The evolution of recombination in self-fertilizing organisms

Roman Stetsenko[*,†] and Denis Roze[*,†]

[*] CNRS, IRL 3614 Evolutionary Biology and Ecology of Algae, 29688 Roscoff, France

[†] Sorbonne Université, Station Biologique de Roscoff, 29688 Roscoff, France

Running title: Selfing and recombination

Keywords: epistasis, genetic interference, meiosis, multilocus model, mating systems

Address for correspondence:

Denis Roze

Station Biologique de Roscoff

Place Georges Teissier, CS90074

29688 Roscoff Cedex

France

Phone: (+33) 2 56 45 21 39

Fax: (+33) 2 98 29 23 24

email: roze@sb-roscoff.fr

ABSTRACT

Cytological data from flowering plants suggest that the evolution of recombination rates is affected by the mating system of organisms, as higher chiasma frequencies are often observed in self-fertilizing species compared with their outcrossing relatives. Understanding the evolutionary cause of this effect is of particular interest, as it may shed light on the selective forces favoring recombination in natural populations. While previous models showed that inbreeding may have important effects on selection for recombination, existing analytical treatments are restricted to the case of loosely linked loci and weak selfing rates, and ignore the effect of genetic interference (Hill-Robertson effect), known to be an important component of selection for recombination in randomly mating populations. In this article, we derive general expressions quantifying the stochastic and deterministic components of selection acting on a mutation affecting the genetic map length of a whole chromosome along which deleterious mutations occur, valid for arbitrary selfing rates. The results show that selfing generally increases selection for recombination caused by interference among mutations as long as selection against deleterious alleles is sufficiently weak. While interference is often the main driver of selection for recombination under tight linkage or high selfing rates, deterministic effects can play a stronger role under intermediate selfing rates and high recombination, selecting against recombination in the absence of epistasis, but favoring recombination when epistasis is negative. Individual-based simulation results indicate that our analytical model often provides accurate predictions for the strength of selection on recombination under partial selfing.

AUTHOR SUMMARY

Understanding the evolutionary advantage of sex and the associated process of meiotic recombination remains a major challenge in evolutionary biology. Useful insights on this question may possibly be gained from exploring the causes of genetic variation for recombination rates within and between species. In particular, a remarkable pattern observed in flowering plants is that self-fertilizing species tend to have higher rates of meiotic crossovers than their outcrossing relatives. In this article, we use analytical and simulation methods to investigate how selfing affects the sign and magnitude of stochastic and deterministic components of selection for recombination in hermaphroditic organisms. The results show that, while selection for recombination vanishes under complete selfing (as recombination has no genetic effect in fully homozygous individuals), it is often maximized for selfing rates slightly below one. This is primarily due to the fact that selfing increases the strength of selective interference among loci in finite populations, and may explain the higher crossover rates observed in inbred plant species.

INTRODUCTION

Genetic recombination lies at the heart of the sexual life cycle, and is often considered as one of the main evolutionary benefits of sexual reproduction [1]. However, considerable variation for the rate and position of meiotic crossovers along chromosomes exists within and between species [2–7], showing that recombination can evolve over short timescales. Rapid changes in recombination rates have also been observed during artificial selection experiments, in response to selection on recombination itself or on other traits (reviewed in [8]). Although substantial progress has been achieved over recent years in our understanding of the molecular mechanisms governing crossover formation [9, 10], the evolutionary forces acting on recombination in natural populations remain elusive. In most species, at least one crossover per bivalent seems required to ensure the proper segregation of chromosomes during meiosis I, while homologs may fail to separate when they are entangled by too many crossovers [11–13]. These mechanistic constraints probably set lower and upper bounds to the genetic map length of chromosomes, but may leave some space for evolutionary change to occur. The evolution of recombination may also be affected by indirect selective forces, stemming from the effect of recombination on genetic variation. In particular, higher recombination rates may be favored when negative linkage disequilibria (LD) between selected loci exist within populations (*i.e.*, when beneficial alleles at some loci tend to be associated with deleterious alleles at other loci), as recombination then increases the variance in fitness among offspring and the efficiency of natural selection [14, 15]. Different possible sources of negative LD have been identified, including epistatic interactions among loci [16, 17] and the Hill-Robertson effect, that tends to generate negative LD between selected loci from the random fluctuations of genotype frequencies occurring in finite populations [18–22]. Analytical and simulation models have shown that the stochastic component of selection for recombination (due to the Hill-Robertson effect) may be stronger than deterministic components generated by epistasis even when population size is rather large, especially when linkage is tight [8, 23, 24].

An interesting pattern observed in several genera of flowering plants is that self-fertilizing species tend to have higher chiasma frequencies than their outcrossing

relatives [25, 26]. Detailed comparisons between the genetic maps of the selfing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata* also point to higher recombination rates in *A. thaliana* [27–29]. A possible explanation for higher recombination rates in selfers could be that interhomolog polymorphism hinders recombination [30]. However, this hypothesis does not stand up to closer scrutiny, as available data suggest that substantial levels of divergence are needed to prevent meiotic recombination [31], while moderate levels of heterozygosity may favor rather than inhibit crossovers [32–34]. Because recombination between homozygous loci has no genetic effect, selfing reduces the efficiency of recombination in breaking LD [35, 36], and one may thus expect that increased rates of recombination could evolve to compensate for this effect. However, indirect selection for recombination should vanish under complete selfing (as heterozygosity should then be extremely rare), and the effect of selfing on selection for recombination may thus be non-monotonic. Furthermore, it is not immediately obvious that models for the evolution of recombination under random mating can be directly transposed to the case of partial selfers. Indeed, simulation models showed that recombination may be favored under different conditions in partially selfing than in outcrossing populations, while the exact mechanisms remained unclear [37–39].

A three-locus model on the effect of partial selfing on the evolution of recombination was analyzed by Roze & Lenormand [25]. The results showed that correlations in homozygosity across loci caused by partial selfing generate a selective force on recombination that is absent under random mating. By breaking correlations in homozygosity, recombination is favored when dominance-by-dominance epistasis is negative (meaning that double homozygotes have a lower fitness than expected based on the fitness of single homozygotes), as recombination increases the mean fitness of offspring produced by selfing. In the absence of dominance-by-dominance epistasis (or when it is positive), recombination is generally disfavored. The analysis also showed that even a small selfing rate has important consequences, the effect of breaking correlations in homozygosity quickly becoming the dominant source of indirect selection on recombination. However, the method used by Roze & Lenormand only holds when effective recombination rates are large, and thus breaks down when the selfing rate is not small, or when loci are tightly linked — yet tightly linked loci should be the

ones contributing most to indirect selection for recombination, unless selfing is strong. Furthermore, this model and the previous simulation models mentioned above are deterministic, considering infinite populations. From previous results on selection for recombination in finite, randomly mating populations [8, 23, 24], and from the fact that the effective size of highly selfing populations may be strongly reduced by interference effects among loci [40, 41], it seems likely that the Hill-Robertson effect should be an important component of selection for recombination in selfing organisms, but its effect remains unexplored.

In this article, we provide a general analysis of selection for recombination caused by interactions among deleterious alleles, in populations with arbitrary selfing rates. In a first step, we revisit Roze & Lenormand's deterministic three locus model, and show that linkage disequilibrium is the main source of selection for recombination in the case of tightly linked loci or under strong selfing, while the effect of correlations in homozygosity stays negligible. In a second step, we explore how the stochastic component of selection for recombination (Hill-Robertson effect) is affected by the mating system, by extending Roze's [24] finite population model to partial selfing. Last, we extrapolate the results from the stochastic and deterministic three-locus models in order to quantify the overall strength of selection acting on a modifier allele increasing the map length of a whole chromosome, and compare the predictions obtained with results from individual-based simulations. The results confirm that the Hill-Robertson effect is often the main component of selection for recombination when the selfing rate is high, usually generating stronger benefits of recombination as selfing increases, but not always. Deterministic effects may become important under intermediate selfing rates and when the chromosomal map length is sufficiently high, and may either increase or decrease selection for recombination depending on the sign and magnitude of the different components of epistasis.

METHODS

Our baseline model is the three-locus deterministic recombination modifier model with partial selfing considered by Roze & Lenormand [25], that we re-analyze

in order to obtain more accurate results for arbitrary selfing and recombination rates. In a second step, this model is extended to include the effect of random drift in finite populations. Finally, extrapolations are used to predict the overall strength of selection acting on a modifier affecting the genetic map length of a whole chromosome.

The three-locus model

Genetic architecture

The model considers two selected loci (each with two alleles, A, a at the first locus and B, b at the second) and a recombination modifier locus (with two alleles M and m). All notations used are summarized in Table 1. Alleles a and b are deleterious and affect fitness by a factor $1 - s$ in homozygotes and $1 - sh$ in heterozygotes (s and h thus correspond to the selection and dominance coefficients of alleles a and b). Deleterious alleles are generated by mutation at a rate u per generation; back mutation is ignored. Throughout the paper, we assume that h is significantly greater than zero and that $u \ll s$, so that the equilibrium frequency of deleterious alleles remains small. As in [25], epistasis between alleles a and b is decomposed into three components (see Table 2): additive-by-additive epistasis $e_{\mathrm{a\times a}}$ represents the effect of the interaction between two deleterious alleles, one at each selected locus (either on the same or on different chromosomes), while additive-by-dominance epistasis $e_{\mathrm{a\times d}}$ represents the effect of the interaction between three deleterious alleles, and dominance-by-dominance epistasis $e_{\mathrm{d\times d}}$ the effect of the interaction between four deleterious alleles. The recombination modifier locus affects the baseline recombination rate r_{ab} between the two selected loci, so that individuals with genotypes MM, Mm and mm at the modifier locus have recombination rates r_{ab}, $r_{ab} + \delta r_{ab} h_m$ and $r_{ab} + \delta r_{ab}$, respectively, with δr_{ab} the effect of the modifier and h_m the dominance coefficient of allele m. Note that because recombination only affects the genotype of meiotic products when it occurs between heterozygous loci, any effect of the modifier on recombination rates between itself and the selected loci (r_{ma}, r_{mb}) will not generate any indirect selection at the modifier locus (e.g., [17, 20]). The results given throughout the paper are valid for any ordering of the three loci along the chromosome (i.e., either $m - a - b$ or $a - m - b$).

Genetic associations

The change in frequency at the modifier locus involves different forms of genetic associations among loci, which are defined as follows. As in [42] and [43], we define indicator variables $X_{j,\varnothing}$ and $X_{\varnothing,j}$ that equal 1 if the lowercase allele at locus j (m, a or b) is present (or 0 if absent) on the maternally ($X_{j,\varnothing}$) or paternally ($X_{\varnothing,j}$) inherited gene of an individual. The average of these indicator variables over all individuals in the population gives the frequency of allele j on maternally and paternally inherited genes, $p_{j,\varnothing}$ and $p_{\varnothing,j}$, respectively. The frequency of allele j in the population, p_j, is thus given by $(p_{j,\varnothing} + p_{\varnothing,j})/2$. Centered variables $\zeta_{j,\varnothing}$ and $\zeta_{\varnothing,j}$ are defined as:

$$\zeta_{j,\varnothing} = X_{j,\varnothing} - p_{j,\varnothing}, \qquad \zeta_{\varnothing,j} = X_{\varnothing,j} - p_{\varnothing,j}. \tag{1}$$

The genetic association between the sets of loci $\mathbb{U}$ and $\mathbb{V}$ present on the maternally and paternally inherited haplotypes of the same individual is given by:

$$D_{\mathbb{U},\mathbb{V}} = \mathbb{E}[\zeta_{\mathbb{U},\mathbb{V}}], \tag{2}$$

with $\mathbb{E}$ the average over all individuals in the population, and with

$$\zeta_{\mathbb{U},\mathbb{V}} = \Big(\prod_{j\in\mathbb{U}} \zeta_{j,\varnothing}\Big)\Big(\prod_{k\in\mathbb{V}} \zeta_{\varnothing,k}\Big). \tag{3}$$

Because our model does not include and sex-of-origin effect, we always have $D_{\mathbb{U},\mathbb{V}} = D_{\mathbb{V},\mathbb{U}}$. Associations between genes on the same haplotype $D_{\mathbb{U},\varnothing} = D_{\varnothing,\mathbb{U}}$ will be denoted $D_{\mathbb{U}}$ for simplicity. For example, $D_{a,a}$ measures the excess of homozygotes (departure from Hardy-Weinberg equilibrium) at locus a, while D_{ab} is the linkage disequilibrium between alleles a and b.

Life cycle

Each generation starts by selection between newly formed diploid individuals, followed by meiosis and syngamy. The effect of selection on allele frequencies and genetic associations can be computed using the multilocus genetics framework of [43]. For this, the fitness or an individual relative to the mean fitness of the population is written as:

$$\frac{W}{\overline{W}} = 1 + \sum_{\mathbb{U},\mathbb{V}} \alpha_{\mathbb{U},\mathbb{V}} \left(\zeta_{\mathbb{U},\mathbb{V}} - D_{\mathbb{U},\mathbb{V}}\right), \tag{4}$$

where the "selection coefficients" $\alpha_{\mathbb{U},\mathbb{V}}$ represent the effect of selection acting on the sets of loci $\mathbb{U}$ and $\mathbb{V}$ present on the maternally and paternally inherited haplotypes of an individual. Since we do not assume any sex-of-origin effect we have $\alpha_{\mathbb{U},\mathbb{V}} = \alpha_{\mathbb{V},\mathbb{U}}$, while $\alpha_{\mathbb{U},\varnothing}$ will be denoted $\alpha_{\mathbb{U}}$ for simplicity. The combined effect of selection at loci a and b can thus be represented by 9 coefficients: α_a, α_b represent the net effect of selection against the deleterious alleles a and b, $\alpha_{a,a}$, $\alpha_{b,b}$ the effect of dominance at the two loci, while α_{ab}, $\alpha_{a,b}$, $\alpha_{ab,a}$, $\alpha_{ab,b}$ and $\alpha_{ab,ab}$ represent epistatic interactions, measured as deviations from additivity. Throughout the paper, we will assume that selection is weak (s is of order ϵ, where ϵ is a small term) while epistasis is weaker ($e_{\mathrm{a}\times\mathrm{a}}$, $e_{\mathrm{a}\times\mathrm{d}}$, $e_{\mathrm{d}\times\mathrm{d}}$ of order ϵ^2). Assuming that deleterious alleles stay at low frequency, and under the fitness matrix given by Table 2, $\alpha_{\mathbb{U},\mathbb{V}}$ coefficients are, to leading order:

$$\begin{aligned}
&\alpha_a = \alpha_b \approx -sh, && \alpha_{a,a} = \alpha_{b,b} \approx -s\left(1-2h\right), \\
&\alpha_{ab} = \alpha_{a,b} \approx e_{\mathrm{a}\times\mathrm{a}} + \left(sh\right)^2, && \alpha_{ab,a} = \alpha_{ab,b} \approx e_{\mathrm{a}\times\mathrm{d}} + s^2h\left(1-2h\right), \\
&\alpha_{ab,ab} \approx e_{\mathrm{d}\times\mathrm{d}} + s^2\left(1-2h\right)^2.
\end{aligned} \tag{5}$$

The effect of selection on genetic associations (in terms of $\alpha_{\mathbb{U},\mathbb{V}}$ coefficients) is given by equations 9 and 15 in [43], while the change in frequency of the modifier is given by:

$$\Delta p_m = \sum_{\mathbb{U},\mathbb{V}} \alpha_{\mathbb{U},\mathbb{V}} D_{m\mathbb{U},\mathbb{V}}. \tag{6}$$

After selection, individuals produce gametes to form the zygotes of the next generation. During syngamy, a proportion σ of fertilizations involves two gametes produced by the same parent (selfing), while a proportion $1-\sigma$ involves gametes sampled at random from the whole population (outcrossing). Recombination and syngamy do not change allele frequencies, but do change genetic associations, their effect being given by equations 13 and 14 in [25]. These equations, together with the equations describing the effect of selection on allele frequencies and genetic associations, have been implemented in a *Mathematica* notebook (available as Supplementary Material) that can be used to automatically generate recursions on these variables.

Approximations

Throughout the paper we assume that the effect of the recombination modifier (δr_{ab}) is weak and compute all results to the first order in δr_{ab}. Our three-locus diploid

model can be described by 36 genotype frequencies (leading to 35 independent variables), or alternatively by 3 allele frequencies (p_m, p_a, p_b) and 32 genetic associations. A separation of timescales argument can be used to reduce this large number of variables: in particular, when selection is weak relative to recombination, allele frequencies change slowly while genetic associations are rapidly eroded by recombination. In this case, one can show that genetic associations quickly reach a quasi-equilibrium value, which can be computed by assuming that allele frequencies remain constant [42–44]. In this quasi-linkage equilibrium (QLE) state, associations can be expressed in terms of allele frequencies and of the parameters of the model, and these expressions can then be plugged into equation 6 to obtain the change in frequency of the modifier. This is the approach used in [25] to quantify the strength of selection for recombination under partial selfing, assuming that recombination rates are sufficiently large (relative to the strength of selection) and that the selfing rate is not too large (as selfing reduces the effect of recombination). However, more accurate expressions can in principle be obtained in situations where deleterious alleles are maintained at mutation – selection balance: indeed, in this case changes in allele frequencies are only driven by the effect of the recombination modifier, and the QLE approximation thus only requires that δr_{ab} is sufficiently small relative to effective recombination rates [24, 45, 46]. This is the approach used in the present paper to obtain approximations that remain valid under low effective recombination (i.e., when recombination rates r_{ij} or when the outcrossing rate $1 - \sigma$ are of order ϵ).

The Hill-Robertson effect

An expression for the strength of selection for recombination due to the Hill-Robertson effect between two deleterious alleles in a randomly mating, diploid population was derived in [24]. Generalizing this analysis to partial selfing would be extremely tedious, as a very large number of stochastic moments of genetic associations and allele frequencies would need to be computed. However, in the case of tightly linked loci (which should be the ones contributing most to selection for recombination, unless the selfing rate is very high), scaling arguments can be used to show that the effects of selection against deleterious alleles, recombination and drift under partial selfing can be predicted by replacing the dominance coefficient h, recombination rates r_{ij}

and the population size N by the effective coefficients $h(1-F)+F$, $r_{ij}(1-F)$ and $N/(1+F)$ in the expressions obtained under random mating, where $F=\sigma/(2-\sigma)$ is the inbreeding coefficient (e.g., [40, 41, 47]). We thus introduced these effective coefficients into the expression derived in [24] in order to explore how self-fertilization affects selection for recombination generated by the Hill-Robertson effect. As we will see, comparisons with multilocus simulations indicate that this approach often yields correct results.

Multilocus extrapolation

Following [24], the three-locus analysis can be extrapolated to predict the overall strength of selection on a modifier affecting the genetic map length R of a whole chromosome (by an amount δR). For simplicity, we assume that the modifier is located at the mid-point of a linear chromosome and that the position of each crossover is sampled in a uniform distribution along the chromosome (no interference). Deleterious mutations occur at a rate U per haploid chromosome per generation, and we assume that all mutations have the same selection and dominance coefficients s and h, the position of each new mutation being sampled in a uniform distribution along the chromosome (infinite site model). Neglecting the effect of interactions between more than two mutations, the strength of indirect selection for recombination can be obtained by integrating the result from the three-locus analysis over the genetic map (see Supplementary Material). When the mean number of deleterious alleles per chromosome (n) is large and with epistasis, more accurate expressions for selection coefficients $\alpha_{\mathbb{U},\mathbb{V}}$ at loci j and k segregating for deleterious alleles are given by:

$$\begin{gathered}\alpha_j \approx -sh + 2ne_{\mathrm{a\times a}}, \quad \alpha_{j,j} \approx -s(1-2h) + 2ne_{\mathrm{a\times d}}, \\ \alpha_{jk} = \alpha_{j,k} \approx e_{\mathrm{a\times a}} + \alpha_j\alpha_k, \quad \alpha_{jk,k} \approx e_{\mathrm{a\times d}} + \alpha_j\alpha_{k,k}, \quad \alpha_{jk,jk} \approx e_{\mathrm{a\times d}} + \alpha_{j,j}\alpha_{k,k}\end{gathered} \tag{7}$$

(see Supplementary Material). An approximate expression for the mean number of deleterious alleles per chromosome n (taking into account the effects of epistasis and selfing) can be obtained from equation 27 in [48]. We also computed the overall strength of selection for recombination under the diploid model of synergistic epistasis considered by Charlesworth et al [49], in which the fitness of individuals is given by $W=\exp\left[-\left(\alpha\tilde{n}+\beta\tilde{n}^2/2\right)\right]$, with $\tilde{n}=hn_{\mathrm{he}}+n_{\mathrm{ho}}$ and where n_{he} and n_{ho} are the numbers

of heterozygous and homozygous mutations present in the genome of the individual. As shown in [48], this is equivalent to setting $e_{\mathrm{a\times a}} = -\beta\, h^2$, $e_{\mathrm{a\times d}} = -\beta\, h\,(1-2h)$ and $e_{\mathrm{d\times d}} = -\beta\,(1-2h)^2$ in the present model.

Individual-based simulations

Our simulation program (written in C++ and available from Zenodo) is equivalent to the program used in [24], to which partial selfing and the different forms of epistasis are added. It represents a population of N diploids, each carrying a pair of linear chromosomes. At each generation, the number of new deleterious mutations per chromosome is drawn from a Poisson distribution with parameter U, and their position along the chromosome is drawn from a uniform distribution between 0 and 1. The fitness of each individual is computed as:

$$W = W_c\,(1-sh)^{n_{\mathrm{he}}}\,(1-s)^{n_{\mathrm{ho}}}\,(1+e_{\mathrm{a\times a}})^{n_2}\,(1+e_{\mathrm{a\times d}})^{n_3}\,(1+e_{\mathrm{d\times d}})^{n_4} \tag{8}$$

where n_{he} and n_{ho} are the number of heterozygous and homozygous deleterious mutations in the genome of the individual, and where n_2, n_3 and n_4 are the number of interactions between 2, 3 and 4 deleterious alleles at two loci, given by:

$$\begin{aligned} n_2 &= \frac{1}{2} n_{\mathrm{he}}\,(n_{\mathrm{he}}-1) + 2n_{\mathrm{he}}n_{\mathrm{ho}} + 2n_{\mathrm{ho}}\,(n_{\mathrm{ho}}-1)\,, \\ n_3 &= n_{\mathrm{he}}n_{\mathrm{ho}} + 2n_{\mathrm{ho}}\,(n_{\mathrm{ho}}-1)\,, \\ n_4 &= \frac{1}{2} n_{\mathrm{ho}}\,(n_{\mathrm{ho}}-1) \end{aligned} \tag{9}$$

(e.g., [48]). The term W_c corresponds to a direct fitness effect of the chromosome map length R; as in [24], it is generally set to $W_c = \exp(-cR)$, so that c measures a direct fitness cost per crossover. This ensures that R does not evolve towards very large values and allows simple comparisons with analytical predictions. Direct stabilizing selection around one crossover per bivalent (i.e., $R = 0.5$) was also implemented by setting $W_c = \exp\left[-c\,(R-0.5)^2\right]$.

To produce individuals of the next generation, parents are sampled according to their fitness, each new individual being produced by selfing with probability σ. The recombination modifier locus is located at the midpoint of the chromosome, with an infinite number of possible alleles coding for different values of R (the map length of

the individual being given by the average of its two modifier alleles). During the first 20,000 generations R is fixed to 1 in order to reach mutation–selection balance for deleterious alleles. Then, for an extra 5×10^6 generations (increased up to 5×10^7 generations for high values of σ), mutations are introduced at a rate 10^{-4} at the modifier locus, each mutation multiplying the value of R by a random number drawn from a Gaussian distribution with mean 1 and variance 0.04. To allow for large effect mutations, a proportion 0.05 of mutations have an additive effect on R drawn from a uniform distribution between -1 and 1 (the new value being set to zero if it is negative). The average map length $\overline{R}$, average fitness $\overline{W}$, average number of deleterious mutations per chromosome n and number of fixed mutations are recorded every 500 generations (fixed mutations are removed from the population in order to reduce execution speed). The equilibrium value of map length is computed as the time average of $\overline{R}$, after removing the first 5×10^5 generations to allow $\overline{R}$ to equilibrate. In order to study the effect of variable selection coefficients of deleterious mutations, we modified the above baseline simulation program so that each new mutation is associated with a value of s drawn from a log-normal distribution (ensuring $s > 0$): the value of $\ln(s)$ is drawn from a Gaussian distribution with variance sd^2 and mean $\ln(\bar{s}) - sd^2/2$ (so that the average selection coefficient $\bar{s}$ stays constant for different values of sd).

RESULTS

The deterministic three-locus model

We first reiterate the result obtained by [25] under high effective recombination in a more general version, and then provide a new analysis for the case of weak effective recombination. All derivations are provided in the *Mathematica* notebook available as Supplementary Material.

High effective recombination

As found by [25], the joint effects of selfing and the recombination modifier generate an association $D_{mab,ab}$ even in the absence of selection. This stems from

the fact that recombination tends to break correlations in homozygosity between loci generated by partial selfing (see also [50]). The association $D_{mab,ab}$ has the sign of $-\delta r_{ab}$, reflecting the fact that the modifier allele increasing recombination tends to be found on genetic backgrounds in which the correlation in homozygosity between loci a and b is relatively weaker. At QLE and to the first order in δr_{ab}, it is given by:

$$D^0_{mab,ab} \approx -\frac{2\delta r_{ab}\left(1-2r_{ab}\right)F\left(1-F\right)\left[H_m\left(1+2F\gamma_{ab}\right)+F\left(\gamma_{ma}+\gamma_{mb}-\gamma_{ab}\right)\right]}{\left[1+F\left(\gamma_{ma}+\gamma_{mb}+\gamma_{ab}\right)\right]\left(1+2F\gamma_{ab}\right)^2}\,pq_{mab}, \tag{10}$$

where the superscript "0" indicates that the effect of selection at loci a and b is neglected in this expression, and where $H_m = h_m + p_m(1-2h_m)$, $\gamma_{ij} = r_{ij}(1-r_{ij})$ and $pq_{mab} = p_m q_m p_a q_a p_b q_b$ (with $q_j = 1-p_j$). Equation 10 generalizes equation 32 in [25] to arbitrary h_m and any ordering of the three loci along the chromosome. It shows that $D^0_{mab,ab}$ vanishes under random mating ($F = 0$) and full selfing ($F = 1$), because in these cases the mating system does not generate any correlation in homozygosity among loci.

The other associations appearing in equation 6 are generated by selection against the deleterious alleles (of order ϵ) and by the modifier effect. At QLE, associations $D_{mi,i}$, $D_{mij,i}$ and $D_{mi,ij}$ (where i, j are either a or b) are of order ϵ, while associations D_{mi}, $D_{m,i}$, D_{mij}, $D_{m,ij}$ and $D_{mi,j}$ are of order ϵ^2 (expressions for these associations are given in Appendix A). As a result, one obtains from equations 5 and 6 that the change in frequency of the modifier is, to leading order:

$$\Delta p_m \approx \alpha_{a,a} D_{ma,a} + \alpha_{b,b} D_{mb,b} + \alpha_{ab,ab} D^0_{mab,ab}, \tag{11}$$

the association $D_{ma,a}$ being given by:

$$D_{ma,a} \approx \alpha_{b,b} F D^0_{mab,ab}, \tag{12}$$

(and symmetrically for $D_{mb,b}$). $D_{ma,a}$ is positive when $\delta r_{ab} > 0$ and $h < 1/2$ (since $\alpha_{b,b}$ and $D^0_{mab,ab}$ are both negative in this case), reflecting the fact that the modifier allele m increasing recombination tends to be found on more homozygous backgrounds at locus a. Indeed when $h < 1/2$, homozygous genotypes at loci a and b have, on average, a lower fitness than heterozygous genotypes, and the correlation in homozygosity generated by partial selfing increases the efficiency of selection, lowering the frequency of homozygous genotypes. As the modifier allele increasing recombination tends to break this correlation, it is associated to a relative excess in homozygosity at each

selected locus. This effect disfavors recombination (since homozygotes have a lower fitness than heterozygotes on average), which is reflected by the first two terms of equation 11. Equations 11 and 12 give:

$$\Delta p_m \approx \left(2\alpha_{a,a}\alpha_{b,b}F + \alpha_{ab,ab}\right) D^0_{mab,ab}, \tag{13}$$

which, using equation 5, becomes:

$$\Delta p_m \approx \left[s^2\left(1-2h\right)^2\left(1+2F\right) + e_{\mathrm{d}\times\mathrm{d}}\right] D^0_{mab,ab}, \tag{14}$$

yielding Equation 36 in [25]. Equations 10 and 14 show that increased recombination is favored when dominance-by-dominance epistasis ($e_{\mathrm{d}\times\mathrm{d}}$) is sufficiently negative. Indeed, in this case, breaking correlations in homozygosity (thus increasing the frequency of genotypes that are homozygous at one selected locus and heterozygous at the other) tends to increase the mean fitness of offspring.

Low effective recombination

In the previous analysis, terms in $r_{mi}\left(1-F\right)$ appear in the denominators of the expressions for D_{mi} and $D_{m,i}$ at QLE (see Appendix A), causing these expressions to diverge (i.e., tend to infinity) as effective recombination rates $r_{mi}\left(1-F\right)$ tends to zero. Similarly, terms in $r_{mab}\left(1-F\right)$ appear in the denominators of D_{mab}, $D_{m,ab}$, $D_{ma,b}$ and $D_{mb,a}$, where $r_{mab} = \left(r_{ma}+r_{mb}+r_{ab}\right)/2$ is the probability that at least one recombination event occurs between the three loci. This indicates that these associations should play a more important role in the case of tightly linked loci. In order to explore this regime, we re-analyzed the model in the case where all recombination rates are of order ϵ (see Supplementary Material). This analysis shows that the effect of the linkage disequilibrium D_{ab} between deleterious alleles (which was negligible under high effective recombination) becomes predominant when loci are tightly linked. A general expression for D_{ab} at equilibrium in terms of $\alpha_{\mathbb{U},\mathbb{V}}$ coefficients is given by:

$$D_{ab} \approx \frac{\tilde{\alpha}_{ab} - \tilde{\alpha}_a\tilde{\alpha}_b + 2FG_{ab}\left(\alpha_a+\alpha_{a,a}\right)\left(\alpha_b+\alpha_{b,b}\right)}{\tilde{r}_{ab} - \tilde{\alpha}_a - \tilde{\alpha}_b}\, pq_{ab}, \tag{15}$$

where the tilde denotes "effective coefficients": $\tilde{r}_{ab} = r_{ab}\left(1-F\right)$, $\tilde{\alpha}_a = \alpha_a\left(1+F\right) + \alpha_{a,a}F$ (and similarly for $\tilde{\alpha}_b$), while $\tilde{\alpha}_{ab} = \alpha_{ab}\left(1+\phi_{ab}\right) + 2\alpha_{a,b}F + \left(\alpha_{ab,a}+\alpha_{ab,b}\right)\left(F+\phi_{ab}\right) + \alpha_{ab,ab}\phi_{ab}$, where ϕ_{ab} is the probability of joint identity-by-descent at the two loci (which is approximately F when r_{ab} is small, see Appendix A). Finally, G_{ab} in equation 15

refers to the identity disequilibrium between the two loci, defined as $G_{ab} = \phi_{ab} - F^2$ [51], and thus approximately equal to $F(1-F)$ under tight linkage. Equation 15 simplifies to $(\alpha_{ab} - \alpha_a\alpha_b)\, pq_{ab} / (r_{ab} - \alpha_a - \alpha_b)$ in the absence of selfing; this is equivalent to the result obtained by [17] under strong recombination (equation 9b in [17]), except that α_a and α_b now appear in the denominator, due to our assumption that r_{ab} is small (of order ϵ). With selfing, two important differences appear: (i): the recombination rate and selection coefficients are replaced by effective coefficients $\tilde{r}_{ab}$, $\tilde{\alpha}_a$, $\tilde{\alpha}_b$, $\tilde{\alpha}_{ab}$ (since increased homozygosity affects both the effects of recombination and of selection, as explained in Appendix B), and (ii): an extra term, involving the identity disequilibrium G_{ab}, appears in the numerator and tends to generate positive linkage disequilibrium between deleterious alleles.

Using the expressions for $\alpha_{\mathbb{U},\mathbb{V}}$ coefficients given by equation 5, equation 15 simplifies to:

$$D_{ab} \approx \frac{s^2 (1-h)^2 (1+2F)\, G_{ab} + \tilde{e}}{\tilde{r}_{ab} + 2s\tilde{h}}\, pq_{ab}, \tag{16}$$

with $\tilde{e} = e_{\mathrm{a\times a}}(1 + 2F + \phi_{ab}) + 2e_{\mathrm{a\times d}}(F + \phi_{ab}) + e_{\mathrm{d\times d}}\phi_{ab}$ and $\tilde{h} = h(1-F) + F$. The fact that partial selfing generates positive linkage disequilibrium between deleterious alleles in the absence of epistasis has been noticed in previous analytical and simulation studies [25, 52], and may be understood as follows. Lineages with different histories of inbreeding coexist in a partially selfing population: lineages that have been inbred for many generations tend to be very homozygous (at all loci), while lineages that have been inbred for fewer generations tend to be less homozygous. This is the basis of correlations in homozygosity among loci, represented by the identity disequilibrium G_{ab}. Because homozygosity increases the efficiency of selection, the frequency of deleterious alleles tends to be lower within lineages that have been inbred for longer (purging), and higher in less inbred lineages, resulting in positive linkage disequilibrium between deleterious alleles. As shown by equation 16, D_{ab} may become negative when epistasis is negative and sufficiently strong, the relative importance of $e_{\mathrm{a\times d}}$ and $e_{\mathrm{d\times d}}$ increasing as the selfing rate increases.

The fact that the allele coding for higher recombination tends to erode D_{ab} more rapidly in turn generates genetic associations between the modifier and the selected

loci. In particular, we have:

$$D_{mab} \approx -\frac{\delta\tilde{r}_{ab} H_m D_{ab}}{\tilde{r}_{mab} - \tilde{\alpha}_a - \tilde{\alpha}_b} pq_m, \qquad D_{ma} \approx \frac{\tilde{\alpha}_b D_{mab}}{\tilde{r}_{ma} - \tilde{\alpha}_a} \tag{17}$$

with $\delta\tilde{r}_{ab} = \delta r_{ab}(1-F)$, $\tilde{r}_{mab} = r_{mab}(1-F)$ (and again $\tilde{\alpha}_a = \tilde{\alpha}_b \approx -s\tilde{h}$), D_{mb} being given by a symmetric expression. These equations take the same form as under random mating (e.g., equations 9c and 11 in [17]) and are interpreted in the same way: when $D_{ab} > 0$, AB and ab haplotypes are in excess in the population, and the allele increasing recombination (allele m if $\delta r_{ab} > 0$) tends to reduce this excess, and thus becomes more associated with Ab, aB haplotypes (which is reflected by a negative value of D_{mab}). By contrast, when $D_{ab} < 0$, the allele increasing recombination becomes more associated with ab, AB haplotypes (which is reflected by a positive value of D_{mab}). When $D_{mab} > 0$, selection is more efficient in the m background (because the frequency of extreme haplotypes AB and ab is higher in this background), and m thus becomes better purged from deleterious alleles (which is reflected by negative values of D_{ma}, D_{mb}). When $D_{mab} < 0$, selection is less efficient in the m background, leading to positive values of D_{ma}, D_{mb} (deleterious alleles are more frequent in the m background).

Under weak recombination, separation of timescale arguments can be used to express the other associations that appear in equation 6 in terms of D_{mab}, D_{ma} and D_{mb} (e.g., [41]); one obtains:

$$\begin{gathered} D_{mab,ab} \approx D_{mab,a} \approx D_{ma,ab} \approx D_{m,ab} \approx D_{ma,b} \approx F D_{mab}, \\ D_{m,a} \approx D_{ma,a} \approx F D_{ma}, \qquad D_{m,b} \approx D_{mb,b} \approx F D_{mb}. \end{gathered} \tag{18}$$

Plugging these into equation 6 yields:

$$\Delta p_m \approx \tilde{\alpha}_a D_{ma} + \tilde{\alpha}_b D_{mb} + \tilde{\alpha}_{ab} D_{mab}, \tag{19}$$

which again takes the same form as under random mating (equation 8 in [17]). From equations 17 and 19, one obtains:

$$\Delta p_m \approx -\frac{\delta\tilde{r}_{ab} H_m D_{ab} pq_m}{\tilde{r}_{mab} - \tilde{\alpha}_a - \tilde{\alpha}_b} \left[\tilde{\alpha}_{ab} + \tilde{\alpha}_a\tilde{\alpha}_b \left(\frac{1}{\tilde{r}_{ma} - \tilde{\alpha}_a} + \frac{1}{\tilde{r}_{mb} - \tilde{\alpha}_b}\right)\right], \tag{20}$$

equivalent to equation 12 in [17]. Given that the equilibrium frequency of deleterious alleles is approximately $u/(s\tilde{h})$, thus of order u/ϵ, equations 16 and 20 show that

under weak effective recombination (i.e., when $\tilde{r}$ coefficients are of order ϵ), selection for recombination generated by the linkage disequilibrium D_{ab} is of order $\delta r_{ab} u^2/\epsilon$, thus stronger than selection for recombination generated by the term in $D^0_{mab,ab}$ seen in the previous subsection (of order $\delta r_{ab} u^2$, see equation 14). When $\tilde{r}$ coefficients are not small (of order 1), however, indirect selection generated by D_{ab} becomes of order $\delta r_{ab} u^2 \epsilon^2$ (and thus negligible). A general expression for the change in frequency of the modifier, valid under both weak and strong effective recombination, can thus be obtained by summing equations 14 and 20. We noticed that in some cases, more accurate expressions can be obtained by taking into account indirect effects of $D^0_{mab,ab}$ on other genetic associations between the modifier and selected loci (these terms should be negligible under both weak and strong effective recombination, but improve the precision of our approximations for intermediate values of recombination rates); these expressions are given in Appendix C. Two further improvements to our approximations are done in Appendix C, leading to more accurate expressions when epistasis is of the same order of magnitude as the strength of selection ($e_{\mathrm{a\times a}}$, $e_{\mathrm{a\times d}}$, $e_{\mathrm{d\times d}}$ of order ϵ) and when the selfing rate is high (see Supplementary Material for derivations).

Figures 1 and S1 show that our approximations provide correct predictions for the change in frequency at the modifier locus, for all values of the selfing rate between 0 and 1. The dots in these figures correspond to the results of deterministic simulations, obtained by iterating exact recurrence equations for the 36 genotype frequencies: allele M is fixed during the first 3000 generations to reach mutation–selection balance at the selected loci, then allele m is introduced in frequency 0.01 and the population is let to evolve for an extra 1000 generations, $\Delta p_m/pq_{mab}$ being averaged over the last 500 generations (see Supplementary Material). As can be seen on Figure 1, indirect selection is mostly driven by D_{ab} when recombination rates are small (left figures) or when the selfing rate approaches 1. In the absence of epistasis, the identity disequilibrium generates positive D_{ab} which disfavors recombination (as breaking positive D_{ab} reduces the variance in fitness, see Figures 1A, 1B). Negative epistasis generates negative D_{ab}, which favors increased recombination when effective recombination rates are sufficiently small (equation 20, Figures 1C-F). The relative effect of the term in $D^0_{mab,ab}$ becomes more important in the case of loosely linked loci and for intermediate selfing rates, either when epistasis is absent (in which case it disfavors recombination, see

equation 14, Figure 1B), or when the dominance-by-dominance component of epistasis is important (in which case it favors increased recombination when $e_{\mathrm{d\times d}} < 0$, Figure 1F). Finally, selection on recombination vanishes under complete selfing ($\sigma = 1$), as recombination is ineffective in that case.

The Hill-Robertson effect

An expression for the strength of selection for recombination caused by the Hill-Robertson effect between two deleterious alleles in a diploid, randomly mating population was derived by Roze [24]. This expression consists in a sum of terms corresponding to different mechanisms generating selection for recombination, that all take the form $\delta r_{ab}\,(sh)^2\,u^2 / \left[N_{\mathrm{e}} \prod_{i=1}^{5} (r_{\mathbb{U}_i} + x_i sh)\right]$ where N_{e} is the effective population size, $\mathbb{U}_i$ is either mab, ma, mb or ab, and x_i equals 1, 2, 3 or 4. In the case of loosely linked loci ($r_{\mathbb{U}_i} \gg sh$), the terms in sh in the denominator may be neglected and the strength of selection for recombination increases as sh increases — being roughly proportional to $(sh)^2$. In the case of tightly linked loci ($r_{\mathbb{U}_i} \ll sh$), however, recombination rates $r_{\mathbb{U}_i}$ may be neglected and selection for recombination now decreases as sh increases — being roughly proportional to $1/\,(sh)^3$.

In order to extend these results to partially selfing populations, we replaced δr_{ab}, $r_{\mathbb{U}_i}$ and h by the effective parameters $\delta\tilde{r}_{ab} = \delta r_{ab}\,(1-F)$, $\tilde{r}_{\mathbb{U}_i} = r_{\mathbb{U}_i}\,(1-F)$ and $\tilde{h} = h\,(1-F) + F$ in the expression derived in [24]. The effect of increasing selfing (thus increasing F) on the strength of selection for recombination due to the Hill-Robertson effect can be understood using the same reasoning as above. When effective recombination rates are large relative to the strength of selection against deleterious alleles (selfing rate not too large, $r_{\mathbb{U}_i}$ not too small so that $\tilde{r}_{\mathbb{U}_i} \gg s\tilde{h}$), selection for recombination becomes approximately proportional to $\delta\tilde{r}_{ab}(s\tilde{h})^2 / \left[N_{\mathrm{e}} \prod_{i=1}^{5} \tilde{r}_{\mathbb{U}_i}\right] = \delta r_{ab}(s\tilde{h})^2 / \left[N_{\mathrm{e}}\,(1-F)^4 \prod_{i=1}^{5} r_{\mathbb{U}_i}\right]$, which increases as F increases (mostly due to the decreased effective recombination rates $\tilde{r}_{\mathbb{U}_i}$ at the denominator, but also to the increase in $\tilde{h}$). When effective recombination rates are small relative to the strength of selection (high selfing and/or tightly linked loci, so that $\tilde{r}_{\mathbb{U}_i} \ll s\tilde{h}$), selection for recombination is approximately proportional to $\delta\tilde{r}_{ab}/[N_{\mathrm{e}}(s\tilde{h})^3] = \delta r_{ab}\,(1-F)\,/[N_{\mathrm{e}}(s\tilde{h})^3]$, which decreases as F increases (mostly due to the decreased effect of the modifier $\delta\tilde{r}_{ab}$, but

also to the increase in $\tilde{h}$). One thus predicts that increasing selfing should generally have a non-monotonic effect on selection for recombination due to the Hill-Robertson effect (selection for recombination first increasing, and then decreasing towards zero as F increases from 0 to 1), the position of the maximum depending on sh and on the values of recombination rates (however, selfing may always decrease selection for recombination when sh is sufficiently strong and/or linkage sufficiently tight). These verbal predictions are illustrated by Figure 2. Note that selfing also has the additional effect of reducing N_e by a factor $1/(1+F)$ [35, 53], thus increasing the strength of the Hill-Robertson effect, but this effect stays minor relative to the effect of selfing on effective recombination rates. Selfing may cause stronger reductions in N_e when deleterious alleles are segregating at many linked loci, however, through background selection effects [40, 41].

These results also provide us with some understanding of the relative importance of stochastic and deterministic sources of selection for recombination: when effective recombination rates are large, indirect selection due to the Hill-Robertson effect is of order $\delta r_{ab} u^2 \epsilon^2 / N_\mathrm{e}$, and should thus be negligible relative to the deterministic component (of order $\delta r_{ab} u^2$). When effective recombination rates are small (of order ϵ), however, selection due to the Hill-Robertson effect is now of order $\delta r_{ab} u^2 / (N_\mathrm{e} \epsilon^3)$, and may thus become of the same order of magnitude or stronger than deterministic terms (which are then of order $\delta r_{ab} u^2 / \epsilon$, as shown in the previous subsection).

Multilocus extrapolation

The three-locus analysis can be extended to compute the overall strength of indirect selection acting on a modifier affecting the genetic map length of a whole chromosome. Neglecting the effect of interactions between deleterious alleles at more than two loci, this can be done by integrating the result from the three-locus model over all possible positions of alleles a and b. As in [24], we consider a linear chromosome with map length R (in Morgans), along which deleterious mutations occur at a rate U per chromosome per generation. The modifier is located at the mid-point of the chromosome, allele m increasing map length by an amount $\delta R/2$ when heterozygous and δR when homozygous (see Methods). The strength of selection for allele m, defined

as $s_m = \Delta p_m / \left(\frac{1}{2} p_m q_m\right)$, can be decomposed into three terms:

$$s_m = s_{\text{direct}} + s_{\text{det}} + s_{\text{HR}} \tag{21}$$

where s_{direct} corresponds to the effect of direct selective pressures acting on R (due to any direct effect of R on the fitness of individuals), s_{det} to indirect selection caused by deterministic interactions between deleterious alleles, and s_{HR} to indirect selection caused by the Hill-Robertson effect. Assuming a direct fitness cost of crossovers so that fitness decreases as e^{-cR} as R increases, the direct selection term is given by $s_{\text{direct}} = -c\,\delta R\,(1+F)$ to the first order in δR (e.g., [46]). From the analysis above, the term s_{det} can be further decomposed into $s_{D_{ab}} + s_{D^0_{mab,ab}}$, where $s_{D_{ab}}$ corresponds to the overall effect of deterministically generated linkage disequilibria (D_{ab}) between deleterious alleles (either by epistasis or by identity disequilibria), given by equations 15 and 20, while $s_{D^0_{mab,ab}}$ corresponds to the overall effect of associations of the form $D_{mab,ab}$ generated by correlations in homozygosity and by the modifier effect, given by equations 10 and 13. Because $s_{D_{ab}}$ should be mostly driven by tightly linked loci, recombination rates are approximated by genetic distances in equations 15 and 20 (before integrating over the genetic map), while δr_{ab} is approximated by $\delta R\, r_{ab}/R$ [24]. By contrast, loosely linked loci can make a stronger contribution to $s_{D^0_{mab,ab}}$, and the recombination rate between loci i and j in equation 10 is thus expressed in terms of the genetic distance d_{ij} between these loci using Haldane's mapping function $r_{ij} = \left[1 - e^{-2d_{ij}}\right]/2$ [54], while δr_{ab} is approximated by $\delta R\, e^{-2d_{ab}}\, d_{ab}/R$ (see Supplementary Material). Note that when U is not small (so that the mean number of deleterious alleles per chromosome n may be large), and with epistasis, the expressions for $\alpha_{\mathbb{U},\mathbb{V}}$ coefficients given by equation 7 must be used instead of equation 5.

Concerning the integration of the term generated by the Hill-Robertson effect, [24] showed that using scaled recombination rates $\rho_{\mathbb{U}} = r_{\mathbb{U}}/(sh)$ eliminates sh from the integrand, sh only appearing in the integration limits, given by $R/(2sh)$. The same method can be used when $r_{\mathbb{U}}$, δr_{ab} and h are changed to effective parameters in order to incorporate the effect of selfing. In particular, defining scaled recombination rates $\tilde{\rho}_{\mathbb{U}} = \tilde{r}_{\mathbb{U}}/(s\tilde{h})$, one obtains that the overall strength of indirect selection generated by the Hill-Robertson effect is given by the same expression as under random mating (equation 3 in [24]), except that the integration limits become $R\,(1-F)/(2s\tilde{h})$ and

that the whole expression is multiplied by a factor $1/\left(1-F\right)^2$. When $R\left(1-F\right) \gg s\tilde{h}$ (which implies that the selfing rate is not too large and s is sufficiently small), the integral can be approximated by the same integral taken between 0 and infinity, which is ≈ 1.8 [24], yielding:

$$s_{\mathrm{HR}} \approx \frac{1.8\,\delta R\,U^2}{N_{\mathrm{e}}\,R^3\left(1-F\right)^2}. \tag{22}$$

Equation 22 is equivalent to equation 1 in [24], replacing R by $\tilde{R} = R\left(1-F\right)$ and δR by $\delta\tilde{R} = \delta R\left(1-F\right)$. Note that this approximation is not valid when F approaches 1, however, as $R\left(1-F\right) \gg s\tilde{h}$ cannot hold in this case. Finally, the effective population size N_{e} may be significantly lowered by background selection effects when U is not small. Assuming that the reduction in N_{e} is mostly due to tightly linked loci, one obtains that $N_{\mathrm{e}} \approx [N/(1+F)]\exp[-2U/(\tilde{R}+2s\tilde{h})]$ (e.g., [41, 55]), where N is the census population size.

The evolutionarily stable map length (R_{ES}) corresponds to the value of R for which direct and indirect selection balance each other, that is, $s_m = 0$. Figure 3 shows R_{ES} as a function of the selfing rate σ, for different values of the direct cost of recombination c. The curves correspond to the analytical predictions, obtained using *Mathematica* by numerically integrating the results of the three-locus model over the genetic map to obtain s_{det} and s_{HR} for a range of values of R, and finding the value of R for which $s_m = 0$ by interpolation (see Supplementary Material). Note that the more precise approximations given in Appendix C have been used to compute s_{det}, but using equations 10, 13, 15 and 20 often yields similar results. Figure 3 shows that the ES map length generally increases as the selfing rate increases, due to stronger indirect selection caused by the Hill-Robertson effect. While indirect selection vanishes under full selfing ($\sigma = 1$), the analytical model predicts that the maximum map length is reached for values of σ very close to 1, in particular when $c = 10^{-3}$ and $c = 10^{-4}$. In the absence of epistasis, the deterministic component of indirect selection selects against recombination ($s_{\mathrm{det}} < 0$). Figure 3 shows that this deterministic component stays negligible relative to the Hill-Robertson effect for parameter values leading to low R_{ES} ($c = 0.01$), while its relative effect becomes more important for parameter values leading to higher R_{ES} ($c = 10^{-4}$). This agrees with the prediction that the Hill-Robertson effect becomes stronger than deterministic effects when effective recombination rates

are sufficiently small. In the absence of epistasis, we generally found that s_{det} is mainly driven by $s_{D_{ab}}$, the effect of $s_{D^0_{mab,ab}}$ staying negligible (not shown). Figure 3 also shows that extrapolations from our three-locus model often provide accurate predictions of the evolutionarily stable map length observed in the simulations, discrepancies appearing for high values of the selfing rate. These discrepancies may be due to the fact that using effective recombination coefficients to transpose the result obtained under random mating to the case of a partially selfing population (as we did to compute s_{HR}) should strictly only hold when loci are sufficiently tightly linked (e.g., [41, 56]), while loosely linked loci may significantly contribute to selection for recombination when the selfing rate is high.

Figure 4 shows the effects of the deleterious mutation rate U, population size N, selection and dominance coefficients of deleterious alleles (s, h) on the evolutionarily stable map length (Figure S2 shows the same results with R on a log scale). As predicted, increasing U and/or decreasing N leads to stronger effects of Hill-Robertson interference between deleterious alleles, favoring higher values of R. As can be seen on Figure 4A, our analytical approximations overestimate the ES map length when U is high: this is probably due to the effect of higher-order associations (involving more than two selected loci) that are not taken into account in the analysis. Furthermore, in some simulations with $U = 0.5$, R fell down to a very small value at some point during the simulation, which led to a quick accumulation of heterozygous mutations, the only surviving individuals being heterozygous for haplotypes carrying different deleterious alleles in repulsion (pseudo-overdominance, e.g., [57–59]). This occurred for $0.5 \leq \sigma \leq 0.7$, in which case the simulation had to be stopped manually, the number of segregating mutations quickly becoming very large. Figure 4C shows that selection for recombination is stronger when deleterious alleles are more weakly selected, as already found by [24]; furthermore, above a given value of s, the equilibrium map length decreases as the selfing rate increases (which can be understood from the reasoning given in the previous subsection and Figure 2). By contrast, the dominance coefficient h of deleterious alleles has only little effect on the ES map length (Figure 4D). Figures S3 and S4 show that the results are not significantly affected by introducing variability in the selection coefficients of deleterious alleles into the simulation program, nor by limiting to 100 or 1000 the number of loci at which deleterious

605 mutations may occur. Figure S5 shows that the strength of selection for recombination due to the Hill-Robertson effect is accurately predicted by equation 22 when selection against deleterious alleles is sufficiently weak, as long as the selfing rate is not too high.

Figure 5 shows the effect of each form of diploid epistasis ($e_{\mathrm{a\times a}}$, $e_{\mathrm{a\times d}}$, $e_{\mathrm{d\times d}}$) and of Charlesworth et al's [49] synergistic epistasis coefficient β (that combines the 610 three forms of epistasis, see Methods) on the ES map length. Note that only negative values of epistasis were considered, as combinations of deleterious mutations quickly become beneficial when epistasis is positive and U is not very small. Negative epistasis tends to lower the strength of the Hill-Robertson effect (dotted curves on Figure 5), by increasing the effective strength of selection against deleterious alleles. Despite that, 615 the deterministic effects generated by negative epistasis increase the overall strength of selection for recombination. Figure S6 shows that this increase is driven by the term in D_{ab} in the case of negative additive-by-additive ($e_{\mathrm{a\times a}}$) and additive-by-dominance ($e_{\mathrm{a\times d}}$) epistasis, the term in $D^0_{mab,ab}$ being negligible when $e_{\mathrm{a\times a}} < 0$, while it disfavors recombination when $e_{\mathrm{a\times d}} < 0$ (one can show that this last effect is generated by the first 620 term of equation 13, since effective dominance coefficients $a_{j,j}$ are increased by $e_{\mathrm{a\times d}}$, as shown by equation 7). In the case of dominance-by-dominance epistasis ($e_{\mathrm{d\times d}}$), the important increase in recombination observed for moderate values of the selfing rate is generated by the term in $D^0_{mab,ab}$, and thus corresponds to the benefit of recombination previously described by [25]. One can note that the model overestimates the strength 625 of selection for recombination in this case. This is possibly due to the fact that, while the effect of $e_{\mathrm{d\times d}}$ on effective dominance coefficients $a_{j,j}$ is negligible as long as epistasis is sufficiently weak, one can show that $a_{j,j}$ increases with $e_{\mathrm{d\times d}}$ when dominance-by-dominance epistasis becomes the main source of selection against deleterious alleles (see Appendix D in [50]), reducing selection for recombination through the first term 630 of equation 13.

DISCUSSION

Digging into the causes of general empirical patterns such as the positive correlation between selfing rate and chiasma frequency observed within several families

of flowering plants may help us to gain a better understanding of the selective forces affecting the evolution of recombination rates. While there is no obvious reason why the mechanistic constraints associated with chromosomal segregation during meiosis should differ between outcrossing and selfing species, the mating system of organisms does affect the benefit of recombination through its effect on genetic variation. Although indirect selective forces acting on recombination are expected to vanish under complete selfing (as heterozygosity should then be very rare), the results presented in this article show that intermediate selfing rates may either increase or decrease selection for recombination caused by interference (Hill-Robertson effect) among deleterious mutations, depending on parameter values. Roughly, selfing leads to stronger selection for increased chromosomal map length as long as $sh \ll R\,(1-F)$ (mostly due to the fact that selfing reduces effective recombination rates, thus increasing the strength of genetic associations), while selfing decreases selection for recombination when $sh \gg R\,(1-F)$, due to the fact that changes in recombination rates have only little effect on genetic associations in this regime. Given that most deleterious mutations seem to have weak fitness effects (e.g., [60]), selection for recombination should thus be increased by selfing, and be maximized for selfing rates slightly below 1 (as illustrated by Figures 3 and 4).

In agreement with previous results [8, 23, 24], we found that interference is often the main driver of selection for recombination when linkage is tight (or when the selfing rate is strong), while deterministic effects tend to become more important when recombination is abundant. In the absence of epistasis, the variance in the degree of inbreeding among individuals caused by partial selfing (associated with a more efficient purging of deleterious alleles in more inbred lineages) generates positive associations among deleterious alleles, selecting against recombination in infinite populations. As a result, the equilibrium map length may be lower under moderate selfing rates than under random mating in large, highly recombining populations, as can be seen on Figure 3 with $c = 10^{-4}$. As in standard models of infinite, randomly-mating populations [16, 17], negative epistasis between mutations tends to favor recombination by generating negative linkage disequilibria between deleterious alleles. This effect is often maximized at high selfing rates, again due to the fact that selfing reduces effective recombination rates, thus increasing the magnitude of linkage disequilibria.

Furthermore, components of epistasis involving dominance ($e_{\mathrm{a\times d}}$, $e_{\mathrm{d\times d}}$) also contribute to generating linkage disequilibrium (through an effective epistasis parameter) when the selfing rate is greater than zero.

Under high effective recombination, Roze & Lenormand [25] showed that correlations in homozygosity among loci generated by partial selfing (identity disequilibria) are the main drivers of indirect selection on recombination rates, favoring higher recombination when dominance-by-dominance epistasis is negative (as recombination then benefits from a short-term advantage, by increasing the mean fitness of offspring). The results of the present article tune down the relative importance of this effect, by showing that it becomes negligible relative to the effect of deterministically generated linkage disequilibria and Hill-Robertson interference under high selfing and in the case of tightly linked loci. Nevertheless, moderate selfing rates may favor a significant increase in chromosomal map length through this short-term benefit of breaking identity disequilibria, provided that dominance-by-dominance interactions are negative and represent a major component of epistasis (Figure 5C). In principle, the average sign and overall importance of dominance-by-dominance effects can be inferred from the shape of the relation between the degree of inbreeding of individuals and their fitness ([61], p. 80), negative $e_{\mathrm{d\times d}}$ causing a faster than linear decline in fitness with inbreeding. This method was used on several plant species and did not yield any clear evidence for negative $e_{\mathrm{d\times d}}$ [62, 63]; however, the methodology used (involving experimental crosses to increase the degree of inbreeding of individuals) generates a bias against finding negative $e_{\mathrm{d\times d}}$, as deleterious alleles may have been purged from the more highly inbred lines. Using an experimental protocol that avoids this bias, Sharp & Agrawal [64] found evidence for negative $e_{\mathrm{d\times d}}$ (on viability) between EMS-induced mutations in *Drosophila melanogaster*. While more work is needed to assess the generality of this result, previous experimental studies showed that epistasis is typically quite variable among pairs of loci [65–67]. As shown by [68], recombination tends to be less favored when epistasis is variable, and it would thus be of interest to extend our model to more realistic fitness landscapes including distributions of epistasis.

The indirect benefits of increased crossover rates can be quite strong when recombination is rare (e.g., [23]) but typically become rather weak under frequent recombination [24], and one may thus wonder to what extent our model can explain a

positive effect of selfing on chiasma frequency when at least one crossover per chromosome occurs during meiosis. When direct selection on map length takes the form of stabilizing selection around 1 crossover per bivalent (corresponding to a map length of 50cM, that is, $R = 0.5$), one indeed observes very limited effects of indirect selection on the evolutionarily stable map length (which always stays very close to 0.5), unless stabilizing selection is extremely weak (Figure 6). Furthermore, in the absence of epistasis, the main effect of indirect selection is to lower the ES map length under intermediate selfing rates (as can be seen for $c = 0.001$ in Figure 6), since deterministic effects (selecting against recombination) become stronger than stochastic effects when R is large. However, our model assumes that crossovers can occur anywhere along the chromosome with the same probability, which is probably an extreme and rather unrealistic situation. In particular, one could imagine that structural constraints may impose that the obligate crossover occurs in a given chromosomal region (for example, crossovers often tend to occur in subtelomeric regions, e.g., [69]), which would increase indirect selection for recombination in the other regions. Another extreme scenario would correspond to a situation where obligate crossovers occur exactly at the telomeres (and thus have no effect on genetic variation), in which case the predicted increase in map length (above 0.5) caused by indirect selection would be given by Figures 3–5 (showing that R increases with the selfing rate in many cases). The reality lies probably somewhere between these two extremes, calling for the need to consider less simplistic models for the distribution of crossovers along chromosomes (also incorporating crossover interference); this will be explored in a future work. Sweeps of beneficial mutations may also increase the strength of indirect selection for recombination [24, 70]. In particular, transitions towards predominant selfing are often followed by a number of phenotypic changes such as reduced flower size ("selfing syndrome", e.g., [71]) that may be seen as adaptations to the new mating system, thus involving the spread of newly beneficial alleles that may (at least transiently) favor higher recombination rates. Recombination rates may also be further increased after transitions to selfing due to stronger Hill-Robertson effects caused by population bottlenecks and extinction/recolonization dynamics that often characterize selfing populations (e.g., [72–74]).

Interestingly, predominantly selfing species often maintain low rates of out-

crossing in nature (e.g., [75, 76]), which may possibly be the consequence of selective forces favoring the maintenance of recombination. Indeed, deterministic and stochastic multilocus models showed that individuals outcrossing at a low rate are often selectively favored over complete selfers in conditions where recombination is advantageous (negative epistasis or finite population size, [49, 52]). However, higher effective recombination rates could in principle be achieved by increasing either the outcrossing rate of individuals or the number of crossovers at meiosis, and one could imagine that one solution or the other may be favored depending on the different types of direct and indirect selective forces that may act on outcrossing vs. recombination modifiers. Exploring the joint evolution of outcrossing and recombination rates in predominantly selfing species would thus be of interest, both from a theoretical and empirical perspective.

Acknowledgements. We thank the bioinformatics and computing service of Roscoff's Biological Station (Abims platform) for computing time, and the Agence Nationale pour la Recherche for funding (SelfRecomb project: ANR-18-CE02-0017-02, and GenAsex project: ANR-17-CE02-0016-01).

LITERATURE CITED

[1] Otto SP. Selective interference and the evolution of sex. J Hered. 2021;112:9-18.

[2] Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010;467:1099-103.

[3] Johnston SE, Bérénos C, Slate J, Pemberton JM. Conserved genetic architecture underlying individual recombination rate variation in a wild population of Soay sheep (*Ovis aries*). Genetics. 2016;203:583-98.

[4] Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Phil Trans Roy Soc (Lond) B. 2017;372:20160455.

[5] Ritz KR, Noor MAF, Singh ND. Variation in recombination rate: adaptive or not? Trends Genet. 2017;33:364-74.

[6] Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. Curr Biol. 2018;28:1289-95.

[7] Samuk K, Manzano-Winkler B, Ritz KR, Noor MAF. Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. Curr Biol. 2020;30:1517-28.

[8] Otto SP, Barton NH. Selection for recombination in small populations. Evolution. 2001;55(10):1921-31.

[9] Gray D, Cohen PE. Control of meiotic crossovers: from double-strand break formation to designation. Ann Rev Gen. 2016;50:175-210.

[10] Zelkowski M, Olson MA, Wang M, Pawlowski W. Diversity and determinants of meiotic recombination landscapes. Trends Genet. 2019;35:359-70.

[11] Koehler KE, Scott Hawley R, Sherman S, Hassold T. Recombination and nondisjunction in human and flies. Hum Mol Genet. 1996;5:1495-504.

[12] Otto SP, Payseur BA. Crossover interference: shedding light on the evolution of recombination. Ann Rev Gen. 2019;53:19-44.

[13] Ottolini CS, Newnham LJ, Capalbo A, Natesan SA, Joshi HA, Cimadomo D, et al. Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. Nat Genet. 2015;47:727-37.

[14] Otto SP, Lenormand T. Resolving the paradox of sex and recombination. Nat Rev Genet. 2002;3:252-61.

[15] Agrawal AF. Evolution of sex: why do organisms shuffle their genotypes? Curr Biol. 2006;16:R696-704.

[16] Charlesworth B. Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet Res. 1990;55:199-221.

[17] Barton NH. A general model for the evolution of recombination. Genet Res. 1995;65:123-44.

[18] Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8(3):269-94.

[19] Felsenstein J. The evolutionary advantage of recombination. Genetics. 1974;78:737-56.

[20] Otto SP, Barton NH. The evolution of recombination: removing the limits to natural selection. Genetics. 1997;147(2):879-906.

[21] Barton NH, Otto SP. Evolution of recombination due to random drift. Genetics. 2005;169:2353-70.

[22] Roze D, Barton NH. The Hill-Robertson effect and the evolution of recombination. Genetics. 2006;173:1793-811.

[23] Keightley PD, Otto SP. Interference among deleterious mutations favours sex and recombination in finite populations. Nature. 2006;443:89-92.

[24] Roze D. A simple expression for the strength of selection on recombination generated by interference among mutations. Proc Natl Acad Sci U S A. 2021;118:e2022805118.

[25] Roze D, Lenormand T. Self-fertilization and the evolution of recombination. Genetics. 2005;170:841-57.

[26] Ross-Ibarra J. Genome size and recombination in angiosperms: a second look. J Evol Biol. 2007;20:800-6.

[27] Kuittinen H, de Haan AA, Vogl C, Oikarinen S, Leppälä J, Koch M, et al. Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics. 2004;168:1575-84.

[28] Hansson B, Kawabe A, Preuss S, Kuittinen H, Charlesworth D. Comparative gene mapping in *Arabidopsis lyrata* chromosomes 1 and 2 and the corresponding *A. thaliana* chromosome 1: recombination rates, rearrangements and centromere location. Genet Res. 2006;87:75-85.

[29] Kawabe A, Hansson B, Forrest A, Hagenblad J, Charlesworth D. Comparative gene mapping in *Arabidopsis lyrata* chromosomes 6 and 7 and *A. thaliana* chromosome IV: evolutionary history, rearrangements and local recombination rates. Genet Res. 2006;88:45-56.

[30] Borts RH, Haber JE. Meiotic recombination in yeast: alteration by multiple heterozygosities. Science. 1987;237:1459-65.

[31] Chen W, Jinks-Robertson S. The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. Genetics. 1999;151:1299-313.

[32] Benavente E, Sybenga J. The relation between pairing preference and chiasma frequency in tetrasomics of rye. Genome. 2004;47:122-33.

825 [33] Ziolkowski PA, Berchowitz LE, Lambing C, Yelina NE, Zhao X, Kelly KA, et al. Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. eLife. 2015;4:e03708.

[34] Blackwell AR, Dluzewska J, Szymanska-Lejman M, Desjardins S, Tock AJ, Kbiri N, et al. MSH2 shapes the meiotic crossover landscape in relation to interhomolog
830 polymorphism in Arabidopsis. EMBO J. 2020;39:e104858.

[35] Nordborg M. Linkage disequilibrium, gene trees and selfing: and ancestral recombination graph with partial self-fertilization. Genetics. 2000;154:923-9.

[36] Wright SI, Ness RW, Foxe JP, Barrett SCH. Genomic consequences of outcrossing and selfing in plants. Int J Plant Sci. 2008;169:105-18.

835 [37] Charlesworth D, Charlesworth B, Strobeck C. Effects of selfing on selection for recombination. Genetics. 1977;86:213-26.

[38] Charlesworth D, Charlesworth B, Strobeck C. Selection for recombination in partially self-fertilizing populations. Genetics. 1979;93:237-44.

[39] Holsinger KE, Feldman MW. Linkage modification with mixed random mating
840 and selfing: a numerical study. Genetics. 1983;103:323-33.

[40] Glémin S, Ronfort J. Adaptation and maladaptation in selfing in outcrossing species: new mutations versus standing variation. Evolution. 2013;67:225-40.

[41] Roze D. Background selection in partially selfing populations. Genetics. 2016;203:937-57.

845 [42] Barton NH, Turelli M. Natural and sexual selection on many loci. Genetics. 1991;127(1):229-55.

[43] Kirkpatrick M, Johnson T, Barton NH. General models of multilocus evolution. Genetics. 2002;161(4):1727-50.

[44] Nagylaki T. The evolution of multilocus systems under weak selection. Genetics.
850 1993;134:627-47.

[45] Roze D. Selection for sex in finite populations. J Evol Biol. 2014;27:1304-22.

[46] Gervais C, Roze D. Mutation rate evolution in partially selfing and partially asexual organisms. Genetics. 2017;207:1561-75.

[47] Nordborg M. Structured coalescent processes on different time scales. Genetics. 1997;146:1501-14.

[48] Abu Awad D, Roze D. Epistasis, inbreeding depression and the evolution of self-fertilization. Evolution. 2020;74:1301-20.

[49] Charlesworth B, Morgan MT, Charlesworth D. Multilocus models of inbreeding depression with synergistic selection and partial self-fertilization. Genet Res. 1991;57:177-94.

[50] Roze D. Diploidy, population structure and the evolution of recombination. Am Nat. 2009;174:S79-94.

[51] Weir BS, Cockerham CC. Mixed self and random mating at two loci. Genet Res. 1973;21:247-62.

[52] Kamran-Disfani A, Agrawal AF. Selfing, adaptation and background selection in finite populations. J Evol Biol. 2014;27:1360-71.

[53] Pollak E. On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics. 1987;117:353-60.

[54] Haldane JBS. The combination of linkage values and the calculation of distances between the loci of linked factors. J Genet. 1919;8:299-309.

[55] Hudson RR, Kaplan NL. Deleterious background selection with recombination. Genetics. 1995;141:1605-17.

[56] Padhukasahasram B, Marjoram P, Wall JD, Bustamante CD, Nordborg M. Exploring population genetics models with recombination using efficient forward-time simulations. Genetics. 2008;178:2417-27.

[57] Charlesworth B, Charlesworth D. Rapid fixation of deleterious alleles can be caused by Muller's ratchet. Genet Res. 1997;70:63-73.

[58] Pálsson S, Pamilo P. The effects of deleterious mutations on linked, neutral variation in small populations. Genetics. 1999;153:475-83.

[59] Waller DM. Addressing Darwin's dilemma: can pseudo-overdominance explain persistent inbreeding depression and load? Evolution. 2021;75:779-93.

[60] Charlesworth B. Causes of natural variation in fitness: Evidence from studies of *Drosophila* populations. Proc Natl Acad Sci U S A. 2015;112:1662-9.

[61] Crow JF, Kimura M. An Introduction to Population Genetics Theory. New York: Harper and Row; 1970.

[62] Willis JH. Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. Evolution. 1993;47:864-76.

[63] Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. Harlow: Addison Wesley Longman; 1996.

[64] Sharp NP, Agrawal AF. The decline in fitness with inbreeding: evidence for negative dominance-by-dominance epistasis in *Drosophila melanogaster*. J Evol Biol. 2016;29:857-64.

[65] de Visser JAGM, Elena SF. The evolution of sex: empirical insights into the roles of epistasis and drift. Nat Rev Genet. 2007;8:139-49.

[66] Kouyos RD, Silander OK, Bonhoeffer S. Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol Evol. 2007;22:308-15.

[67] Martin G, Elena SF, Lenormand T. Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat Genet. 2007;39:555-60.

[68] Otto SP, Feldman MW. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. Theor Popul Biol. 1997;51(2):134-47.

[69] Haenel Q, Laurentino TG, Roesti M, Berner D. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol Ecol. 2018;27:2477-97.

[70] Hartfield M, Otto SP, Keightley PD. The role of advantageous mutations in enhancing the evolution of a recombination modifier. Genetics. 2010;184:1153-64.

[71] Cutter AD. Reproductive transitions in plants and animals: selfing syndrome, sexual selection and speciation. New Phytol. 2019;224:1080-94.

[72] Guo YL, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D, et al. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. Proc Natl Acad Sci U S A. 2009;106:5246-51.

[73] Willi Y, Fracassetti M, Zoller S, Van Buskirk J. Accumulation of mutational load at the edges of a species range. Mol Biol Evol. 2018;35:781-91.

[74] Orsucci M, Milesi P, Hansen J, Girodolle J, Glémin S, Lascoux M. Shift in ecological strategy helps marginal populations of shepherd's purse (*Capsella bursa-pastoris*) to overcome a high genetic load: competition avoidance and colonization. Proc Roy Soc (Lond) B. 2020;287:20200463.

[75] Bonnin I, Ronfort J, Wozniak F, Olivieri I. Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). Mol Ecol. 2001;10:1371-83.

[76] Bomblies K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N, et al. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. PLoS Genetics. 2010;6:e1000890.

APPENDIX A: HIGH EFFECTIVE RECOMBINATION

Under high effective recombination, the associations between the modifier and the selected loci involved in equation 6 are given by (see *Mathematica* notebook for derivation):

$$D_{mab,a} \approx D_{ma,ab} \approx F\left(\alpha_b + \alpha_{b,b}\right) D^0_{mab,ab} \tag{A1}$$

$$D_{ma} \approx \frac{\left(1+2F\,r_{ma}\right)\left[2F\left(\alpha_a+\alpha_{a,a}\right)\alpha_{b,b}+\alpha_{ab,b}+\alpha_{ab,ab}\right]}{r_{ma}\left(1-F\right)} D^0_{mab,ab} \tag{A2}$$

$$D_{m,a} \approx \frac{F\left(1+2r_{ma}\right)}{1+2F\,r_{ma}} D_{ma} \tag{A3}$$

$D_{mab,b}$, $D_{mb,ab}$, D_{mb} and $D_{m,b}$ being given by symmetric expressions. D_{mab}, $D_{m,ab}$ and $D_{ma,b}$ are given by:

$$\begin{aligned} D_{mab} \approx\ & \frac{\left(1+2F\,r_{ma}\right)\left(1+2F\,r_{mb}\right)}{\left(1-F\right)\left[r_{mab}+2F\left(r_{ma}r_{mb}+r_{ma}r_{ab}+r_{mb}r_{ab}\right)+6F^2 r_{ma}r_{mb}r_{ab}\right]} \\ & \times\Big[\left(1+2F\,r_{ab}\right)T_{ab}\,D^0_{mab,ab} \\ & -\delta r_{ab}\,H_m\left(D_{ab}-D_{a,b}\right)pq_m-\delta r_{ab}\left(1-H_m\right)\left(D_{mab,m}-D_{ma,mb}\right)\Big] \end{aligned} \tag{A4}$$

$$D_{m,ab} \approx \frac{F\left[\left(1+2F\,r_{ma}\right)\left(1+2F\,r_{mb}\right)+\left(1-F\right)\left(r_{ma}+r_{mb}+4F\,r_{ma}r_{mb}\right)\right]}{\left(1+2F\,r_{ma}\right)\left(1+2F\,r_{mb}\right)} D_{mab} \tag{A5}$$

$$\begin{aligned} D_{ma,b} \approx\ & \frac{F}{\left(1-F\right)\left[r_{mab}+2F\left(r_{ma}r_{mb}+r_{ma}r_{ab}+r_{mb}r_{ab}\right)+6F^2 r_{ma}r_{mb}r_{ab}\right]} \\ & \times\Big[\left(1+2F\,r_{ma}\right)\Big[\left(1+2F\,r_{mb}\right)\left(1+2F\,r_{ab}\right) \\ & \qquad +\left(1-F\right)\left(r_{mb}+r_{ab}+4F\,r_{mb}r_{ab}\right)\Big]T_{ab}\,D^0_{mab,ab} \\ & -\left(1+2F\,r_{mb}\right)\left[1-\left(1-3F\right)r_{ma}\right]\Big[\delta r_{ab}\,H_m\left(D_{ab}-D_{a,b}\right)pq_m \\ & \qquad +\delta r_{ab}\left(1-H_m\right)\left(D_{mab,m}-D_{ma,mb}\right)\Big]\Big] \end{aligned} \tag{A6}$$

($D_{mb,a}$ being given by a symmetric expression), with $H_m = h_m + \left(1-2h_m\right)p_m$, and:

$$T_{ab} = \alpha_{ab}+\alpha_{ab,a}+\alpha_{ab,b}+\alpha_{ab,ab}+2F\left(\alpha_a+\alpha_{a,a}\right)\left(\alpha_b+\alpha_{b,b}\right) \tag{A7}$$

$$D_{ab}-D_{a,b} \approx \frac{\left(1-F\right)\left[\tilde{\alpha}_{ab}-\tilde{\alpha}_a\tilde{\alpha}_b+2F\,G_{ab}\left(\alpha_a+\alpha_{a,a}\right)\left(\alpha_b+\alpha_{b,b}\right)\right]}{r_{ab}\left(1-F\right)} pq_{ab} \tag{A8}$$

$$D_{mab,m}-D_{ma,mb} \approx \frac{F\left(r_{ma}+r_{mb}-r_{ab}-2r_{ma}r_{mb}\right)}{1+F\left(r_{ma}+r_{mb}+r_{ab}-2r_{ma}r_{mb}\right)}\left(D_{ab}-D_{a,b}\right)pq_m. \tag{A9}$$

The identity disequilibrium G_{ab} is given by $\phi_{ab}-F^2$, with:

$$\phi_{ab} = F\left[1-\frac{2\left(1-F\right)r_{ab}\left(1-r_{ab}\right)}{1+2F\,r_{ab}\left(1-r_{ab}\right)}\right]. \tag{A10}$$

APPENDIX B: INTERPRETING EFFECTIVE SELECTION COEFFICIENTS

The form of the coefficients $\tilde{\alpha}_a$ and $\tilde{\alpha}_{ab}$ representing the effective strength of selection against deleterious alleles and the effective epistasis between pairs of deleterious alleles in a partially selfing population can be understood as follows. Assuming that deleterious alleles stay rare in the population ($u \ll s$), we can neglect homozygotes for deleterious alleles produced by outcrossing, and consider that homozygosity for deleterious alleles necessarily implies identity-by-descent among these alleles. Then, if an individual carries allele a on one of its haplotypes, with probability F it also carries allele a on its second haplotype. In that case, the fitness effect of the deleterious allele (α_a) is increased by α_a due to the presence of a on the second haplotype, and by $\alpha_{a,a}$ due to the interaction among those two deleterious alleles: therefore, the effective strength of selection experienced by deleterious alleles is $\tilde{\alpha}_a = (1+F)\,\alpha_a + F\,\alpha_{a,a}$ (which, using equation 5, yields $\tilde{\alpha}_a \approx -s\tilde{h}$). A similar reasoning can be used to compute the effective epistasis coefficient. With probability ϕ_{ab}, an individual is identical-by-descent at both loci a and b; with probability $F - \phi_{ab}$ it is identical-by-descent at the first locus and not at the second (or at the second and not at the first), while with the complementary probability $1 - 2F + \phi_{ab}$ it is identical-by-descent at neither locus. Therefore, an individual carrying alleles a and b on one of its haplotypes is $aabb$ with probability ϕ_{ab}, $aaBb$ (or $Aabb$) with probability $F - \phi_{ab}$, and $AaBb$ with probability $1 + 2F - \phi_{ab}$. In the first case ($aabb$), the fitness effect of the interaction between a and b (α_{ab}) is increased by the interaction between a and b present on the other haplotype (α_{ab}), twice the interaction between deleterious alleles in trans ($\alpha_{a,b}$), twice the additive-by-dominance interactions ($\alpha_{ab,a}$, $\alpha_{ab,b}$), and by the dominance-by-dominance interaction ($\alpha_{ab,ab}$). In the second case ($aaAb$ or $Aabb$), it is increased by the interaction between deleterious alleles in trans ($\alpha_{a,b}$) and by the additive-by-dominance interaction ($\alpha_{ab,a}$ or $\alpha_{ab,b}$). This yields:

$$\begin{aligned}\tilde{\alpha}_{ab} &= \phi_{ab}\,(2\alpha_{ab} + 2\alpha_{a,b} + 2\alpha_{ab,a} + 2\alpha_{ab,b} + \alpha_{ab,ab}) \\ &\quad + (F - \phi_{ab})\,(2\alpha_{ab} + 2\alpha_{a,b} + \alpha_{ab,a} + \alpha_{ab,b}) + (1 - 2F + \phi_{ab})\,\alpha_{ab} \\ &= (1+\phi_{ab})\,\alpha_a + 2F\,\alpha_{a,b} + (F+\phi_{ab})\,(\alpha_{ab,a} + \alpha_{ab,b}) + \phi_{ab}\,\alpha_{ab,ab}.\end{aligned} \tag{B1}$$

APPENDIX C: GENERAL QLE APPROXIMATIONS

The following expressions combine approximations obtained under different regimes (high effective recombination, weak recombination, high selfing, epistasis of order ϵ^2 or ϵ, see Supplementary Material). The different associations affecting the change in frequency of the modifier (equation 6) are given by (with $D^0_{mab,ab}$ given by equation 10):

$$D_{mab,ab} \approx D^0_{mab,ab} + F\, D_{mab} \tag{C1}$$

$$D_{mab,a} \approx D_{ma,ab} \approx F\left[(\alpha_b + \alpha_{b,b})\, D^0_{mab,ab} + D_{mab}\right] \tag{C2}$$

$$D_{m,ab} \approx D_{ma,b} \approx D_{mb,a} \approx F\, D_{mab} \tag{C3}$$

$$D_{ma,a} \approx F\left[\alpha_{b,b} D^0_{mab,ab} + D_{ma}\right] \tag{C4}$$

$$D_{m,a} \approx F\, D_{ma} \tag{C5}$$

$$\begin{aligned} D_{mab} \approx \frac{(1+2F\,r_{ma})\,(1+2F\,r_{mb})}{Y_{mab}} \Big[&(1+2F\,r_{ab})\, T_{ab}\, D^0_{mab,ab} \\ &- \delta r_{ab}\, H_m\, (D_{ab} - D_{a,b})\, pq_m - \delta r_{ab}\,(1-H_m)\,(D_{mab,m} - D_{ma,mb})\Big] \end{aligned} \tag{C6}$$

$$D_{ma} \approx \frac{(1+2F\,r_{ma})\left[\left[2F\,(\alpha_a + \alpha_{a,a})\,\alpha_{b,b} + \alpha_{ab,b} + \alpha_{ab,ab}\right] D^0_{mab,ab} + (\tilde{\alpha}_b + \tilde{\alpha}_{ab})\, D_{mab}\right]}{r_{ma}\,(1-F) - (1+2F\,r_{ma})\,\tilde{\alpha}_a} \tag{C7}$$

($D_{mab,b}$, $D_{mb,ab}$, $D_{mb,b}$, $D_{m,b}$ and D_{mb} being given by symmetric expressions), with:

$$T_{ab} = \alpha_{ab} + \alpha_{ab,a} + \alpha_{ab,b} + \alpha_{ab,ab} + 2F\,(\alpha_a + \alpha_{a,a})\,(\alpha_b + \alpha_{b,b}) \tag{C8}$$

$$\begin{aligned} Y_{mab} = (1-F)\left[r_{mab} + 2\,(r_{ma}r_{mb} + r_{ma}r_{ab} + r_{mb}r_{ab}) + 6\,r_{ma}r_{mb}r_{ab}\right] \\ - (1+2F\,r_{ma})\,(1+2F\,r_{mb})\,(1+2F\,r_{ab})\,(\tilde{\alpha}_a + \tilde{\alpha}_b + \tilde{\alpha}_{ab}) \end{aligned} \tag{C9}$$

$$D_{ab} - D_{a,b} \approx \frac{(1-F)\left[\tilde{\alpha}_{ab} - \tilde{\alpha}_a\tilde{\alpha}_b + 2F\,G_{ab}\,(\alpha_a + \alpha_{a,a})\,(\alpha_b + \alpha_{b,b})\right]}{r_{ab}\,(1-F) - (1+2F\,r_{ab})\,(\tilde{\alpha}_a + \tilde{\alpha}_b + \tilde{\alpha}_{ab})}\, pq_{ab} \tag{C10}$$

$$D_{mab,m} - D_{ma,mb} \approx \frac{F\,(r_{ma} + r_{mb} - r_{ab} - 2r_{ma}r_{mb})}{1 + F\,(r_{ma} + r_{mb} + r_{ab} - 2r_{ma}r_{mb})}\,(D_{ab} - D_{a,b})\, pq_m. \tag{C11}$$

The change in frequency of the modifier is given by $\Delta p_m = \sum_{\mathbb{U},\mathbb{V}} \alpha_{\mathbb{U},\mathbb{V}} D_{m\mathbb{U},\mathbb{V}}$ where the double sum is over all elements of the set $\{\varnothing, a, b, ab\}$, and thus decomposes into a term generated by $D^0_{mab,ab}$ (that becomes predominant under high effective recombination) and a term generated by $D_{ab} - D_{a,b}$ (that becomes predominant under weak recombination or strong selfing).

Table 1: Parameters and variables of the model.

σ	Selfing rate
s, h, $\tilde{h}$	Selection, dominance and effective dominance coefficients of deleterious alleles
$e_{\mathrm{a\times a}}$, $e_{\mathrm{a\times d}}$, $e_{\mathrm{d\times d}}$, $\tilde{e}$	Additive-by-additive, additive-by-dominance, dominance-by-dominance epistasis and effective epistasis between deleterious alleles
β	Charlesworth et al's [49] synergistic epistasis coefficient
ϵ	Strength of selection (scaling parameter)
u, U	Deleterious mutation rate per locus, and per chromosome
r_{ij}, $\tilde{r}_{ij}$	Recombination rate and effective recombination rate between loci i and j
δr_{ab}, $\tilde{\delta} r_{ab}$	Effect and effective effect of the modifier on r_{ab}
h_m	Dominance coefficient of allele m at the modifier locus
N, N_{e}	Census and effective population size
R, R_{ES}	Chromosome map length and evolutionarily stable map length
δR	Effect of the modifier on R
$D_{\mathbb{U},\mathbb{V}}$	Genetic association between the sets $\mathbb{U}$ and $\mathbb{V}$ of loci present on the maternally and paternally inherited haplotypes of an individual
$\alpha_{\mathbb{U},\mathbb{V}}$	Effect of selection on the sets $\mathbb{U}$ and $\mathbb{V}$ of loci present on the maternally and paternally inherited haplotypes of an individual
W, $\overline{W}$	Fitness of an individual and average fitness
$F = \sigma/(2-\sigma)$	Inbreeding coefficient (probability of identity by descent at one locus)
ϕ_{ab}	Joint probability of identity by descent at loci a and b
$G_{ab} = \phi_{ab} - F^2$	Identity disequilibrium between loci a and b
p_j, q_j	Frequencies of the lower- and uppercase allele at locus j
n	Mean number of deleterious alleles per chromosome

Table 2: Fitness matrix in the three-locus model.

	AA	Aa	aa
BB	1	$1-hs$	$1-s$
Bb	$1-hs$	$(1-hs)^2+e_{\mathrm{a\times a}}$	$(1-hs)(1-s)+2e_{\mathrm{a\times a}}+e_{\mathrm{a\times d}}$
bb	$1-s$	$(1-hs)(1-s)+2e_{\mathrm{a\times a}}+e_{\mathrm{a\times d}}$	$(1-s)^2+4e_{\mathrm{a\times a}}+4e_{\mathrm{a\times d}}+e_{\mathrm{d\times d}}$

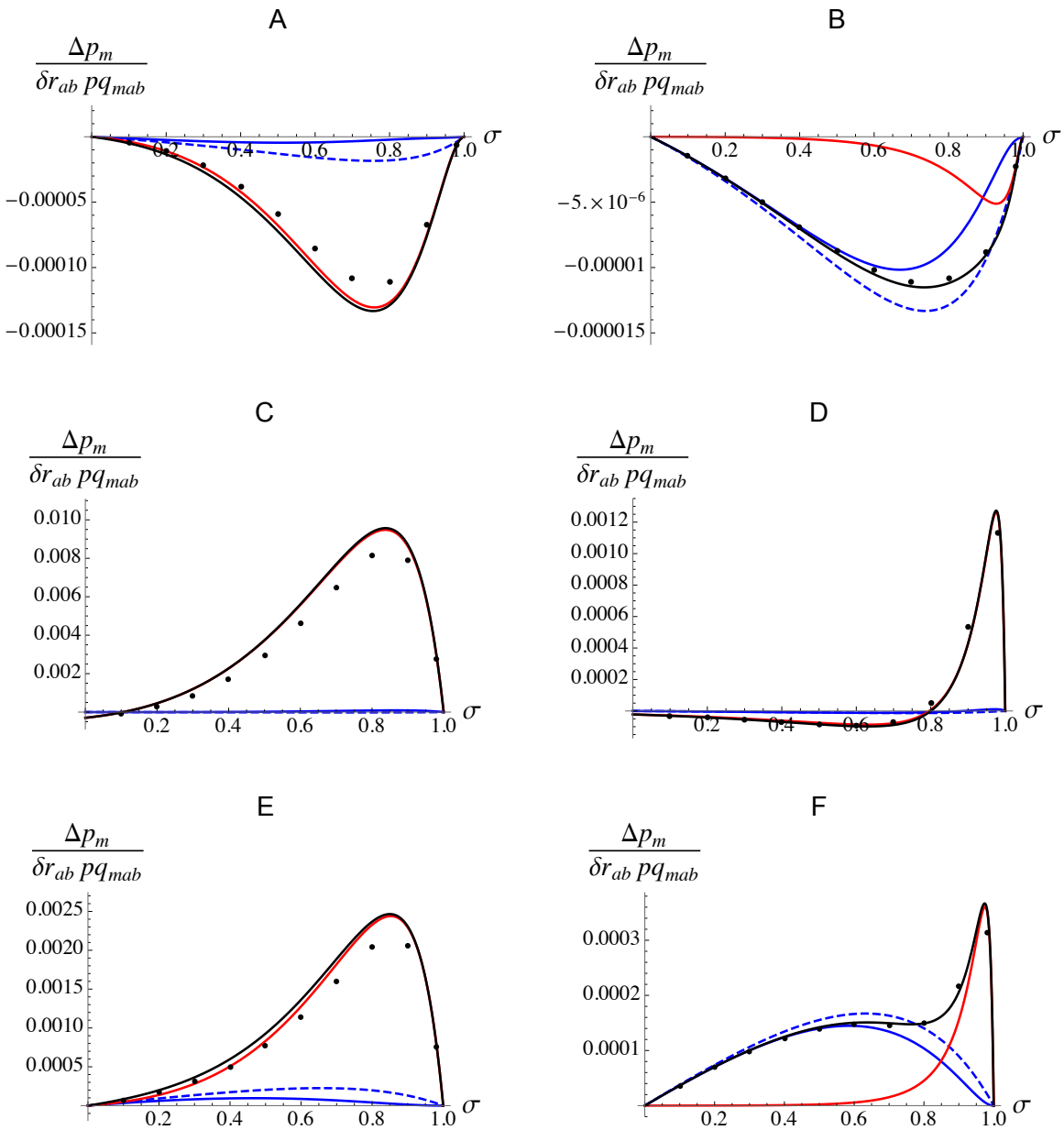

Figure 1. Effect of the selfing rate σ on the change in frequency of allele m in the deterministic three-locus model (scaled by $\delta r_{ab}\, pq_{mab}$) with no epistasis (A and B), negative additive-by-additive epistasis ($e_{\mathrm{a\times a}} = -0.001$; C and D) and negative dominance-by-dominance epistasis ($e_{\mathrm{d\times d}} = -0.001$; E and F), and for different recombination rates (A, C, E: $r_{ma} = r_{ab} = 0.01$; B, D, F: $r_{ma} = r_{ab} = 0.1$). Dots correspond to deterministic simulation results (see text for details) and black curves to the result obtained from equations 6 and C1 – C11, which is the sum of a term generated by $D_{ab} - D_{a,b}$ (red curves) and a term generated by $D^0_{mab,ab}$ (blue curves);

dashed blue curves correspond to the strong recombination approximation (Equation 11). Note that the red and black curves are nearly undistinguishable in C, D. The effect of additive-by-dominance epistasis ($e_{\mathrm{a\times d}}$) is similar to the effect of additive-by-additive epistasis ($e_{\mathrm{a\times a}}$) for these parameter values, and is shown in Figure S1. Loci are in the order $m - a - b$, parameters values are: $s = 0.01$, $h = 0.2$, $h_m = 0.5$, while $\delta r_{ab} = 0.01$ and $u = 10^{-5}$ in the simulations.

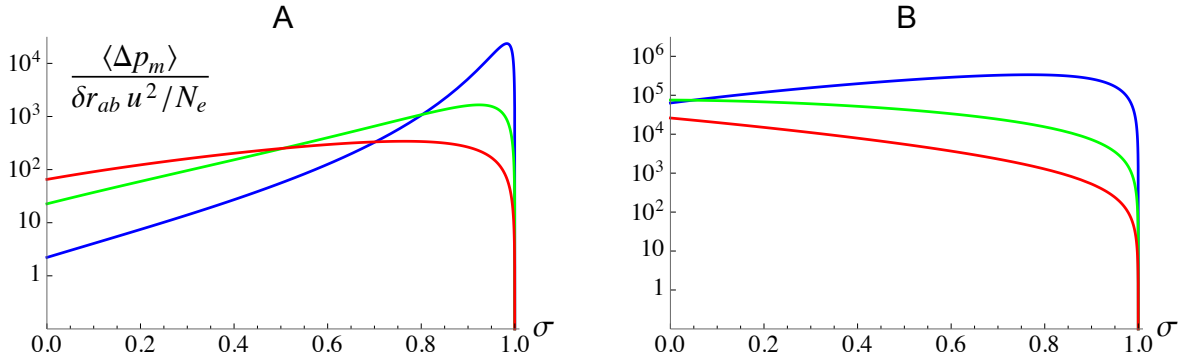

Figure 2. Effect of the selfing rate σ on the expected change in frequency of allele m (scaled by $\delta r_{ab}\, u^2/N_{\mathrm{e}}$ and on a log scale) generated by the Hill-Robertson effect between alleles a and b, for different values of the strength of selection against deleterious alleles (blue: $s = 0.005$; green: $s = 0.02$; red: $s = 0.05$) and recombination rates (A: $r_{ma} = r_{ab} = 0.05$; B: $r_{ma} = r_{ab} = 0.005$). The value of σ maximizing selection for recombination decreases as s increases and as recombination decreases. When selection is sufficiently strong and linkage sufficiently tight, selection for recombination decreases monotonously as σ increases (as can in seen in B for $s = 0.02$, $s = 0.05$). Loci are in the order $m - a - b$, the dominance coefficient of deleterious alleles is set to $h = 0.2$.

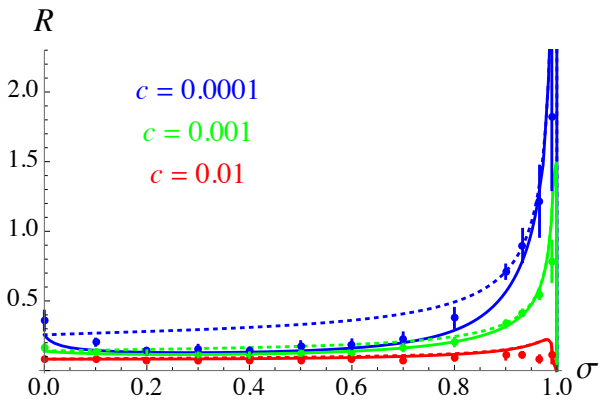

Figure 3. Effect of the selfing rate σ on the evolutionarily stable map length R_{ES}, for different values of the cost of recombination c. Solid curves correspond to the analytical predictions obtained by solving $s_{\mathrm{direct}} + s_{\mathrm{det}} + s_{\mathrm{HR}} = 0$ for R, where s_{det} and s_{HR} are obtained by integrating the expressions for the strength of indirect selection generated by deterministic effects (for s_{det}) and by the Hill-Robertson effect (for s_{HR}) over the genetic map (see Supplementary Material). Dotted curves correspond to the predictions obtained when ignoring indirect selection caused by deterministic effects (i.e., solving $s_{\mathrm{direct}} + s_{\mathrm{HR}} = 0$ for R). Dots correspond to individual-based simulation results; in this and the following figures, error bars are obtained by dividing the simulation output (after removing the first 5×10^5 generations) into 10 batches and calculating the variance of the average map length per batch, error bars measuring ± 1.96 SE. Parameter values are $N = 20{,}000$, $U = 0.1$, $s = 0.02$, $h = 0.2$, $e_{\mathrm{a\times a}} = e_{\mathrm{a\times d}} = e_{\mathrm{d\times d}} = 0$.

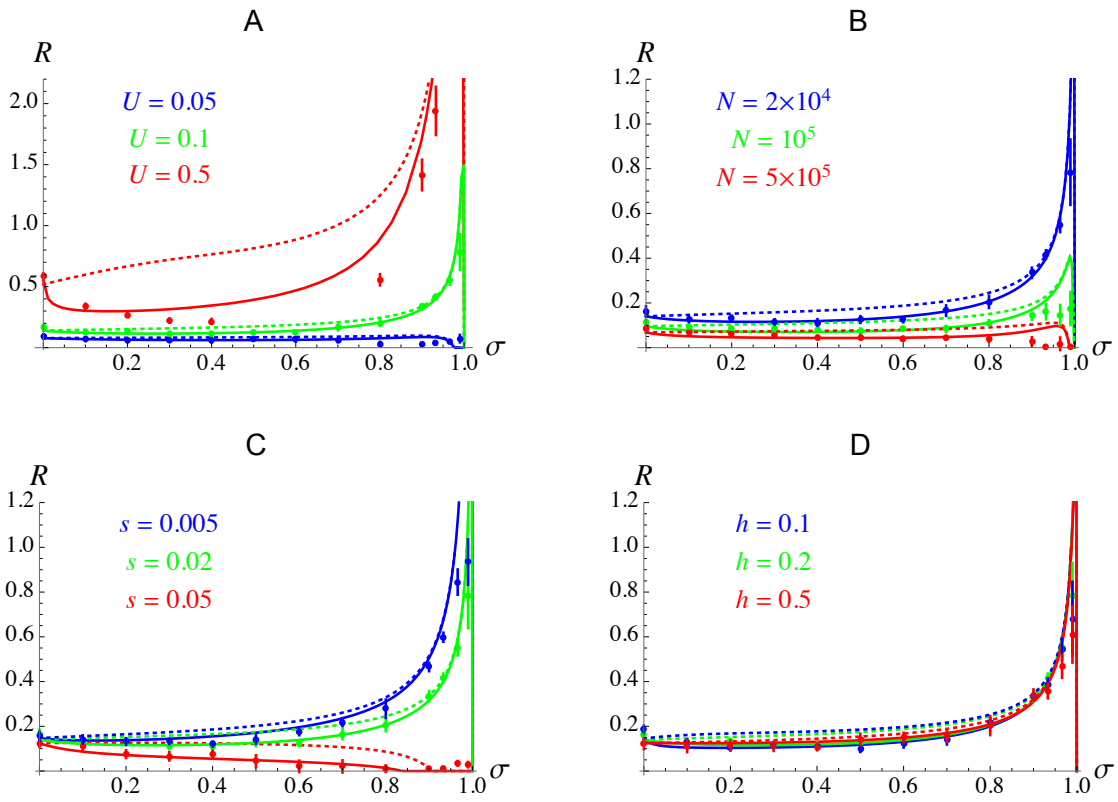

Figure 4. Effect of the selfing rate σ on the evolutionarily stable map length R_{ES}, for different values of the mutation rate U, population size N, strength of selection and dominance coefficient of deleterious alleles (s, h). Dots correspond to simulation results, curves have the same meaning as in Figure 3, and default parameter values are as in Figure 3 with $c = 0.001$ (no epistasis).

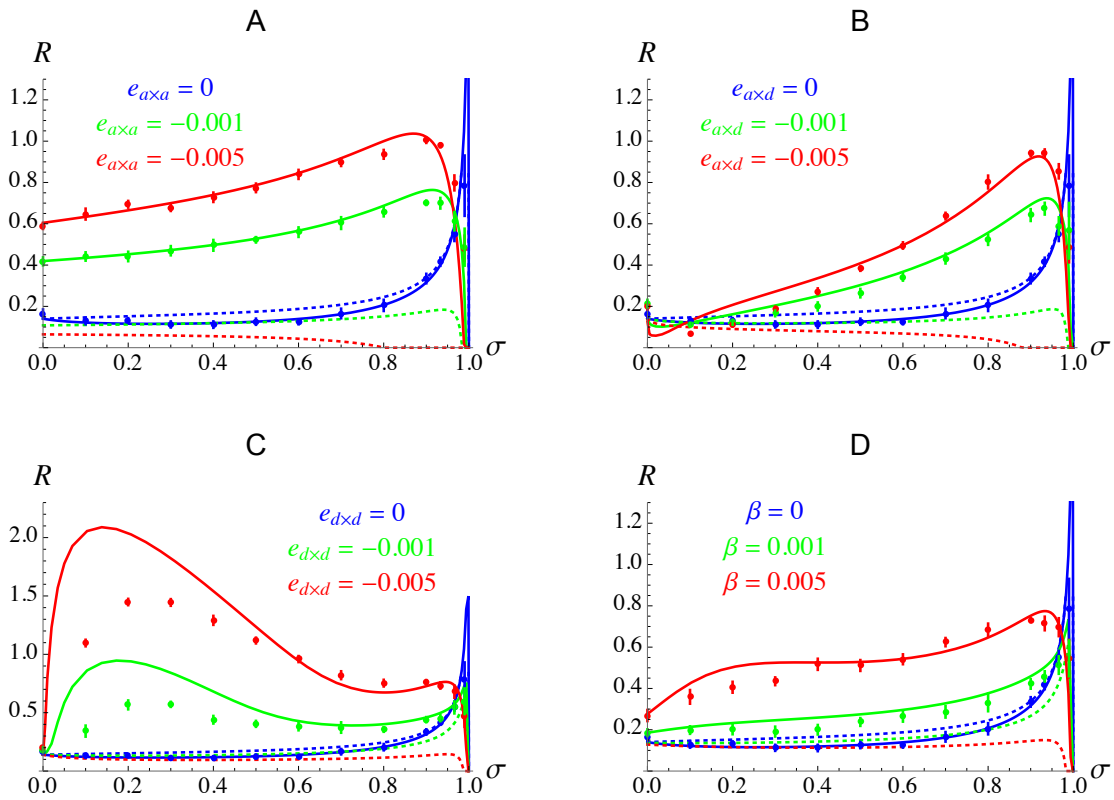

Figure 5. Effect of the selfing rate σ on the evolutionarily stable map length R_{ES}, for different values of additive-by-additive (A), additive-by-dominance (B) and dominance-by-dominance (C) epistasis, and different values of Charlesworth et al's [49] synergistic epistasis coefficient β (D). Note the different scale of the y-axis in C. Dots correspond to simulation results, curves have the same meaning as in Figure 3, and default parameter values are as in Figure 3 with $c = 0.001$.

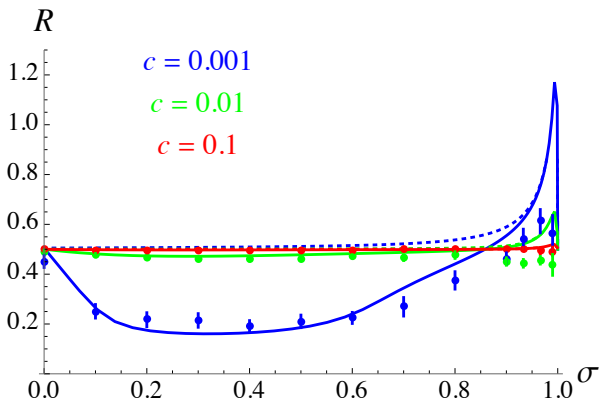

Figure 6. Effect of the selfing rate σ on the evolutionarily stable map length R_{ES} under direct stabilizing selection around $R = 0.5$ (1 crossover per bivalent) of varying strength c (stabilizing selection takes the form $\exp[-c(R-0.5)^2]$). Dots correspond to simulation results, curves have the same meaning as in Figure 3, and default parameter values are as in Figure 3.