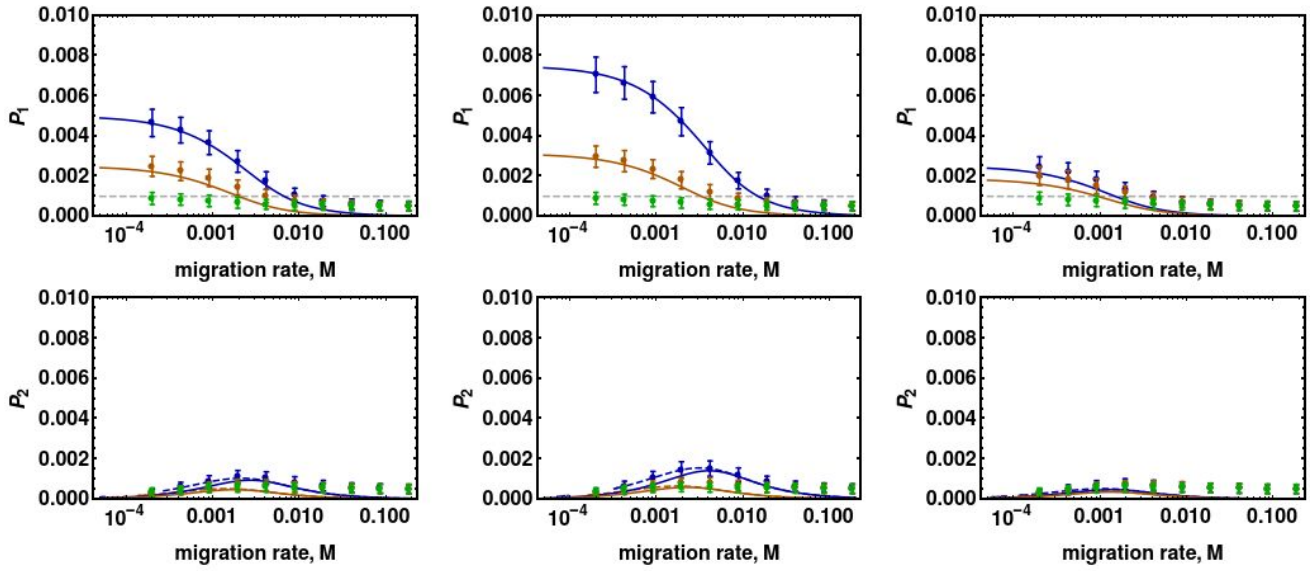


Figure A11: Comparison of analytical solution to simulations when selection acts on male fecundity and only seed disperses. Left column: codominant case ($h_1^{♂} = h_2^{♂} = 1/2$); Middle column: dominant case ($h_1^{♂} = h_2^{♂} = 3/4$); Right column: recessive case ($h_1^{♂} = h_2^{♂} = 1/4$). Upper and lower panels depict the establishment probability conditioning on allele emerging in favored and disfavored patch, respectively. Parameters: $s_1^{♂} = -s_2^{♂} = 0.01$, $M_{12} = M_{21} = M$.

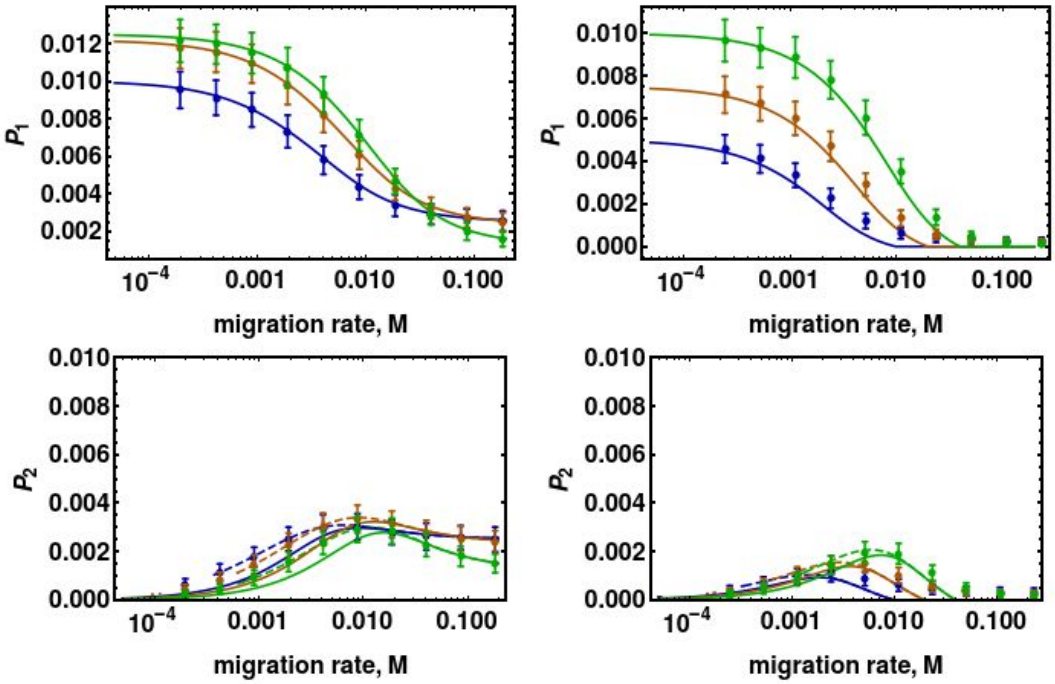


Figure A12: Comparison of analytical solution to simulations when selection acts on the female fecundity, only seed disperses, and selection or migration are asymmetrical. Left column: asymmetric selection ($s_1^{♀} = 0.01$, $s_2^{♀} = -0.02$, $h_1^{♀} = h_2^{♀} = 1/2$); Right column: asymmetric migration ($s_1^{♀} = -s_2^{♀} = 0.01$, $h_1^{♀} = h_1^{♀} = 1/2$, $M_{12} = M$, and $M_{21} = 1.25M$).

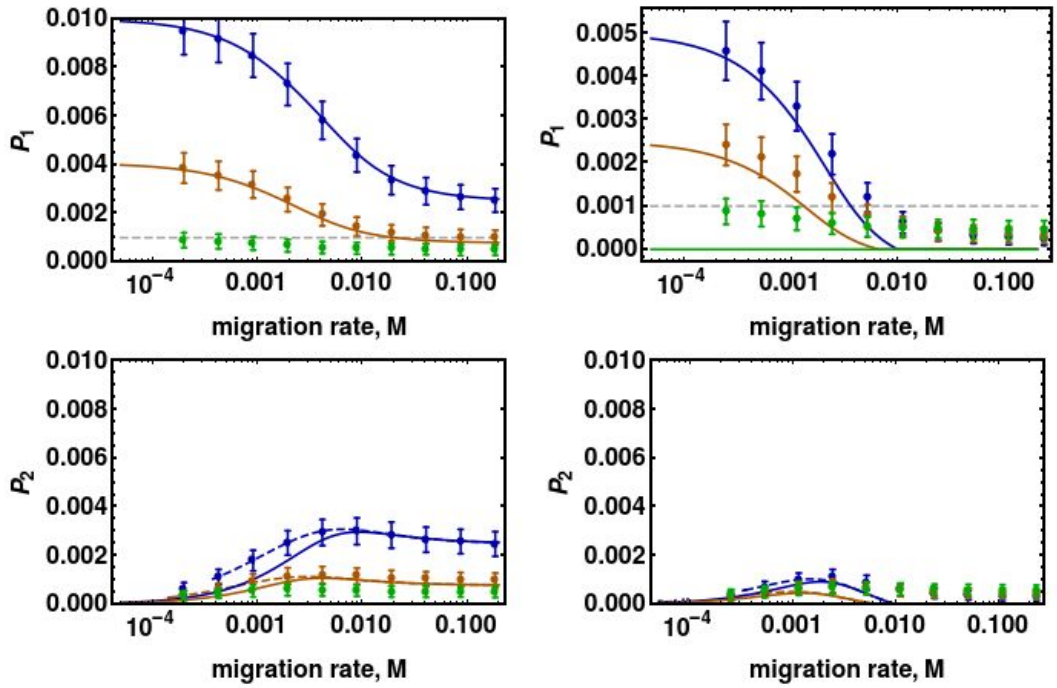


Figure A13: Comparison of analytical solution to simulations when selection acts on the male fecundity, only seed migrates, and selection or migration are asymmetrical. Left column: asymmetric selection ($s_1^{♂} = 0.01$, $s_2^{♂} = -0.02$, $h_1^{♂} = h_2^{♂} = 1/2$); Right column: asymmetric migration ($s_1^{♂} = -s_2^{♂} = 0.01$, $h_1^{♂} = h_1^{♂} = 1/2$, $M_{12} = M$, and $M_{21} = 1.25M$).

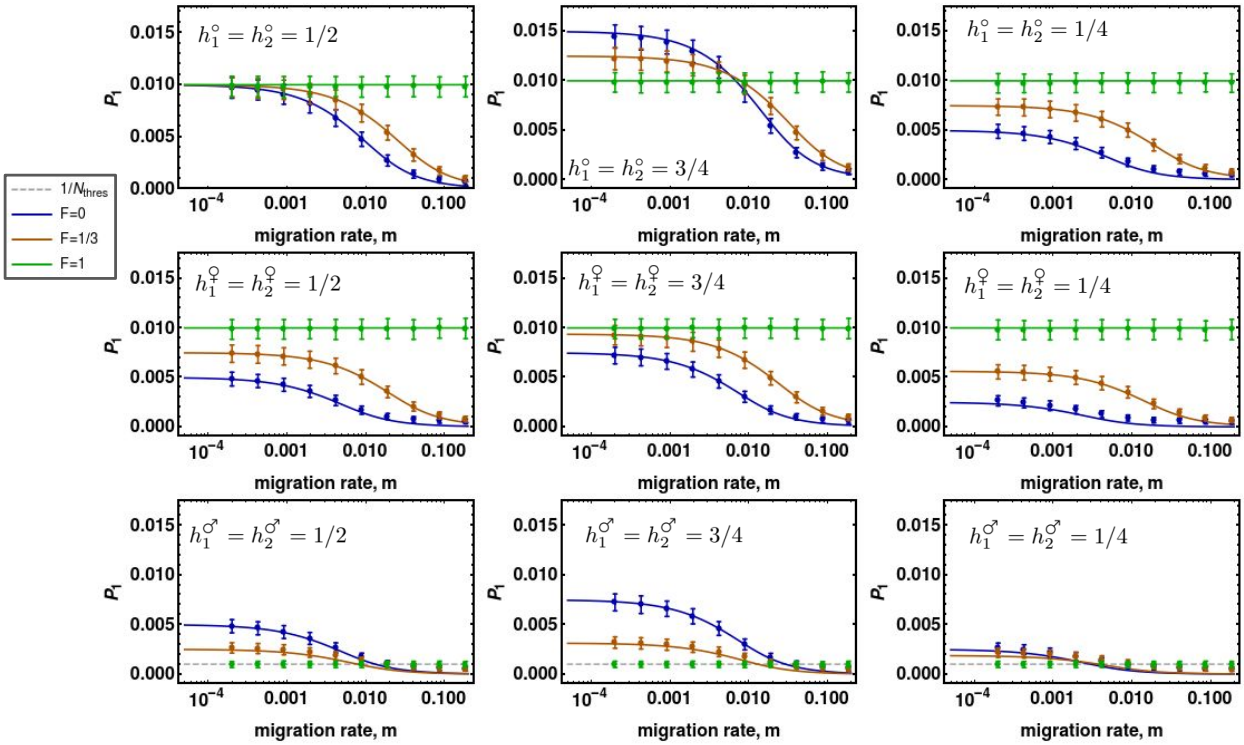


Figure A14: Comparison of analytical solution to simulations when selection acts on the viability and only pollen disperses. Allele starts in the favored patch. Top row: viability selection; Middle row: female fecundity selection; Bottom row: Male fecundity selection. Dominances reported in panels. Other parameters: corresponding seelection coefficients are always $s_1 = -s_2 = 0.01$, and $m_{12} = m_{21} = m$.

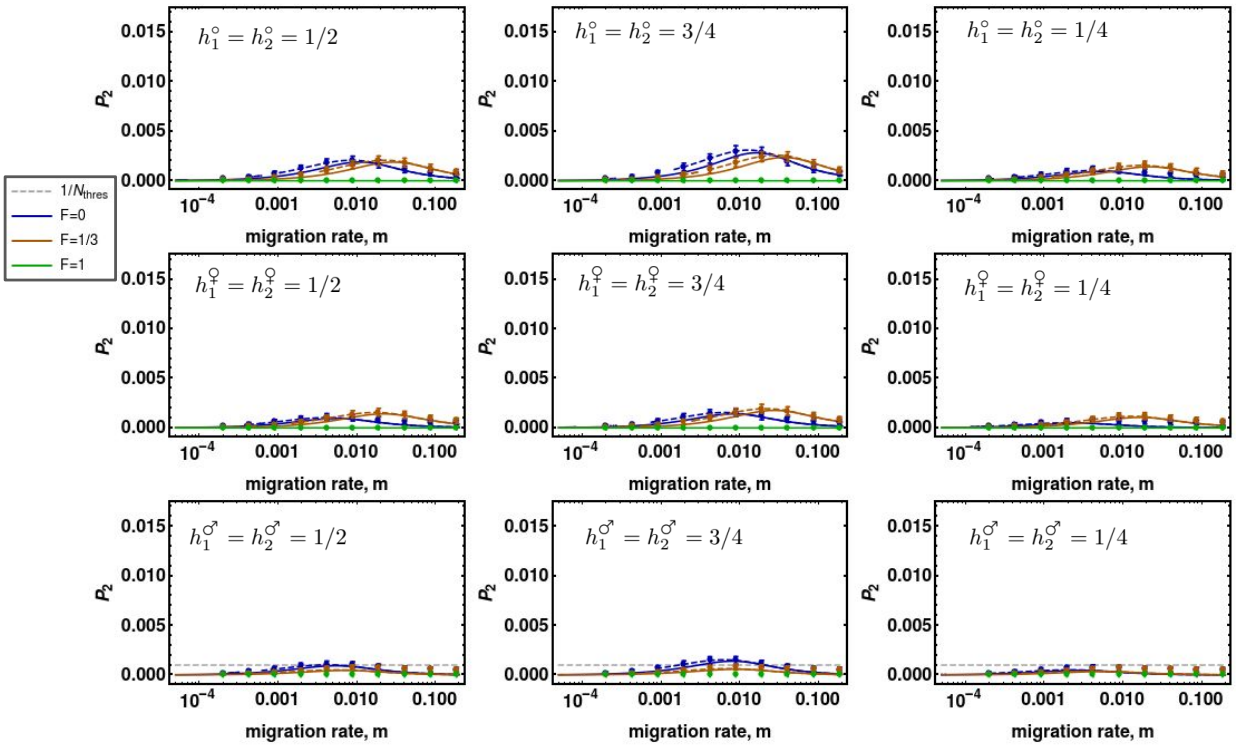


Figure A15: Same as above, but allele starts in the disfavored patch.

F Procedure for computing β indicators

Suppose one wants to derive β indicator when selection operates on j^{th} fitness component (where $j \in \{V, F, M\}$) in k^{th} patch (where $k \in \{1, 2\}$). Selfing affects the establishment probability through three factors: the effective population size ($\beta_k(\tilde{N})$), effective favored and disfavored dominance ($\beta_k(\tilde{h}_1)$ and $\beta_k(\tilde{h}_2)$), and the effective migration rate ($\beta_k(\tilde{M})$). Starting from equations (A16) and (A17), expressions for β are obtained using the following steps in a sequential manner:

- $\beta_k(\tilde{N})$: Parameterize migration according to eqns. 12a–12c, and then set $F = 0$ to exclude the effect via $\tilde{M}$. Let $\tilde{s}_i^{\{j\}} = s_i^{\{j\}}\tilde{h}_i^{\{j\}}$ for $i \in \{1, 2\}$. Next, set $F = 0$ to exclude the effect on dominances. Take a derivative in F and evaluate at $F = 0$.

- $\beta_k(\tilde{h}_1)$: Parameterize migration according to eqns. 12a–12c, and then set $F = 0$ to exclude the effect via $\tilde{M}$. Set $F = 0$ to exclude the effect via $\tilde{N}$. Let $\tilde{s}_2^{\{j\}} = s_2^{\{j\}}\tilde{h}_2^{\{j\}}$ and then set $F = 0$ to exclude the effect via $\tilde{h}_2$. Let $\tilde{s}_1^{\{j\}} = s_1^{\{j\}}\tilde{h}_1^{\{j\}}$. Take a derivative in F and evaluate at $F = 0$.

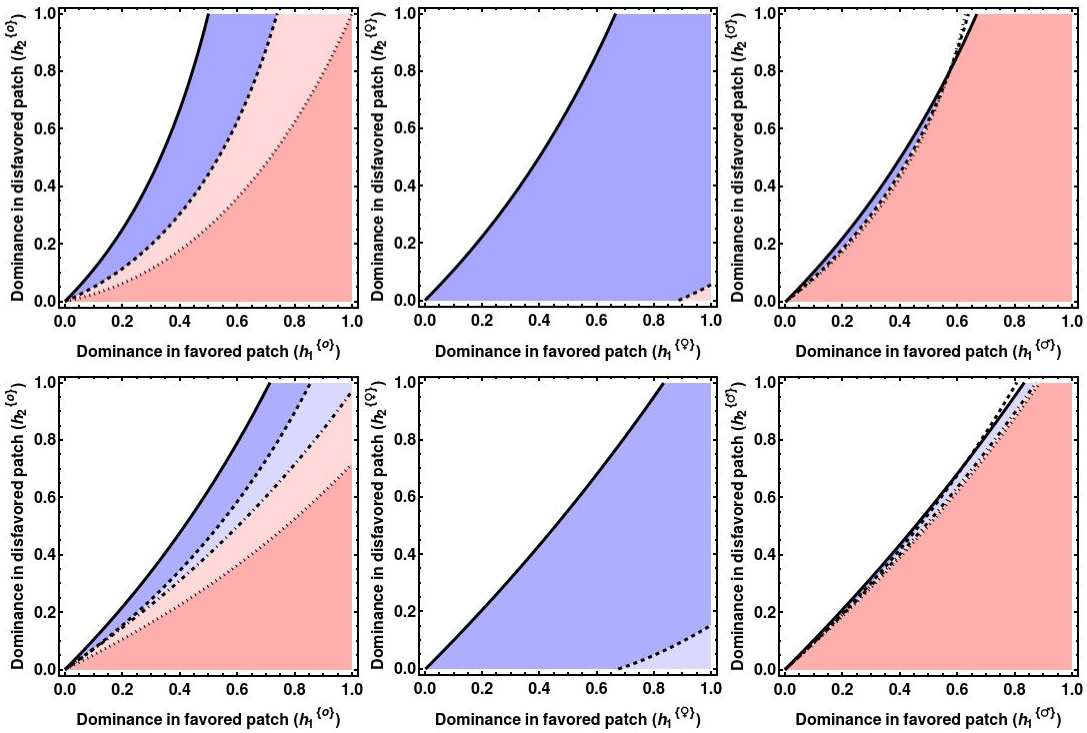


Figure A16: Zones of selfing's effect on the probability of establishment across viability (left panel), female fecundity (middle panel), and male fecundity (right panel), conditioning on allele appearing in the disfavored patch. Only seed migrates. Parameters are identical to those used in Figure 4. Upper row: scenarios with seed migration; Bottom row: scenarios with pollen migration.

- $\beta_k(\tilde{h}_2)$: Parameterize migration according to eqns. 12a–12c, and then set $F = 0$ to exclude the effect via $\tilde{M}$. Set $F = 0$ to exclude the effect via $\tilde{N}$. Let $\tilde{s}_1^{\{j\}} = s_1^{\{j\}}\tilde{h}_1^{\{j\}}$ and then set $F = 0$ to exclude the effect via $\tilde{h}_1$. Let $\tilde{s}_2^{\{j\}} = s_2^{\{j\}}\tilde{h}_2^{\{j\}}$. Take a derivative in F and evaluate at $F = 0$.

- $\beta_k(\tilde{M})$: Set $F = 0$ to exclude the effect via $\tilde{N}$. Let $\tilde{s}_i^{\{j\}} = s_i^{\{j\}}\tilde{h}_i^{\{j\}}$ for $i \in \{1, 2\}$. Then set $F = 0$ to exclude the effect via $\tilde{h}_1$, and $\tilde{h}_2$. Next, parameterize migration according to eqns. 12a–12c. Take a derivative in F and evaluate at $F = 0$.

Parameters $\tilde{h}_i^{\circ}$, $\tilde{h}_i^{♀}$, and $\tilde{h}_i^{♂}$ are given by equations (9a)–(9c). Whenever a $\beta_i(x)$ is greater than zero, a shift to selfing increases the probability that an allele becomes established, conditioning on starting in patch i. Evaluating the derivative in point other than $F = 0$ will change the results quantitatively, but not qualitatively.

G The consequences of shift to selfing on local adaptation

Where fecundity selection is examined, the establishment probability equations were parameterized with $s_i^{♀}$, and $h_i^{♀}$, while selection on male fitness component was done by parameterizing with $s_i^{♂}$, and $h_i^{♂}$. Other than that, parameters were quantitatively identical to the viability selection case. Migration rates were also kept constant across different scenarios, and only the migration type has changed.

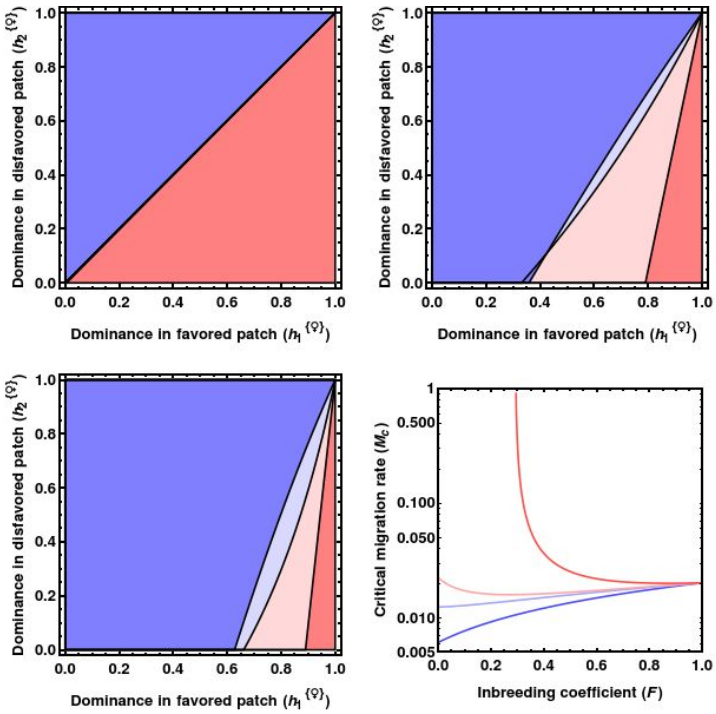


Figure A17: The consequences of a shift to selfing on establishment of local adaptation under female fecundity selection and seed dispersal. Color-coding and parameters as in the main text and parameters as in the main text.

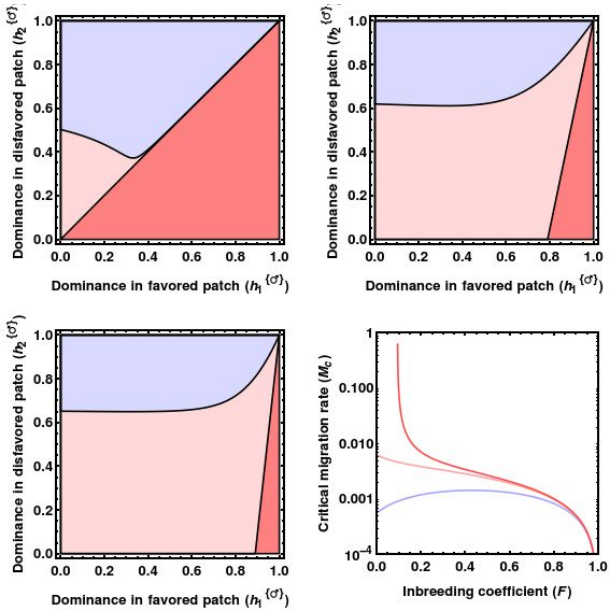


Figure A18: The consequences of a shift to selfing on establishment of local adaptation under male fecundity selection and seed dispersal. Color-coding and parameters as in the main text and parameters as in the main text.

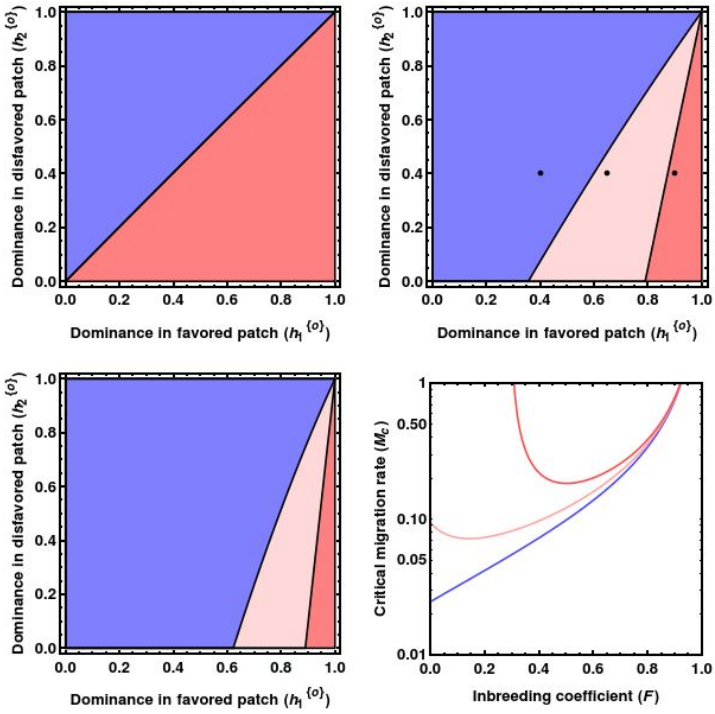


Figure A19: The consequences of a shift to selfing on establishment of local adaptation under viability selection and pollen dispersal.

H Notation glossary

For the sake of clarity, Table 1 indicates the meaning of symbols used in the derivation of the deterministic model. As a general rule, parameters with tilde signify an *effective* parameter that accounts for the effect of selfing, selection mode, and migration type on the dynamic. In other words, the selfing population with a specific parameter behaves like an outcrossing population with an effective parameter. Upper-case symbols are reserved for processes affecting the diploid phase of a life cycle, whereas the lower-case represents the process of the haploid phase.

Symbol	Meaning
$x_i(t)$	Frequency of allele A in patch i at the start of the life cycle (generation t)
$x_i^{♀}$	Frequency of allele A in patch i after female fecundity selection
$x_i^{♂}$	Frequency of allele A in patch i after male fecundity selection
$g_i^{♂}$	Frequency of allele A in patch i after pollen migration
X_i	Frequency of genotype AA at the start of the life cycle
Y_i	Frequency of genotype Aa at the start of the life cycle
X'_i	Frequency of genotype AA after mating
Y'_i	Frequency of genotype Aa after mating
X''_i	Frequency of genotype AA after seed migration
Y''_i	Frequency of genotype Aa after seed migration
N_i^*	Population size in patch i
N_i	The number of invading alleles in patch i
$\tilde{N}$	The effective population size
M_{ij}	The fraction of seed (or the probability of seed) in patch i originating from patch j
m_{ij}	The fraction of pollen (or the probability of pollen) in patch i originating from patch j
$\Delta\mathcal{W}_{0,i}$	The intergenerational A frequency change in patch i, when A rare (generic selection)
$\Delta V_{0,i}$	The intergenerational A frequency change in patch i, when A rare (only viability selection)
$\Delta W_{0,i}^{♀}$	The intergenerational A frequency change in patch i, when A rare (only female fecundity selection)
$\Delta W_{0,i}^{♂}$	The intergenerational A frequency change in patch i, when A rare (only male fecundity selection)
$V_{i,k}$	The progeny number left by genotype k ($k \in \{AA, Aa, aa\}$) that survives to maturity in patch i
$W_{i,k}^{♀}$	Number of female gametophytes produced by genotype k ($k \in \{AA, Aa, aa\}$) in patch i
$W_{i,k}^{♂}$	Number of male gametophytes produced by genotype k ($k \in \{AA, Aa, aa\}$) in patch i
$\overline{V}_i$	Mean population viability fitness in patch i
$\overline{W}_i^{♀}$	Mean population fecundity fitness in patch i
$\overline{W}_i^{♂}$	Mean population male sexual fitness in patch i
F	Equilibrium population inbreeding coefficient. Set to $S/(2-S)$
S	Fraction of the population that reproduces by selfing (i.e., the population selfing rate)
$s_i^{\{k\}}$	The relative selective advantage of invading A, with selection via k^{th} ($k \in \{V, F, M\}$) fitness component
$h_i^{\{k\}}$	The dominance coefficient of allele A, when selection acts only via k^{th} fitness component
$\tilde{M}_{ij}$	The effective fraction of seed (or the probability of seed) in patch i that originates from patch j
$\tilde{h}_i^{\{k\}}$	The effective dominance coefficient of allele A, when selection acts only via k^{th} fitness component
ω	Correction for the effective dominance coefficient for male selection and pollen dispersal
$\tilde{s}_i^{\{k\}}$	The effective selective advantage of invading allele A with selection via k^{th} fitness component

Table 1: List of symbols used in the text and their meaning.