A simple expression for the strength of selection on recombination generated by interference among mutations

Denis Roze

A simple expression for the strength of selection on recombination generated by interference among mutations

Denis Roze[*,†]

* CNRS, UMI 3614, 29680 Roscoff, France

† Sorbonne Université, Station Biologique de Roscoff, France

Classification: Biological Sciences, Evolution

Keywords: evolution of recombination, genetic architecture, genetic interference, meiosis, multilocus population genetics

Address for correspondence: Denis Roze, Station Biologique de Roscoff, Place Georges Teissier, CS 90074, 29688 Roscoff Cedex, France

Phone: (+33) 2 56 45 21 39, Fax: (+33) 2 98 29 23 24, Email: roze@sb-roscoff.fr

SIGNIFICANCE STATEMENT

Recombination between parental chromosomes during meiosis represents an important source of genetic novelty, and is thought to be the main evolutionary benefit of sexual reproduction. However, the evolutionary forces driving the rapid evolution of recombination rates demonstrated by comparisons between populations or closely related species remain obscure. This article provides the first mathematical quantification of the selective advantage of a mutation increasing the genetic map length (average number of crossovers occurring at meiosis) of a whole genome, due to the increased efficiency of selection against deleterious alleles. It shows that the advantage of recombination can be expressed as a simple expression of the mutation rate per unit map length, providing a simple way of evaluating its plausible order of magnitude.

ABSTRACT

One of the most widely cited hypotheses to explain the evolutionary maintenance of genetic recombination states that the reshuffling of genotypes at meiosis increases the efficiency of natural selection by reducing interference among selected loci. However, and despite several decades of theoretical work, a quantitative estimation of the possible selective advantage of a mutant allele increasing chromosomal map length (the average number of crossovers at meiosis) remains difficult. This article derives a simple and accurate expression for the strength of selection acting on a modifier gene affecting the genetic map length of a whole chromosome or genome undergoing recurrent mutation. In particular, it shows that indirect selection for recombination caused by interference among mutations is proportional to $(N_{\mathrm{e}}U)^2 / (N_{\mathrm{e}}R)^3$, where N_{e} is the effective population size, U the deleterious mutation rate per chromosome and R the chromosome map length. Indirect selection is relatively insensitive to the fitness effects of deleterious alleles, epistasis, or the genetic architecture of recombination rate variation, and may compensate for substantial costs associated with recombination when linkage is tight. However, its effect generally stays weak in large, highly recombining populations.

INTRODUCTION

Genetic variation for rates of crossing over at meiosis has been reported in several species [1–6], showing that recombination landscapes may evolve by selection or drift: accordingly, differences in recombination rates have been observed between closely related species [7–11] and over broader taxonomic scales [12, 13]. It has been recognized for long that both direct and indirect selective forces may drive the evolution of recombination [14–16]. Direct selection stems in particular from molecular constraints acting on the number of crossovers: in particular, it is usually thought that in most species, at least one crossover per bivalent is required to ensure proper chromosomal disjunction and segregation at meiosis; for example, in humans low recombination is associated with the production of aneuploid gametes and infertility [17–21]. Too many crossovers may also be detrimental, as it may lead to disjunction failure during the first meiotic division [22] and to elevated mutation rates [23]. Indirect selection corresponds to the potential benefits associated with the production of novel genotypes by recombination [14, 24]. In particular, recombination increases the efficiency of natural selection in the presence of negative linkage disequilibria (LD) between selected loci, that is, when beneficial alleles tend to be associated with deleterious alleles at other loci. Negative LD may be the consequence of epistatic interactions (on fitness) among loci [25, 26], but is also predicted to arise in any finite population under selection (a phenomenon known as the Hill-Robertson effect, or selective interference) [27–32].

The strength of indirect selection has been quantified under different scenarios using three-locus modifier models, representing a neutral modifier locus affecting the

rate of recombination between two selected loci (e.g., [25, 26, 29–35]). In general, these models show that indirect selection on a recombination modifier should mostly stem from its effect on selected loci to which it is tightly linked (as the modifier remains longer associated with the beneficial combinations it contributed to create). However, evaluating the overall strength of indirect selection on a modifier affecting the genetic map length of a whole genome or chromosome remains challenging. This is partly due to the fact that the contribution of higher-order disequilibria between selected loci (associations between 3, 4 or more loci) is difficult to assess, and also to the fact that the mathematical approximations used often break down in the case of tightly linked loci (corresponding to the situation in which indirect selection should be strongest). Multilocus simulation models have offered important insights [30, 36–39], showing that indirect selection caused by selective interference among many loci may be rather strong when linkage is tight. However, these simulations are necessarily restricted to limited ranges of parameters (in particular, they often focus on situations in which recombination is very rare) and therefore, how the strength of selection for recombination scales with the different parameters describing mutation and selection remains unclear. Another limitation of current theory is that most models on selective interference consider haploid organisms, while many eukaryotic species are diploid. As a consequence, we are still lacking general expressions quantifying the possible strength of selection for recombination at the level of a whole genome, and applicable to most extent species.

This article presents analytical expressions for the strength of selection on a modifier locus affecting the genetic map length R of a linear chromosome, in a diploid, randomly mating population of N individuals. The model assumes that deleterious

mutations occur at a rate U per haploid chromosome per generation at a very large number of possible sites, each mutation decreasing fitness by a factor $1 - hs$ when heterozygous and $1 - s$ when homozygous (however, we will see that some of the results extend to more general situations). The mathematical analysis of the model proceeds in two steps (detailed in the Methods and in the Supplementary Material). In a first step, the strength of indirect selection acting at the recombination modifier locus due to interference between two deleterious alleles (labelled a and b) at different loci is quantified (the expression obtained staying valid even when selected loci are tightly linked). In a second step, the result of this three-locus model is integrated over all possible positions of deleterious alleles along the chromosome, in order to predict the overall strength of selection for recombination as a function of N, s, h, U and R. Analytical predictions are compared with the results of individual-based, multi-locus simulations in which R evolves during a large number of generations. Various extensions including distributions of fitness effects of deleterious alleles, multiple recombination modifiers, multiple chromosomes, beneficial mutations and epistasis have also been explored, as explained in the Methods. A direct fitness cost associated with recombination is introduced in the simulation program, by assuming that the fitness of individuals is proportional to $\exp(-cR)$ (c may thus be considered as the fitness cost per crossover). Indeed, this provides a straightforward way of evaluating mathematical expressions by comparing the predicted map length at equilibrium (at which indirect selection exactly balances the cost of recombination) to its value observed in simulations, as well as a simple visualization of the effect of indirect selection for different parameter values.

RESULTS

The Hill-Robertson effect in diploids. While the Hill-Robertson effect generates negative linkage disequilibrium between deleterious alleles in finite haploid populations [31, 40], the present model shows that in diploids, the average LD between two deleterious alleles a and b (denoted $\langle D_{ab} \rangle$) may be either positive or negative depending on the dominance coefficient h of these alleles: $\langle D_{ab} \rangle$ is negative when $h > 0.25$, and positive when $h < 0.25$. This result is confirmed by two-locus simulations (Figure 1A). As explained in the Supplementary Material, positive $\langle D_{ab} \rangle$ stems from the fact that although deleterious alleles tend to decrease in frequency when they are in coupling, selection against those alleles becomes weaker as they reach lower frequencies (if they are partially recessive), allowing them to persist longer in the population (while deleterious alleles in coupling are more efficiently eliminated from the population in the absence of dominance). Although the average LD between two deleterious alleles stays very small (proportional to the product of their frequencies in the population), the sum of all pairwise LD between mutations occurring along a whole chromosome may significantly affect the variance in fitness, in particular when chromosomal map length becomes small. In this case, interference between each pair of loci is further amplified by the reduced effective population size N_e caused by selection acting at linked loci (background selection, e.g., [41], Figure 1B). Figures 1C and 1D show that extrapolations from the two-locus analytical result match reasonably well the multilocus simulation results when R is sufficiently large, while important discrepancies appear under tight linkage: in particular, the sum of all pairwise LD is always negative in the simulations when R is small, even for $h < 0.25$. These discrepancies must be due to

higher order interactions (involving three or more loci) affecting pairwise LD, that are not taken into account in the analysis.

The strength of selection for increased map length. Although the positive LD observed for intermediate values of R and $h < 0.25$ tends to disfavor recombination (as breaking positive LD decreases the variance in fitness and reduces the efficiency of selection), the mathematical analysis of the three-locus model shows that indirect selection on recombination involves at least 14 different mechanisms (corresponding to the different paths generating $\langle D_{ma} \rangle$ on Figure S1), of which only one involves $\langle D_{ab} \rangle$. All of these mechanisms favor recombination in the absence of dominance at the selected loci ($h = 0.5$), while dominance generate effects that disfavor recombination (for example, through its effect on $\langle D_{ab} \rangle$ just discussed) and other effects that favor recombination. Interestingly, these different effects of dominance tend to compensate each other (as shown by Figures S5, S6), so that the net effect of interference favors increased map length for most parameter values, and is often well approximated by ignoring the terms generated by dominance (at least as long as $h \geq 0.2$). In that case, the strength of indirect selection becomes equivalent as in a haploid population of size $2N$ in which mutations have an effect sh on fitness (results for haploids are derived in a *Mathematica* notebook available as Supplementary Material). Furthermore, when the fitness effect of deleterious alleles is sufficiently weak ($sh \ll R$) selection for recombination is mostly caused by segregating mutations located in the chromosomal vicinity of the recombination modifier. In that case, the strength of indirect selection on an additive modifier increasing map length by an amount δR is found to be approximately

$\delta R\, s_{\rm ind}$, with

$$s_{\rm ind} \approx 1.8\, \frac{(N_{\rm e}U)^2}{(N_{\rm e}R)^3} \tag{1}$$

independently of s and h, and with $N_{\rm e} \approx N \exp(-2U/R)$ under the model's assumptions (a more accurate result for higher values of sh or lower values of R can be obtained by numerical integration over the genetic map, as explained in the Methods and Supplementary Material).

The evolutionarily stable (ES) map length corresponds to the value of R for which indirect selection caused by interference exactly compensates the cost of recombination, that is, $s_{\rm ind} = c$. Figure 2 shows that the analytical model often provides accurate predictions of the ES map length, discrepancies appearing when the chromosomal mutation rate U is high, for parameter values leading to low equilibrium values of R (in particular, when the cost of recombination is strong). As explained in the Supplementary Material, the model predicts that the strength of indirect selection on recombination should scale with NR, NU and Ns (so that the ES value of NR should not depend on N as long as NU and Ns stay constant): this is confirmed by the simulation results shown on Figure S2A. Figure 2 also confirms that the selection and dominance coefficients of deleterious alleles have little effect on the magnitude of indirect selection as long as s is small; as a consequence, the results are robust to the introduction of a distribution of fitness effects of mutations, as illustrated by Figure S2C.

Because the model assumes that mutation and recombination events occur uniformly along the chromosome, and because indirect selection on the modifier is mostly caused by nearby loci, selection for recombination should not be much affected by the physical position of the modifier as long as map length R is not too small. Similarly,

equation 1 should still hold when map length is a polygenic trait coded by several loci located at various positions along the chromosome. Indeed, Figure S2D confirms that the same equilibrium map length is reached when R is coded by a single locus or by 100 loci with additive effects (adjusting parameters so that the mutational variance on R stays the same). The results also extend to the case of a genome consisting of multiple chromosomes (Figures S2E, S2F). Indeed, the evolution of a local recombination modifier affecting the map length of its own chromosome is not affected much by the presence of other chromosomes (as their only effect is to cause a modest reduction in N_e, by a factor $\sim \exp(-8shU)$ per extra chromosome), while indirect selection on a global modifier affecting the map length of all chromosomes mostly stems from its local effect, and is thus still approximately given by equation 1.

Including beneficial mutations. Obtaining analytical predictions for the equilibrium map length when beneficial and deleterious mutations co-occur remains challenging. Approximations for the strength of selection for recombination generated by interference between two beneficial alleles have been derived for the case of haploid populations, but in many cases, accurate predictions can only be obtained numerically [29, 32]. Furthermore, no simple expression exists for the effective population size and for the probability of fixation of beneficial mutations when both beneficial and deleterious alleles segregate at many loci. Therefore, the extra effect of beneficial mutations on selection for recombination was only explored by simulation (assuming a constant rate U_{ben} of mutation towards beneficial alleles, all with the same selection and dominance coefficients s_{ben}, h_{ben}).

As shown by Figure 3, higher rates of recombination evolve when beneficial mu-

tations co-occur with deleterious alleles, in particular when the deleterious mutation rate U is low. When U is high, selection for recombination is mostly caused by deleterious alleles, and the extra effect of beneficial mutations generally stays minor (Figure S3 shows that similar results are obtained when the rate of beneficial mutation U_{ben} is proportional to U). The strength of indirect selection caused by beneficial mutations increases with their heterozygous effect $s_{\text{ben}}h_{\text{ben}}$ (Figure 3B), while their dominance coefficient has only little effect as long as $s_{\text{ben}}h_{\text{ben}}$ stays constant (Figure 3C). As in the case of deleterious alleles, the strength of selection for recombination caused by beneficial alleles scales with NR, NU_{ben} and Ns_{ben} (Figure 3D).

Epistasis. Negative epistasis among mutations is known to generate a deterministic force favoring recombination [25, 26]. In order to asses its potential importance, the analytical and simulation models were extended to include pairwise negative epistasis among deleterious alleles, by assuming that each interaction between two deleterious alleles at different loci decreases fitness by a factor $1 + e$ (with $e < 0$). Increasing the magnitude of negative epistasis increases the effective strength of selection against mutations (thus potentially affecting interference among mutations), and the selection coefficient s is thus decreased as e becomes more negative in order to maintain a constant effective strength of selection (also ensuring that the average number of mutations per chromosome and the additive variance in fitness in the population remain constant). For a given effective strength of selection against deleterious alleles (corresponding to the fitness effect of a heterozygous mutation in an average genetic background), epistasis cannot be lower than a limit value (at which $s = 0$ and selection only stems from epistatic interactions) that depends on the mutation rate U, and

corresponds to the lowest values on the x-axes of Figure 4 (see Methods). Because selection for recombination due to interference depends on the effective strength of selection against deleterious alleles, it is predicted to stay constant along each curve of Figure 4. As can be seen from Figure 4, the effect of negative epistasis on selection for recombination often remains small relative to the effect of interference (as the equilibrium map length is not affected much by e), even for population sizes as large as 10^5. Figure S4 confirms that the average number of deleterious alleles per chromosome stays approximately constant in the simulations as e varies (due to the scaling of s), while mean fitness increases as epistasis becomes more negative [42]. As shown by Figure 4B, the effect of epistasis on the ES value of R becomes more important for high effective strengths of selection against deleterious alleles.

DISCUSSION

The observation that recombination rates may evolve over fast timescales raises the question of the relative importance of the different types of selective forces that may drive such evolution. As mentioned in introduction, mechanistic constraints associated with chromosomal segregation probably generate stabilizing selection around an optimal number of crossovers per bivalent [16, 43], whose exact shape and strength remain difficult to evaluate from current data. However, it is not immediately clear why such constraints would differ between closely related species, and one can imagine that, if not too strong, stabilizing selection caused by direct fitness effects may leave some room for evolutionary changes in recombination rates generated by indirect effects, as suggested by artificial selection experiments during which map length

increased as a correlated response (e.g., Table 1 in [30]). Although a large body of theoretical work has explored the possible selective advantages of recombination, assessing the plausible order of magnitude of indirect selection acting on chromosomal map length stays difficult, as it is generally not obvious how mathematical results from 3-locus modifier models extend to more realistic situations involving many genes. The results presented in this article show that extrapolations from 3-locus models accurately predict the overall strength of indirect selection acting on a modifier affecting the map length of a chromosome in finite diploid populations, as long as map length is not too small relative to the chromosomal mutation rate (roughly, when $U < R$). Under tight linkage ($U > R$), the analytical model tends to overestimate the strength of indirect selection (as can be seen from Figures 2 and 4): therefore, the approximations presented here may not accurately quantify selection for recombination in populations with very low (or no) recombination, but provide correct predictions in situations where recombination is already frequent, as in most sexual species. The fact that the model performs poorly when $U > R$ may be caused by higher-order interactions among selected loci, and also by the assumption that deleterious alleles stay near mutation–selection balance, which does not hold when $sh \ll 1/N_\mathrm{e}$ (N_e being greatly reduced by background selection when $U > R$, as shown by Figure 1B). While an analytical description of this regime remains challenging (e.g., [44]), simulation approaches are also problematic as mutations may accumulate at a high rate when selection is ineffective, and the equilibrium map length of a population whose mean fitness declines rapidly is probably not biologically meaningful. Possible compensatory effects among mutations should be taken into account when dealing with such situations [45], which would imply extending the model to incorporate distributions of epistasis.

Current estimates of the distribution of fitness effects of mutations indicate that most deleterious alleles have weak fitness effects (e.g., [46]). Interestingly, the model shows that in this regime (and as long as $sh > 1/N_e$ for most mutations), the strength of indirect selection generated by interference among mutations does not depend much on the details of the genetic architecture of fitness (selection and dominance coefficients of deleterious alleles), and can be approximated by a simple expression of $N_e U$ and $N_e R$ (equation 1). This stands in contrast with the evolution of sex modifiers (affecting the rate of sex in partially clonal organisms) which is more dependent on dominance: in particular, the simulation results of [47] showed that obligate asexuality is often favored when $h \leq 0.25$ (see Figure 7 in [47]). This difference probably stems from the fact that, unlike recombination modifiers, sex modifiers have a direct effect on heterozygosity among offspring (see also [48]). In agreement with previous results [30, 37], the effect of epistasis among mutations stays relatively small (and is well predicted by an extension of the model presented in [26]) even when population size is large (up to 10^5 in Figure 4A). Approximation 1 also shows that the $N_e s_{ind}$ product (determining to what extent indirect selection is efficient relative to drift) does not depend on N_e. From classical diffusion results, one thus predicts that the fixation probability of a recombination modifier (relative to the fixation probability of a neutral allele) should not depend on N_e, since this relative fixation probability is approximately $2N_e s_{ind}$ (e.g., p. 426 in [49]). This seems to contradict the simulation results obtained by Keightley & Otto [37], showing that the relative fixation probability of a recombination modifier increases with population size. This discrepancy is probably due to the fact that Keightley & Otto mostly considered situations in which $U \gg R$, while the present approximations break down in this regime (and also possibly from the fact that the classical diffusion result

for the fixation probability may not hold under strong interference). Interestingly, Keightley & Otto's results indicate that the relative fixation probability of the modifier may not depend much on population size N when $U = R = 0.1$ and N is not too small, however (Figure 1d in [37]), in agreement with the present results.

Present estimates of the rate of deleterious mutation per diploid genome are of the order 1 – 2 in organisms such as *Drosophila* and humans [46, 50], although these values are associated with considerable uncertainty. According to the present results (equation 1), the corresponding mutation rates per chromosome U may generate strong selection for increased map length in populations with very low recombination (allowing recombination to be maintained even in the presence of strong direct costs). However, indirect selection should generally stay rather weak when $R \approx 0.5$ (one crossover per bivalent). For example, Figure 5 shows the effect of the deleterious mutation rate U on the equilibrium value of R when direct selection takes the form of stabilizing selection around $R = 0.5$ (the direct fitness component being given by $\exp\left[-c\,(R-0.5)^2\right]$, with $c = 0.1$ so that an increase from $R = 0.5$ to $R = 1$ causes a fitness drop of about 2.5%). As can be seen on Figure 5, indirect selection only causes a modest increase in map length above $R = 0.5$ for these parameter values, in particular when population size is large. Yet, several factors may increase the strength of indirect selection. A first is that crossovers are generally not uniformly distributed along chromosomes, but tend to occur preferentially at the chromosome peripheries (at least in plants and animals), which may stem from constraints associated with the pairing of homologs during the first meiotic division [51]. While gene density is also higher at the chromosome peripheries in plants, this is not particularly the case in animals [51], and the local deleterious mutation rate per unit map length should thus

be higher in the central part of chromosomes, increasing the magnitude of indirect selection on recombination modifiers located in the central part. Second, sweeps of beneficial alleles may increase selection for recombination during periods of adaptation. While the results shown on Figures 3 and S3 indicate that the effect of beneficial alleles stays negligible when the beneficial mutation rate is very small relative to U ($U_{\text{ben}} < 10^{-3}\,U$), map length may be significantly increased by selective sweeps under higher values of U_{ben}, in particular when the fitness effect of advantageous mutations is not too small. Similarly, fluctuating selection acting at several loci may reinforce the overall effect of indirect selection [31]. Last, many populations present some form of spatial structure, increasing interference effects and selection for recombination due to local drift [52, 53]. Comparisons between populations or species presenting different demographies or degrees of spatial structure may thus yield further insights on the potential role of indirect selection in the evolution of recombination.

METHODS

Analytical three-locus model. The model represents a diploid population of size N with discrete generations, and considers three loci: a recombination modifier locus (with two alleles M and m) and two selected loci (each with two alleles, A, a at the first locus and B, b at the second). Alleles a and b are deleterious, reducing fitness by a factor $1 - h_i\,s_i$ when heterozygous (where i stands for a or b), and $1 - s_i$ when homozygous. The effects of deleterious alleles are multiplicative across loci (no epistasis): for example, the fitness of a double heterozygote is $(1 - s_a\,h_a)\,(1 - s_b\,h_b)$. Mutations towards deleterious alleles occur at a rate u per generation. Back mutations

are ignored, but their effect should be negligible as long as deleterious alleles stay rare in the population. Diploid parents produce a very large number of gametes (in proportion to their fitness) which fuse at random to produce zygotes (including the possibility of selfing), among which N are sampled randomly to form the next adult generation. At meiosis, the recombination rate between loci i and j in individuals with genotype MM, Mm and mm at the modifier locus is r_{ij}, $r_{ij}+h_m\delta r_{ij}$ and $r_{ij}+\delta r_{ij}$, respectively: δr_{ij} thus measures the effect of allele m on the recombination rate between loci i and j, while h_m is the dominance coefficient of this allele. In the Supplementary Material, an expression for the expected change in frequency at the modifier locus (valid for any ordering of the three loci along the chromosome) is derived to the first order in δr_{ij}, under the assumptions that selection coefficients and recombination rates are small (of order ϵ, where ϵ is a small term), drift is weak relative to selection ($1/N \ll \epsilon$) and $u \ll \epsilon$ so that the frequencies of deleterious alleles remain small. As in [31], the general principle of the method consists in deriving expressions for different moments of allele frequencies and linkage disequilibria. As long as selected loci are near mutation–selection balance, changes in allele frequencies remain small (of order $1/N \ll \epsilon$), so that quasi-linkage equilibrium approximations can be used even when recombination rates are small, yielding expressions that do not diverge under tight linkage and that may thus be integrated over the genome (see also [40, 54]). In the case of an additive recombination modifier ($h_m = 1/2$), the expected change in frequency of the modifier takes the form:

$$\langle \Delta p_m \rangle \approx \frac{\delta r_{ab}}{N} f(r_{ma}, r_{mb}, r_{ab}, s_a, h_a, s_b, h_b)\, \tilde{p}_a \tilde{p}_b p_m q_m \tag{2}$$

where f is a function of recombination rates, selection and dominance coefficients, and

where $\tilde{p}_a$, $\tilde{p}_b$ correspond to the frequencies of deleterious alleles at mutation-selection balance (see Supplementary Material and *Mathematica* notebook for derivations).

Multilocus extrapolation. The result from the three-locus model can be extrapolated to the case of a modifier affecting the map length R of a linear chromosome, along which deleterious mutations occur at a given rate U per generation. For simplicity, I assume that the modifier is located at the mid-point of the chromosome, that the density of mutations and crossovers is uniform along the chromosome, and that all deleterious alleles have the same selection and dominance coefficients s and h. Under these assumptions, one obtains that the strength of indirect selection at the modifier locus is given by:

$$s_{\text{ind}} \approx \frac{4U^2}{N_{\text{e}} R^3} \left[\int_0^{\frac{R}{2sh}} \int_0^{\frac{R}{2sh}} (x+y)\, g(x,y,x+y)\, dx\, dy + \int_0^{\frac{R}{2sh}} \int_0^{\frac{R}{2sh}} |x-y|\, g(x,y,|x-y|)\, dx\, dy \right] \tag{3}$$

where $g(\rho_{ma}, \rho_{mb}, \rho_{ab})$ is a function of scaled recombination rates $\rho_{ma} = r_{ma}/(sh)$, $\rho_{mb} = r_{mb}/(sh)$, $\rho_{ab} = r_{ab}/(sh)$ that can be found in the *Mathematica* notebook available as Supplementary Material. The first double integral in equation 3 corresponds to the overall effect of pairs of selected loci located on opposite sides of the modifier locus on the chromosome, and the second to the overall effect of pairs of loci located on the same side of the modifier locus. N_{e} corresponds to the effective population size, which is reduced by background selection effects. When R is sufficiently large, N_{e} remains approximately constant along the chromosome and given by $N_{\text{e}} \approx N \exp(-2U/R)$ [55]. When $R/(sh)$ is large, indirect selection mostly stems from the effect of loci located in the chromosomal vicinity of the modifier, and the

integrals in equation 3 may be approximated by the same integrals taken between zero and infinity, which yields equation 1. Note that, because the number of loci at which mutations can occur is effectively infinite in this extrapolation (infinite sites model), a given mutation occurs only once and does not reach mutation–selection balance. Nevertheless, the three-locus model (which assumes an equilibrium frequency of $u/(sh)$ for each mutation) still provides correct predictions for the strength of indirect selection in this limit (see also [40, 54]). Presumably, this is because a small tract of chromosome with mutation rate dU (and over which the mean number of deleterious alleles per haplotype is $\approx dU/(sh)$) behaves similarly as a locus in the three-locus model.

Epistasis. The analysis of [26] on the effect of epistasis on selection for recombination can be extended to the case of tightly linked loci segregating for deleterious alleles, and integrated over the genetic map (see Supplementary Material for more details). Assuming that epistasis e is weak (of order ϵ^2) relative to the strength of selection (of order ϵ), one obtains that the deterministic change in frequency at the modifier locus generated by epistasis is given by:

$$\Delta p_m \approx \sum_i a_i \, D_{mi} + \sum_{i<j} (a_i a_j + e) \, D_{mij} \tag{4}$$

where $a_i \approx -sh + 2e \sum_{j \neq i} p_j$ represents the effective strength of selection against the deleterious allele at locus i, p_j is the frequency of the deleterious allele at locus j and e is epistasis, while 2 and 3-locus linkage disequilibria are given by:

$$D_{ij} \approx \frac{e \, \tilde{p}_i \tilde{p}_j}{r_{ij} - a_i - a_j}, \tag{5}$$

$$D_{mij} \approx \frac{-\delta r_{ij} \, (h_m + d_m p_m) \, D_{ij}}{r_{mij} - a_i - a_j} \, p_m q_m, \quad D_{mi} \approx \sum_{j \neq i} \frac{a_j \, D_{mij}}{r_{mi} - a_i}, \tag{6}$$

with $d_m = 1 - 2h_m$, and where r_{mij} is the probability that at least one recombination event occurs between the three loci. In Figure 4, the effective strength of selection against deleterious alleles ($a_i < 0$, the same for all loci) is kept constant as epistasis varies, in order to maintain a constant average number of deleterious alleles per genome and constant additive variance in fitness. Th calculations detailed in the Supplementary Material show that for a given effective strength of selection a_i, the minimal possible value of epistasis e is $-{a_i}^2/(2U)$, while sh is given by $-(a_i + 2Ue/a_i)$, varying between 0 (when $e = -{a_i}^2/(2U)$ and selection is entirely due to epistatic interactions) and $-a_i$ (when $e = 0$).

Simulation model. The multilocus simulation program represents a population of N individuals carrying two copies of a linear chromosome. Each generation, the number of new deleterious mutations per chromosome is drawn from a Poisson distribution with parameter U, while the position of each new mutation on the chromosome is drawn from a uniform distribution between 0 and 1 (the number of loci at which mutations can occur is thus effectively infinite). The fitness of each individual is computed as $W = (1 - sh)^{n_{\mathrm{he}}} (1 - s)^{n_{\mathrm{ho}}} \exp(-cR)$, where n_{he} and n_{ho} are the numbers of heterozygous and homozygous mutations present in its genome, and R the chromosome map length coded by its recombination modifier locus. Gametes are produced by recombining the two chromosomes of the parent, the number of crossovers being drawn from a Poisson distribution with parameter R (the chromosome map length of the parent), while the position of each crossover along the chromosome is drawn from a uniform distribution (no interference). Map length R is determined by a modifier locus located at the mid-point of the chromosome, with an infinite number of possible alleles coding

for different values of R (if the individual is heterozygous at the modifier locus, R is given by the average between the values coded by its two alleles). Mutation occurs at the modifier locus at a rate μ per generation (generally set to 10^{-4}); when a mutation occurs, with probability 0.95 the value of the allele is multiplied by a random number drawn from a Gaussian distribution with average 1 and variance ${\sigma_{\mathrm{m}}}^2$ (generally set to 0.04), while with probability 0.05 a number drawn from a uniform distribution between -1 and 1 is added to the value of the allele (to allow for large effect mutations), the new value being set to zero if it is negative. During the first 20,000 generations, map length does not evolve and is fixed to $R = 1$; mutations are then introduced at the modifier locus and the population is let to evolve (generally during 5×10^6 generations, the value of the average map length usually reaching an equilibrium during the first 5×10^5 generations). The average map length, average fitness, average number of deleterious mutations per chromosome and number of fixed mutations are recorded every 500 generations (fixed mutations are removed from the population is order to minimize execution speed). Different modifications and extensions of the program were considered (including multiple modifier loci, multiple chromosomes, beneficial mutations and epistasis) and are described in the Supplementary Material.

Data availability. *Mathematica* notebooks showing derivations of the indirect selection gradient in the case of haploid and diploid populations, as well as the C++ simulation program are available from Dryad.

Acknowledgements. I thank Nick Barton, Thomas Lenormand, Henrique Teotónio and two anonymous reviewers for helpful comments, the bioinformatics and computing

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

REFERENCES

[1] Coop, G., Wen, X., Ober, C., Pritchard, J.K., and Przeworski, M. (2008). High resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science *319*, 1395–1398.

[2] Comeron, J.M., Ratnappan, R., and Bailin, S. (2012). The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. *8*, e1002905.

[3] Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A., Bragi Walters, G., Jonasdottir, A., Gylfason, A., Kristinsson, K.T., et al. (2010). Fine-scale recombination rate differences between sexes, populations and individuals. Nature *467*, 1099–1103.

[4] Kong, A., Thorleifsson, G., Frigge, M.L., Masson, G., Gudbjartsson, D.F., Villemoes, R., Magnusdottir, E., Olafsdottir, S.B., Thorsteinsdottir, U., and Stefansson, K. (2014). Common and low-frequency variants associated with genome-wide recombination rate. Nat. Genet. *46*, 11–18.

[5] Johnston, S.E., Bérénos, C., Slate, J., and Pemberton, J.M. (2016). Conserved genetic architecture underlying individual recombination rate variation in a wild population of Soay sheep (*Ovis aries*). Genetics *203*, 583–598.

[6] Samuk, K., Manzano-Winkler, B., Ritz, K.R., and Noor, M.A.F. (2020). Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. Curr. Biol. *30*, 1517–1528.

[7] True, J.R., Mercer, J.M., and Laurie, C.C. (1996). Differences in crossover frequency and distribution among three sibling species of Drosophila. Genetics *142*, 507–523.

[8] Ptak, S.E., Hinds, D.A., Koehler, K., Nickel, B., Patil, N., Ballinger, D.G., Przeworski, M., Frazer, K.A., and Pääbo, S. (2005). Fine-scale recombination patterns differ between chimpanzees and humans. Nat. Genet. *37*, 429–434.

[9] Winckler, W., Myers, S.R., Richter, D.J., Onofrio, R.C., McDonald, G.J., Bontrop, R.E., McVean, G.A.T., Gabriel, S.B., Reich, D., Donnelly, P., et al. (2005). Comparison of fine-scale recombination rates in humans and chimpanzees. Science *308*, 107–111.

[10] Smukowski, C.S. and Noor, M.A.F. (2011). Recombination rate variation in closely related species. Heredity *107*, 496–508.

[11] Brand, C.L., Cattani, M.V., Kingan, S.B., Landeen, E.L., and Presgraves, D.C. (2018). Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. Curr. Biol. *28*, 1289–1295.

[12] Dumont, B.L. and Payseur, B.A. (2007). Evolution of the genomic rate of recombination in mammals. Evolution *62*, 276–294.

[13] Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., and Smadja, C.M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Phil. Trans. Roy. Soc. (Lond.) B *372*, 20160455.

[14] Otto, S.P. and Lenormand, T. (2002). Resolving the paradox of sex and recombination. Nat. Rev. Genet. *3*, 252–261.

[15] Dapper, A.L. and Payseur, B.A. (2017). Connecting theory and data to understand recombination rate evolution. Phil. Trans. Roy. Soc. (Lond.) B *372*, 20160469.

[16] Ritz, K.R., Noor, M.A.F., and Singh, N.D. (2017). Variation in recombination rate: adaptive or not? Trends Genet. *33*, 364–374.

[17] Gonsalves, J., Sun, F. Schlegel, P.N., Turek, P.J., Hopps, C.V., Greene, C., Martin, R.H., and Reijo Pera, R.A. (2004). Defective recombination in infertile men. Hum. Mol. Genet. *13*, 2875–2883.

[18] Kong, A., Barnard, J., Gudbjartsson, D.F., Thorleifsson, G., Jonsdottir, G., Sigurdardottir, G., Richardsson, B., Jonsdottir, J., Thorgeirsson, T., Frigge, M.L., et al. (2004). Recombination rate and reproductive success in humans. Nat. Genet. *36*, 1203–1206.

[19] Ferguson, K.A., Chan Wong, E., Chow, V., Nigro, M., and Ma, S. (2007). Abnormal meiotic recombination in infertile men and its association with sperm aneuploidy. Hum. Mol. Genet. *16*, 2870–2879.

[20] Fledel-Alon, A., Wilson, D.J., Broman, K., Wen, X., Ober, C., Coop, G., and Przeworski, M. (2009). Broad-scale recombination patterns underlying proper disjunction in humans. PLoS Genet. *5*, e1000658.

[21] Ottolini, C.S., Newnham, L.J., Capalbo, A., Natesan, S.A., Joshi, H.A., Cimadomo, D., Griffin, D.K., Sage, K., Summers, M.C., Thornhill, A.R., et al. (2015). Genome-wide maps of recombination and chromosome segregation in hu-

498 man oocytes and embryos show selection for maternal recombination rates. Nat.
499 Genet. *47*, 727–737.

500 [22] Koehler, K.E., Scott Hawley, R., Sherman, S., and Hassold, T. (1996). Recombi-
501 nation and nondisjunction in human and flies. Hum. Mol. Genet. *5*, 1495–1504.

502 [23] Arbeithuber, B., Betancourt, A.J., Ebner, T., and Tiemann-Bogge, I. (2015).
503 Crossovers are associated with mutation and biased gene conversion at recombi-
504 nation hotspots. Proc. Natl. Acad. Sci. U. S. A. *112*, 2109–2114.

505 [24] Otto, S.P. and Michalakis, Y. (1998). The evolution of recombination in changing
506 environments. Trends Ecol. Evol. *13*, 145–151.

507 [25] Charlesworth, B. (1990). Mutation-selection balance and the evolutionary advan-
508 tage of sex and recombination. Genet. Res. *55*, 199–221.

509 [26] Barton, N.H. (1995). A general model for the evolution of recombination. Genet.
510 Res. *65*, 123–144.

511 [27] Hill, W.G. and Robertson, A. (1966). The effect of linkage on limits to artificial
512 selection. Genet. Res. *8*, 269–294.

513 [28] Felsenstein, J. (1974). The evolutionary advantage of recombination. Genetics
514 *78*, 737–756.

515 [29] Otto, S.P. and Barton, N.H. (1997). The evolution of recombination: removing
516 the limits to natural selection. Genetics *147*, 879–906.

517 [30] Otto, S.P. and Barton, N.H. (2001). Selection for recombination in small popula-
518 tions. Evolution *55*, 1921–1931.

519 [31] Barton, N.H. and Otto, S.P. (2005). Evolution of recombination due to random
520 drift. Genetics *169*, 2353–2370.

521 [32] Roze, D. and Barton, N.H. (2006). The Hill-Robertson effect and the evolution
522 of recombination. Genetics *173*, 1793–1811.

523 [33] Nei, M. (1967). Modification of linkage intensity by natural selection. Genetics
524 *57*, 625–641.

525 [34] Feldman, M.W., Christiansen, F.B., and Brooks, L.D. (1980). Evolution of recom-
526 bination in a constant environment. Proc. Natl. Acad. Sci. U. S. A. *77*, 4838–4841.

527 [35] Charlesworth, B. (1993). Directional selection and the evolution of sex and re-
528 combination. Genet. Res. *61*, 205–224.

529 [36] Iles, M.M., Walters, K., and Cannings, C. (2003). Recombination can evolve in
530 large finite populations given selection on sufficient loci. Genetics *165*, 2249–2258.

531 [37] Keightley, P.D. and Otto, S.P. (2006). Interference among deleterious mutations
532 favours sex and recombination in finite populations. Nature *443*, 89–92.

533 [38] Gordo, I. and Campos, P.R.A. (2008). Sex and deleterious mutations. Genetics
534 *179*, 621–626.

535 [39] Hartfield, M., Otto, S.P., and Keightley, P.D. (2010). The role of advantageous
536 mutations in enhancing the evolution of a recombination modifier. Genetics *184*,
537 1153–1164.

538 [40] Roze, D. (2014). Selection for sex in finite populations. J. Evol. Biol. *27*, 1304–
539 1322.

[41] Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. Genetics *134*, 1289–1303.

[42] Kimura, M. and Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. Genetics *54*, 1337–1351.

[43] Otto, S.P. and Payseur, B.A. (2019). Crossover interference: shedding light on the evolution of recombination. Ann. Rev. Gen. *53*, 19–44.

[44] Good, B.H., Walczak, A.M., Neher, R.A., and Desai, M.M. (2014). Genetic diversity in the interference selection limit. PLoS Genetics *10*, e1004222.

[45] Poon, A. and Otto, S.P. (2000). Compensating for our load of mutations: freezing the meltdown of small populations. Evolution *54*, 1467–1479.

[46] Charlesworth, B. (2015). Causes of natural variation in fitness: Evidence from studies of *Drosophila* populations. Proc. Natl. Acad. Sci. U. S. A. *112*, 1662–1669.

[47] Roze, D. and Michod, R.E. (2010). Deleterious mutations and selection for sex in finite, diploid populations. Genetics *184*, 1095–1112.

[48] Otto, S.P. (2003). The advantages of segregation and the evolution of sex. Genetics *164*, 1099–1118.

[49] Crow, J.F. and Kimura, M. (1970). An Introduction to Population Genetics Theory (New York: Harper and Row).

[50] Keightley, P.D. (2012). Rates and fitness consequences of new mutations in humans. Genetics *190*, 295–304.

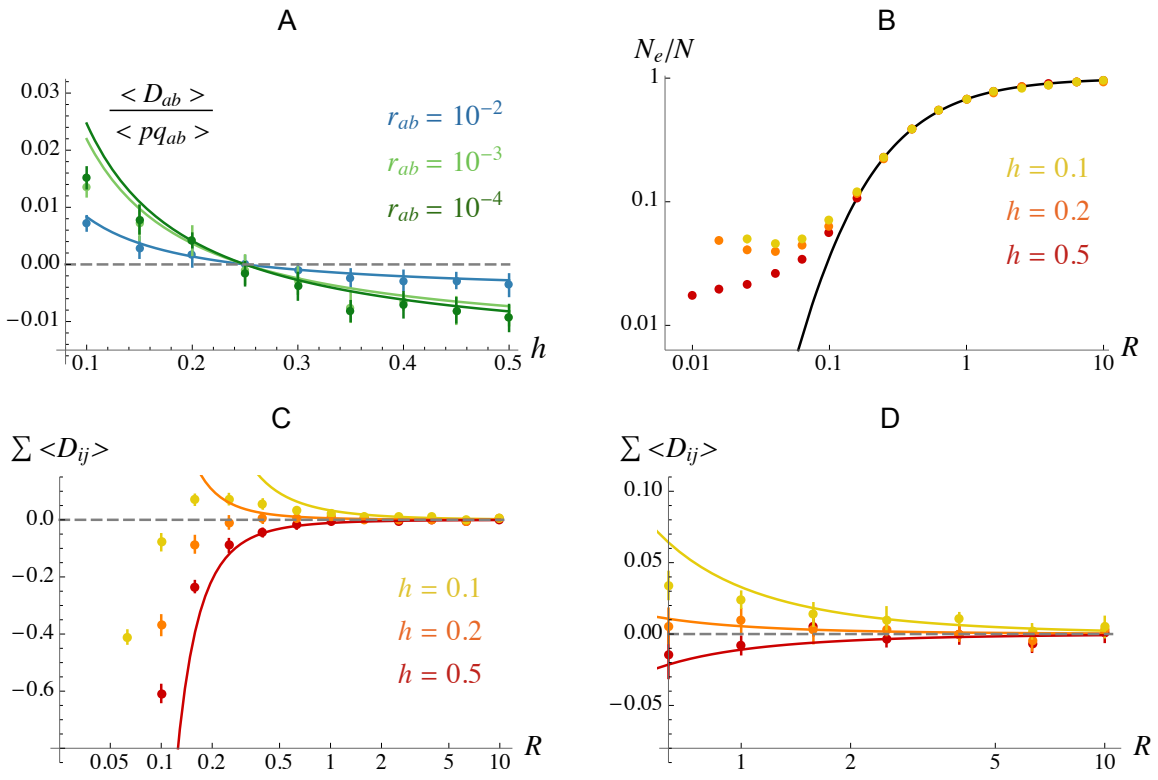
[51] Haenel, Q., Laurentino, T.G., Roesti, M., and Berner, D. (2018). Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol. Ecol. *27*, 2477–2497.

[52] Martin, G., Otto, S.P., and Lenormand, T. (2006). Selection for recombination in structured populations. Genetics *172*, 593–609.

[53] Roze, D. (2009). Diploidy, population structure and the evolution of recombination. Am. Nat. *174*, S79–S94.

[54] Roze, D. (2016). Background selection in partially selfing populations. Genetics *203*, 937–957.

[55] Hudson, R.R. and Kaplan, N.L. (1995). Deleterious background selection with recombination. Genetics *141*, 1605–1617.



571

572 **Figure 1.** A: average linkage disequilibrium between two deleterious alleles at mutation-
573 selection-drift balance (scaled by $\langle p_a q_a p_b q_b \rangle$) as a function of their dominance coeffi-
574 cient h, for different recombination rates r_{ab} between deleterious alleles (population
575 size $N = 1{,}000$, heterozygous effect of mutations sh kept constant at 0.01). Dots cor-
576 respond to two-locus simulation results (see Supplementary Material), and curves to
577 the analytical prediction $s^2 h\,(1-4h) \,/\, \left[2N\,(r_{ab}+2sh)^2\,(r_{ab}+3sh)\right]$ (from equation 5
578 in the Supplementary Material). B: effective population size N_e divided by the census
579 size N (on log scale) at the mid-point of a linear chromosome, as a function of the
580 chromosome map length R (on log scale), and for different values of the dominance
581 coefficient of deleterious alleles h (which occur at a rate $U = 0.2$ per chromosome).
582 The sh product is kept constant at 0.01. Curve: prediction from equation 22 in
583 the Supplementary Material; dots: multilocus simulation results (see Methods) with

584 $N = 10^4$. C, D: sum of all pairwise linkage disequilibria between deleterious alleles
585 as a function of the chromosome map length R, and for different values of h. Dots
586 correspond to simulation results (same simulations as in B) and curves to the analyt-
587 ical prediction $0.095\,(1-4h)\,\overline{n}^2/\,(N_{\mathrm{e}}R\,h)$, where $\overline{n} = U/\,(sh)$ is the mean number of
588 deleterious alleles per chromosome (equation 33 in the Supplementary Material). D
589 shows a magnification of the right part of C (higher values of R).

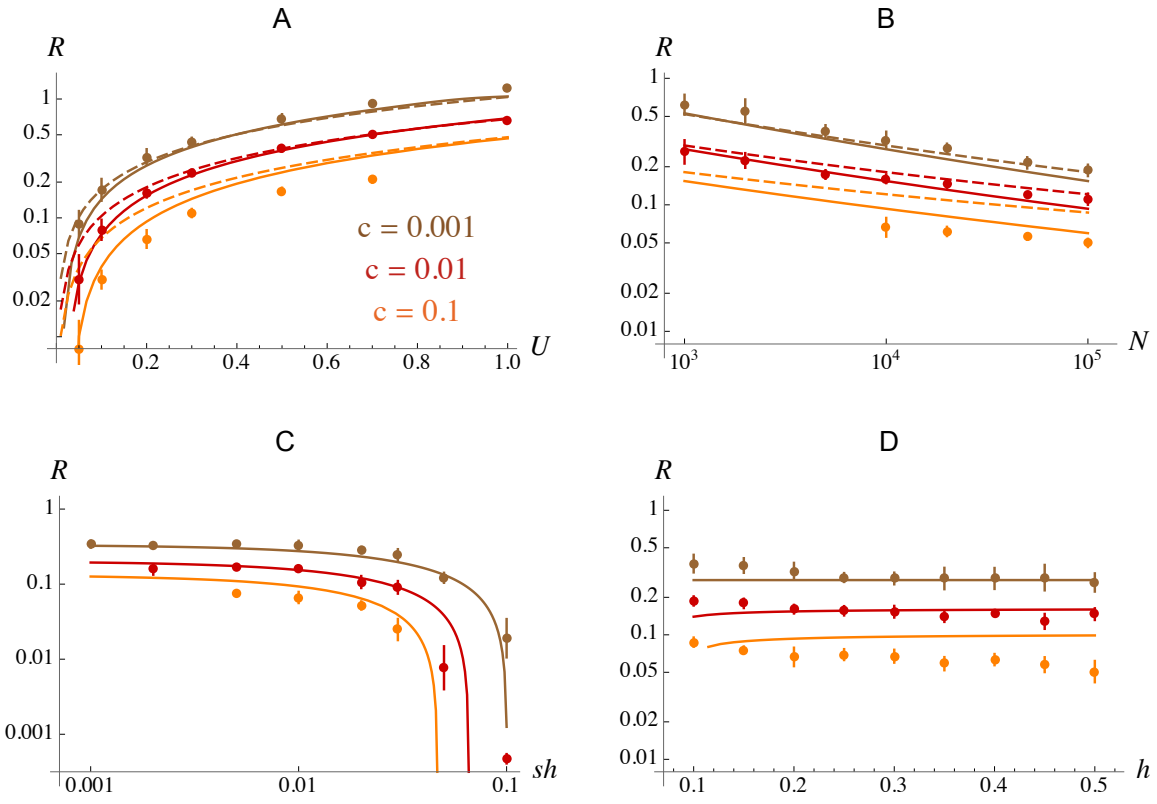
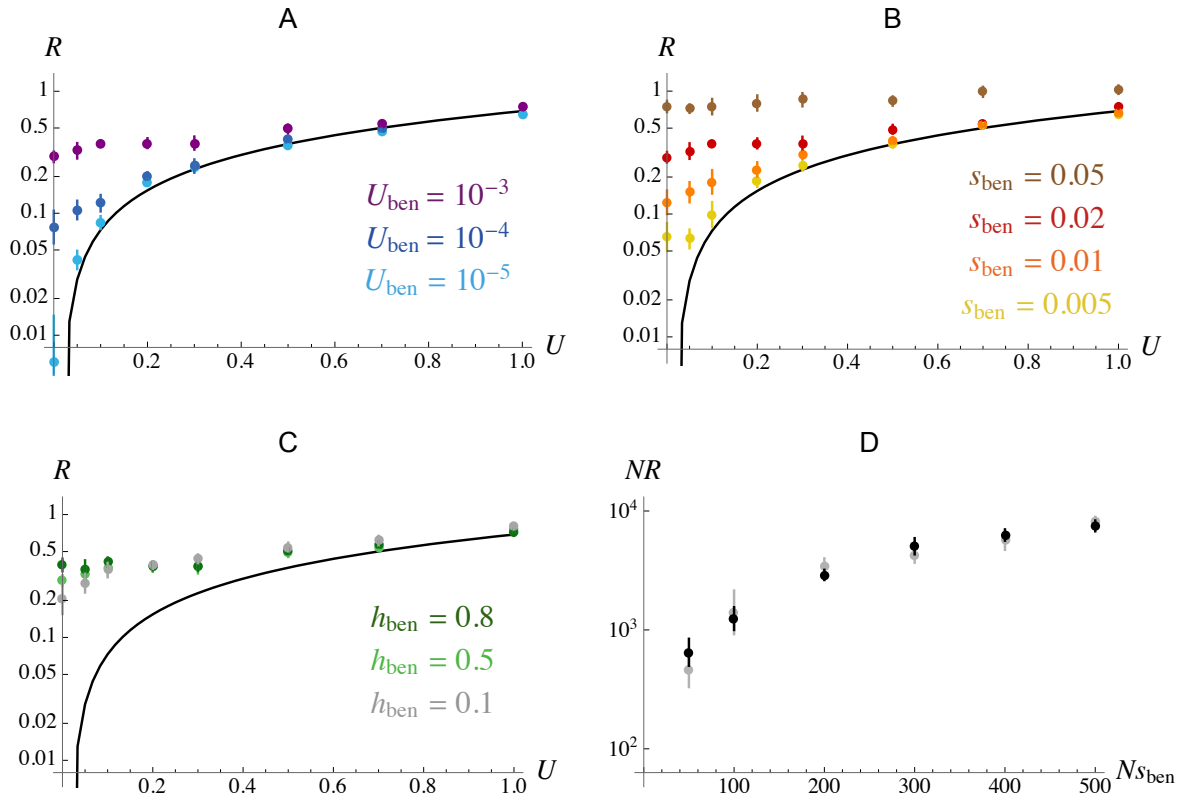


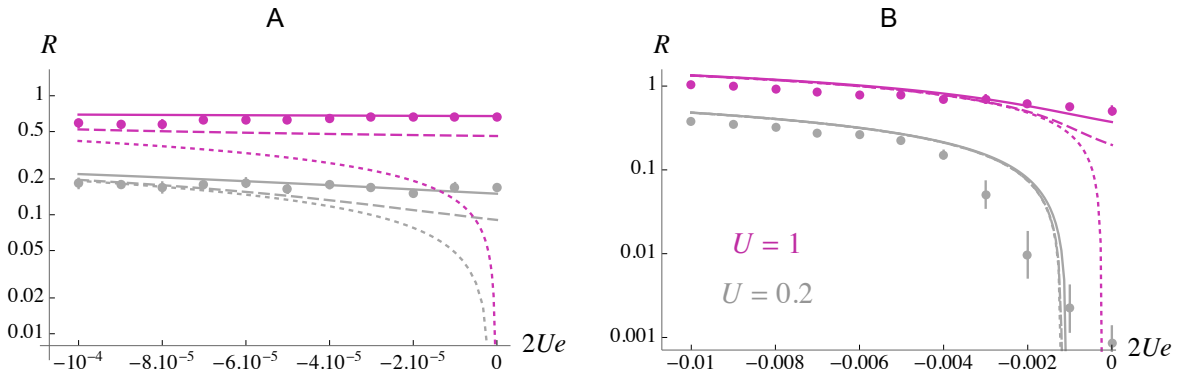
Figure 2. Equilibrium chromosome map length R (on log scale) for different values of the cost of recombination c, as a function of the deleterious mutation rate per haploid chromosome U (A), population size N (on log scale, B), fitness effect of heterozygous mutations sh (on log scale, C) and dominance coefficient h of deleterious alleles (D). Curves correspond to the analytical prediction obtained by extrapolation of the three-locus model (solid curves are obtained by numerical integration over the genetic map as explained in the Methods, while dashed curves in A, B correspond to the predictions from equation 1, also corresponding to the limits of the curves in C for low sh); dots correspond to simulation results (see Methods). Default parameter values are $N = 10^4$, $U = 0.2$, $s = 0.05$, $h = 0.2$. In C, h is kept constant at 0.2, while in D sh is kept constant at 0.01 (by adjusting s as h changes). In some of the simulations with $c = 0.1$,

deleterious alleles accumulated in the heterozygous state over time and the program had to be stopped, explaining why data points for high U, low N and low sh are missing (mutation accumulation also occurred for $c = 0.01$ and $sh = 0.001$ in C).



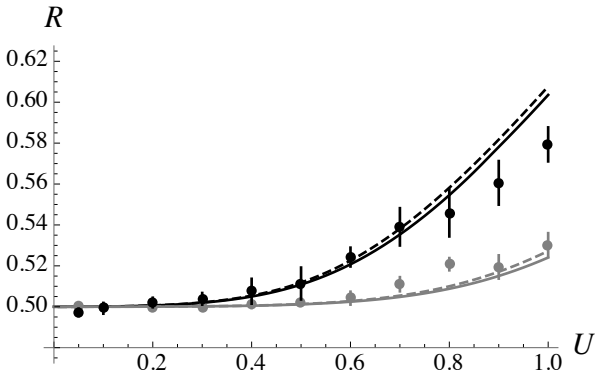
605

606 **Figure 3.** A, B, C: Equilibrium chromosome map length R (on log scale) as a function
607 of the deleterious mutation rate per haploid chromosome U, for different values of the
608 rate of beneficial mutation U_{ben} (A), fitness effect s_{ben} (B) and dominance coefficient
609 h_{ben} of beneficial alleles (C). The black curve corresponds to the analytical prediction
610 in the absence of beneficial allele ($U_{\text{ben}} = 0$). Default parameter values are $c = 0.01$,
611 $N = 10^4$, $s = 0.05$, $h = 0.2$, $U_{\text{ben}} = 10^{-3}$, $s_{\text{ben}} = 0.02$, $h_{\text{ben}} = 0.5$. In B the dominance
612 coefficient of beneficial mutations is fixed at $h_{\text{ben}} = 0.5$, while in C the product $s_{\text{ben}}h_{\text{ben}}$
613 is kept constant at 0.01 as h_{ben} varies (by adjusting s_{ben}). D: scaling with population
614 size: NR at equilibrium as a function of Ns_{ben}, for $NU_{\text{ben}} = 10$, $h_{\text{ben}} = 0.5$, $U = 0$
615 (no deleterious mutation) and $c = 0.01$. Black and grey dots correspond to simulation
616 results for $N = 10^4$ and $N = 10^5$, respectively.



617

618 **Figure 4.** Effect of negative epistasis: equilibrium chromosome map length R (on
619 log scale) as a function of the coefficient of epistasis between deleterious alleles (e)
620 multiplied by $2U$, for $U = 0.2$ (grey) and $U = 1$ (magenta). The overall strength of
621 selection against heterozygous mutations is kept constant (at 0.01 in A, and 0.1 in
622 B) by adjusting s as e varies (see Methods; note that for each strength of selection,
623 $2Ue$ cannot be lower than the left-most values on x-axes, for which $s = 0$). Curves
624 correspond to analytical predictions for $N = 10^4$ (solid), $N = 10^5$ (dashed) and for the
625 case of an infinite population ($N = \infty$, dotted); dots correspond to simulation results
626 for $N = 10^4$. Other parameter values are $c = 0.01$ and $h = 0.2$.



627

628 **Figure 5.** Equilibrium chromosome map length R as a function of the deleterious
629 mutation rate per haploid chromosome U, under direct stabilizing selection around
630 $R = 0.5$ (of the form $W_c = e^{-c(R-0.5)^2}$, with $c = 0.1$). Dashed curves correspond
631 to the predictions obtained by solving $-2c\,(R-0.5) + 1.8\,(N_{\rm e}U)^2\,/\,(N_{\rm e}R)^3 = 0$ with
632 $N_{\rm e} = N\,e^{-2U/R}$, while solid curves are obtained by numerical integration of the three-
633 locus model over the genetic map; dots correspond to simulation results. Parameter
634 values: $s = 0.05$, $h = 0.2$, $N = 10^4$ (black), $N = 10^5$ (grey).

SUPPLEMENTARY MATERIAL

Analytical three-locus model

Model parameters and assumptions. The model represents a diploid population of size N with discrete generations, and considers three loci: a recombination modifier locus (with two alleles M and m) and two selected loci (each with two alleles, A, a at the first locus and B, b at the second). Alleles a and b are deleterious, reducing fitness by a factor $1 - h_i s_i$ when heterozygous (where i stands for a or b), and $1 - s_i$ when homozygous. The effects of deleterious alleles are multiplicative across loci (no epistasis): for example, the fitness of a double heterozygote is $(1 - s_a h_a)(1 - s_b h_b)$. Mutations towards deleterious alleles occur at a rate u per generation. Back mutations are ignored, but their effect should be negligible as long as deleterious alleles remain rare in the population. Diploid parents produce a very large number of gametes (in proportion to their fitness) which fuse at random to produce zygotes (including the possibility of selfing), among which N are sampled randomly to form the next adult generation. At meiosis, the recombination rate between loci i and j in individuals with genotype MM, Mm and mm at the modifier locus is r_{ij}, $r_{ij} + h_m \delta r_{ij}$ and $r_{ij} + \delta r_{ij}$, respectively: δr_{ij} thus measures the effect of allele m on the recombination rate between loci i and j, while h_m is the dominance coefficient of this allele. Throughout the following, I will assume that the modifier only has weak effects on recombination rates, and compute results to the first order in δr_{ij}. Because recombination only has an effect in double heterozygotes (on the frequencies of the different types of gametes produced), recombination between loci m and a only matters in heterozygous individuals at locus

m (the recombination rate being $r_{ma} + h_m \delta r_{ma}$ in those individuals): therefore, δr_{ma} does not generate any selection for (or against) allele 1 at the modifier locus (since the recombination rate between loci m and a in MM and mm individuals is irrelevant), while r_{ma} and δr_{ma} should only affect the results through the quantity $r_{ma} + h_m \delta r_{ma}$ (and similarly for r_{mb}, δr_{mb}). Indirect selection at the modifier locus will only be driven by its effect on the recombination rate between the selected loci a and b, and the first-order approximation for the strength of indirect selection will thus be proportional to δr_{ab}. In this expression, additional terms in δr_{ij} will be neglected (as they would generate second-order terms in the modifier effect), and the final result will thus not depend on δr_{ma}, δr_{mb}. The results given below are valid for any ordering of the three loci along the chromosome (*i.e.*, $m - a - b$ or $a - m - b$). Because indirect selection on the modifier should mostly stem from its effect on closely linked selected loci, I will assume that recombination rates are small (of order ϵ, where ϵ is a small term), while the strength of selection against deleterious allele will also be assumed small (s_a, s_b are of order ϵ). Finally, I assume that drift is weak relative to selection ($1/N \ll \epsilon$) so that the frequency of each deleterious alleles stays close to its deterministic mutation-selection equilibrium value, and will derive all results to the first order in $1/N$. The per-locus mutation rate u is also assumed smaller than ϵ, so that the frequencies of deleterious alleles remain small. I assume throughout that h_a and h_b are significantly greater than zero, so that these frequencies are approximately $u/(s_i h_i)$.

Variables and general method. Because gametes fuse at random, the population can be censused in the haploid phase of the life cycle, just before gamete fusion. Defining X_j as an indicator variable that equals 1 in gametes carrying a lowercase allele (m, a

or b) at locus j, and 0 in gametes carrying an uppercase allele, the frequency of the lowercase allele at locus j is given by $p_j = \mathrm{E}[X_j]$ (where E stands for the average over all gametes), while the linkage disequilibrium between loci i and j (D_{ij}) corresponds to the covariance between X_i and X_j, that is, $D_{ij} = \mathrm{E}[(X_i - p_i)(X_j - p_j)]$. The three-locus linkage disequilibrium is defined similarly as $D_{mab} = \mathrm{E}[(X_m - p_m)(X_a - p_a)(X_b - p_b)]$ (e.g., [1]). Throughout the following, $\langle T \rangle$ will denote the expectation (over the stochastic process) of the quantity T: for example, $\langle D_{ab} \rangle$ is the average linkage disequilibrium between the selected loci at mutation-selection-drift balance.

The general method used to compute recursions on moments of allele frequencies and linkage disequilibria has been described elsewhere [2, 3] and will not be repeated here. General expressions have been implemented in a *Mathematica* notebook (available as Supplementary Material) that can be used to automatically generate recursions describing the effects of selection, recombination (with genotype-dependent recombination rates) and drift on any moment of allele frequencies and linkage disequilibria, to the first order in δr_{ij}, ϵ, $1/N$, $\tilde{p}_a$ and $\tilde{p}_b$ (the frequencies of deleterious alleles at mutation-selection balance). A separation of timescale argument (quasi-linkage equilibrium or QLE) can then be used to express all moments involving linkage disequilibria (LD) in terms of allele frequencies and of the parameters of the model [2, 3]. Indeed, the strength of recombination breaking linkage disequilibria is of order ϵ, while allele frequency changes are caused by drift and by the modifier effect, which are assumed much weaker ($1/N$, $\delta r_{ab} \ll \epsilon$); one may thus neglect changes in allele frequencies over the number of generations needed for moments involving LD to reach their equilibrium values, for the current allele frequencies. The results given below thus provide expressions for such moments in terms of the current allele frequencies of alleles m and

M in the population (p_m and q_m), and of the equilibrium frequencies of deleterious alleles $\tilde{p}_a$ and $\tilde{p}_b$. A similar method was used by Barton and Otto to compute the strength of indirect selection acting on the recombination modifier in a haploid model [4]; however, their derivations assume that selection is much weaker than recombination ($s_i \ll r_{jk}$ for all i, j, k), a necessary assumption for the QLE to hold in the case where beneficial alleles at loci A and B are sweeping through the population. The results shown below thus take similar forms as equations B3 and S2.3 in [4], except that they do not diverge when recombination rates tend to zero. As explained below, selection for recombination is generated by a variety of effects involving selection and drift, which are summarized in Figure S1.

Moments generated by selection and drift. Selection on the recombination modifier ultimately stems from the moments $\langle {D_{ab}}^2 \rangle$ and $\langle {D_{mab}}^2 \rangle$, which are generated by drift. At QLE and under the assumptions detailed above, they are given by:

$$\left\langle {D_{ab}}^2 \right\rangle \approx \frac{\tilde{p}_a \tilde{p}_b}{4N\left(r_{ab} + s_a h_a + s_b h_b\right)} \tag{1}$$

$$\left\langle {D_{mab}}^2 \right\rangle \approx \frac{\tilde{p}_a \tilde{p}_b p_m q_m}{4N\left(r_{mab} + s_a h_a + s_b h_b\right)} \tag{2}$$

where r_{mab} is the probability that at least one recombination event occurs between the three loci, given by $\left(r_{ma} + r_{mb} + r_{ab}\right)/2$ for any ordering of the loci along the chromosome. Equations 1 and 2 represent the fact that drift generates a variance in D_{ab} and D_{mab}. A positive value of D_{ab} corresponds to an excess of AB and ab haplotypes, while a negative value of D_{ab} corresponds to an excess of Ab and aB haplotypes. A positive value of D_{mab} means that allele m tends to be associated with a relative excess of AB and ab haplotypes (allele M being associated with a relative

excess of Ab and aB haplotypes), while a negative value of D_{mab} means the opposite.

The variances in D_{ab} and D_{mab} combine with the effect of selection against deleterious alleles to generate negative values of the moments $\langle p_a D_{ab}\rangle$, $\langle p_b D_{ab}\rangle$, $\langle D_{ma} D_{mab}\rangle$ and $\langle D_{mb} D_{mab}\rangle$:

$$\langle p_a D_{ab}\rangle \approx -\frac{s_b h_b}{r_{ab} + 2s_a h_a + s_b h_b} \left\langle {D_{ab}}^2 \right\rangle \tag{3}$$

$$\langle D_{ma} D_{mab}\rangle \approx -\frac{s_b h_b}{r_{ma} + r_{mab} + 2s_a h_a + s_b h_b} \left\langle {D_{mab}}^2 \right\rangle \tag{4}$$

$\langle p_b D_{ab}\rangle$ and $\langle D_{mb} D_{mab}\rangle$ being given by symmetric expressions. The negative value of $\langle p_a D_{ab}\rangle$ corresponds to the fact that when D_{ab} is positive, allele a is associated with the deleterious allele b, and thus tends to decrease in frequency (p_a decreases); conversely when $D_{ab} < 0$, allele a is associated with the better allele B, causing p_a to increase. Negative values of $\langle D_{ma} D_{mab}\rangle$, $\langle D_{mb} D_{mab}\rangle$ stem from the fact that when D_{mab} is positive, selection against deleterious alleles is more efficient in the background of allele m than in the background of allele M (because the variance in fitness is higher in the background of allele m), causing lower frequencies of deleterious alleles in the background of allele m (that is, D_{ma}, $D_{mb} < 0$). Conversely when D_{mab} is negative, selection against deleterious alleles is less efficient in the background of allele m, generating positive associations D_{ma}, D_{mb}.

The previous moments in turn generate the moments $\langle D_{ab}\rangle$ and $\langle D_{ma} D_{mb}\rangle$:

$$\langle D_{ab}\rangle \approx \frac{s_a (2h_a - d_a) \langle p_a D_{ab}\rangle + s_b (2h_b - d_b) \langle p_b D_{ab}\rangle}{r_{ab} + s_a h_a + s_b h_b} \tag{5}$$

where $d_a = 1 - 2h_a$, $d_b = 1 - 2h_b$, while

$$\langle D_{ma} D_{mb}\rangle \approx -\frac{s_a h_a \langle D_{ma} D_{mab}\rangle + s_b h_b \langle D_{mb} D_{mab}\rangle}{r_{ma} + r_{mb} + s_a h_a + s_b h_b}\,. \tag{6}$$

In the absence of the terms d_a, d_b representing dominance effects, $\langle D_{ab}\rangle$ would have the same sign as $\langle p_a D_{ab}\rangle$, $\langle p_b D_{ab}\rangle$ and would thus be negative. This corresponds to

the classical Hill-Robertson effect: the deleterious alleles are efficiently removed from the population when $D_{ab} > 0$ (causing D_{ab} to vanish), while they are maintained at higher frequencies when $D_{ab} < 0$ because selection is less efficient, causing D_{ab} to be negative on average. As shown by equation 5, partial recessivity of the deleterious alleles (d_a, $d_b > 0$) opposes this effect. This is due to the fact that the strength of selection against deleterious alleles becomes weaker as they become rarer (since they are less frequently present in the homozygous state), thus opposing the elimination of deleterious alleles from the population when $D_{ab} > 0$. According to equation 5, this effect prevails when dominance coefficients are less than 0.25, generating positive $\langle D_{ab} \rangle$. By contrast, the moment $\langle D_{ma} D_{mb} \rangle$ is always positive: as explained above, a positive value of D_{mab} generates a lower frequency of deleterious alleles in the background of allele m (D_{ma} and D_{mb} are both negative), while a negative value of D_{mab} generates a higher frequency of deleterious alleles in the background of allele m (D_{ma} and D_{mb} are both positive). Similarly, a positive covariance between p_m and D_{mab} is generated by the moments $\langle D_{ma} D_{mab} \rangle$, $\langle D_{mb} D_{mab} \rangle < 0$, from the fact that the frequency of allele m tends to increase when D_{ma}, $D_{mb} < 0$ (due to its association with the better alleles A and B):

$$\langle p_m D_{mab} \rangle \approx -\frac{s_a h_a \langle D_{ma} D_{mab} \rangle + s_b h_b \langle D_{mb} D_{mab} \rangle}{r_{mab} + s_a h_a + s_b h_b} . \tag{7}$$

Moments generated by the modifier effect. The effect of the recombination modifier combines with the effects just described to generate other moments, involving a single m index. We have in particular:

$$\langle D_{ab} D_{mab} \rangle \approx -\frac{\delta r_{ab} (h_m + d_m p_m) (\langle {D_{mab}}^2 \rangle + p_m q_m \langle {D_{ab}}^2 \rangle)}{r_{mab} + r_{ab} + 2 s_a h_a + 2 s_b h_b} \tag{8}$$

with $d_m = 1 - 2h_m$. Equation 8 shows that the variance in D_{ab} and the variance in D_{mab} both generate a negative covariance between D_{ab} and D_{mab} when allele m increases recombination ($\delta r_{ab} > 0$). Indeed, when $D_{ab} > 0$ the allele increasing recombination tends to produce more Ab, aB combinations, generating a negative D_{mab} (while when $D_{ab} < 0$ the allele increasing recombination becomes associated with a relative excess of AB, ab combinations). The effect of the variance in D_{mab} can be understood similarly: when $D_{mab} > 0$, the linkage disequilibrium between a and b is positive in the background of allele m, and negative in the background of allele M. The fact that linkage disequilibrium is eroded more rapidly in the background of allele m generates negative D_{ab} in the population (conversely, under negative D_{mab} the effect of the modifier generates positive D_{ab} in the population).

Moments $\langle D_{ma} D_{ab} \rangle$, $\langle D_{mb} D_{ab} \rangle$ are generated by the moment $\langle D_{ab} D_{mab} \rangle$ and by the effect of selection, as well as by the moments $\langle D_{ma} D_{mab} \rangle$, $\langle D_{mb} D_{mab} \rangle$ given by equation 4. We have:

$$\langle D_{ma} D_{ab} \rangle \approx -\frac{\delta r_{ab} \left(h_m + d_m p_m\right) \langle D_{ma} D_{mab} \rangle + s_b h_b \langle D_{ab} D_{mab} \rangle}{r_{ma} + r_{ab} + 2 s_a h_a + s_b h_b} \tag{9}$$

$\langle D_{mb} D_{ab} \rangle$ being given by a symmetric expression. Equation 4 above shows that $\langle D_{ma} D_{mab} \rangle$ is negative: when D_{mab} is negative, D_{ma} tends to be positive. As we have just seen, a negative D_{mab} leads to positive D_{ab} in the population (when allele m increases recombination), generating a positive covariance between D_{ma} and D_{ab}. Given that a negative D_{mab} leads to a positive D_{ma}, the negative moment $\langle D_{ab} D_{mab} \rangle$ also generates a positive $\langle D_{ma} D_{ab} \rangle$. Similarly, the moments $\langle p_a D_{mab} \rangle$, $\langle p_b D_{mab} \rangle$ are given by:

$$\langle p_a D_{mab} \rangle \approx -\frac{\delta r_{ab} \left(h_m + d_m p_m\right) p_m q_m \langle p_a D_{ab} \rangle + s_b h_b \langle D_{ab} D_{mab} \rangle}{r_{mab} + 2 s_a h_a + s_b h_b} \tag{10}$$

which can be understood in the same way (*e.g.*, positive D_{ab} generates negative D_{mab} through the modifier effect, and to a lower frequency of allele a through the effect of selection).

The average three-locus association $\langle D_{mab} \rangle$ plays a critical role in selection for recombination, and is generated by a variety of effects:

$$\begin{aligned}\langle D_{mab} \rangle \approx & \frac{1}{r_{mab} + s_a h_a + s_b h_b} \\ & \times \Big[\delta r_{ab} \left(h_m + d_m p_m \right) \left(\langle D_{ma} D_{mb} \rangle - p_m q_m \langle D_{ab} \rangle \right) \\ & + \delta r_{ab} d_m \left(1 - 2p_m \right) \left(\langle D_{ma} D_{mb} \rangle - \langle p_m D_{mab} \rangle \right) \\ & + s_a \left(2h_a - d_a \right) \langle p_a D_{mab} \rangle + s_b \left(2h_b - d_b \right) \langle p_b D_{mab} \rangle \\ & + 2 s_a h_a \langle D_{ma} D_{ab} \rangle + 2 s_b h_b \langle D_{mb} D_{ab} \rangle \Big].\end{aligned} \tag{11}$$

First, an increase in recombination caused by allele m tends to generate an association D_{mab} of opposite sign to D_{ab}: if the population harbors an excess of Ab and aB haplotypes, the allele increasing recombination will be more associated with AB and ab haplotypes. Second, the positive covariance between D_{ma} and D_{mb} (generated by the variance in D_{mab}, as shown above) tends to produce positive D_{mab}, by increased recombination between a and b in mm individuals (first term between the brackets of equation 11). This effect depends on dominance interactions between modifier alleles and on their frequencies: for example, it may be cancelled in the case of a rare dominant modifier increasing recombination, due to its effect in Mm individuals (second term between the brackets of equation 11). The effect of the moments $\langle p_a D_{mab} \rangle$ and $\langle p_b D_{mab} \rangle$ (third term between the brackets of equation 11) corresponds to the fact that situations in which $D_{mab} < 0$ tend to be transient, as the effect of the modifier generates positive D_{ab} leading to a better elimination of deleterious alleles, while situations in which $D_{mab} > 0$ tend to persist longer (causing positive D_{mab}, on average). As in

the case of $\langle D_{ab} \rangle$ discussed above, recessivity of deleterious alleles (d_a, $d_b > 0$) opposes this effect, by decreasing the strength of selection against rare deleterious alleles. Last, equation 11 shows that the moments $\langle D_{ma} D_{ab} \rangle$ and $\langle D_{mb} D_{ab} \rangle > 0$ also tend to produce positive D_{mab}. This effect is more difficult to understand intuitively. When D_{ma} is positive, D_{ab} tends to be also positive (as shown by equations 4, 8 and 9), leading to a relative excess of MAB and mab genotypes. The MAB genotype contributes negatively to D_{mab}, and the mab genotype positively. When D_{ma} is negative, D_{ab} tends to be also negative, leading to a relative excess of MaB and mAb genotypes; the MaB genotype contributes positively to D_{mab}, and the mAb genotype negatively. Selection tends to reduce the frequency of allele a, and one can show that the overall effect of this reduced frequency is to decrease the overall contribution of terms generating negative D_{mab}, while increasing the overall contribution of terms generating positive D_{mab} (so that the net effect is to produce positive $\langle D_{mab} \rangle$).

The moments $\langle p_a D_{mab} \rangle$ and $\langle D_{ma} D_{ab} \rangle$ also generate a negative covariance between p_a and D_{ma}:

$$\langle p_a D_{ma} \rangle \approx -\frac{s_b h_b \left(\langle p_a D_{mab} \rangle + \langle D_{ma} D_{ab} \rangle \right)}{r_{ma} + 2 s_a h_a} . \tag{12}$$

Indeed, positive values of D_{mab} generates negative values of D_{ma} (since selection against deleterious alleles is more efficient in the background of allele m when $D_{mab} > 0$), while positive values of D_{ab} lead to lower frequencies of deleterious alleles. Finally, the expected D_{ma} is given by:

$$\langle D_{ma} \rangle \approx -\frac{s_b h_b \langle D_{mab} \rangle + s_b d_b \langle p_b D_{mab} \rangle - s_a \left(2h_a - d_a \right) \langle p_a D_{ma} \rangle}{r_{ma} + s_a h_a} \tag{13}$$

(and similarly for $\langle D_{mb} \rangle$). Positive D_{mab} tends to generate negative D_{ma} as explained previously: selection against allele a is more efficient in the background of allele m,

when both deleterious alleles are positively associated in this background ($D_{mab} > 0$). When allele b is partially recessive, this effect is enhanced by the fact that the frequency of this allele in the population tends to be higher when $D_{mab} > 0$ (*i.e.*, $\langle p_b D_{mab} \rangle > 0$), leading to more efficient selection against it (term in $d_b \langle p_b D_{mab} \rangle$). Last, the negative covariance between p_a and D_{ma} (*i.e.*, $\langle p_a D_{ma} \rangle < 0$) indicates that $D_{ma} > 0$ when allele a tends to be more efficiently eliminated from the population, while $D_{ma} < 0$ when it reaches higher frequencies, causing the average value of D_{ma} to be negative. Again, recessivity of the deleterious allele a ($d_a > 0$) opposes this effect by sheltering it from selection at lower frequencies.

Change in frequency at the modifier locus. To leading order, the expected change in frequency of allele m is given by:

$$\langle \Delta p_m \rangle \approx -s_a h_a \langle D_{ma} \rangle - s_b h_b \langle D_{mb} \rangle \tag{14}$$

where $\langle D_{ma} \rangle$ and $\langle D_{mb} \rangle$ can be expressed in terms of p_m, $\tilde{p}_a$ and $\tilde{p}_b$ and of the different parameters of the model from equations 1 – 13 above. Note that all moments generated by the modifier effect are of order $\delta r_{ab} \tilde{p}_a \tilde{p}_b / (N\epsilon^2)$, so that the expected change in frequency of the modifier is of order $\delta r_{ab} \tilde{p}_a \tilde{p}_b / (N\epsilon)$. In the case of an additive recombination modifier ($h_m = 1/2$), it takes the form:

$$\langle \Delta p_m \rangle \approx \frac{\delta r_{ab}}{N} f(r_{ma}, r_{mb}, r_{ab}, s_a, h_a, s_b, h_b) \tilde{p}_a \tilde{p}_b p_m q_m \tag{15}$$

where f is a function of recombination rates, selection and dominance coefficients. This function contains terms involving only $s_a h_a$, $s_b h_b$, which always favor recombination, and terms in $d_a = 1 - 2h_a$, $d_b = 1 - 2h_b$ representing dominance effects. While dominance effects shown in Figure S1 (dashed lines) tend to disfavor recombination when

h_a, $h_b < 0.5$, the direct effect of $\langle p_b D_{mab} \rangle$ on $\langle D_{ma} \rangle$ (equation 13) favors recombination (see Figures S5, S6). Figure S6 shows that these different effects of dominance tend to compensate each other (at least as long as h is not too small and linkage not too tight), so that selection for recombination is often well predicted when ignoring terms in d_a, d_b altogether: indeed, the multilocus simulation results confirm that s and h mostly affect selection for recombination through the sh product (Figure 2). When terms in d_a, d_b are ignored, the strength of selection for recombination becomes equivalent as in a haploid model with a population size twice as large, and where the strength of selection against deleterious alleles is $s_a h_a$, $s_b h_b$ (a *Mathematica* notebook presenting the analysis of the haploid model is available as Supplementary Material).

Multilocus extrapolation

The results from the three-locus model can be extrapolated to the case of a modifier affecting the map length R of a linear chromosome, along which deleterious mutations occur at a given rate U per generation. For simplicity, I assume that the modifier is located at the mid-point of the chromosome, that the density of mutations and crossovers is uniform along the chromosome, and that all deleterious alleles have the same selection and dominance coefficients s and h. A direct cost of recombination c (representing for example an energetic cost associated with crossover formation) is introduced by assuming that the fitness of individuals is proportional to $W_c = \exp(-cR)$. Assuming that alleles at the modifier locus have additive effects on map length, so that the map lengths coded by MM, Mm and mm genotypes are R, $R+\delta R/2$ and $R+\delta R$, the change in frequency of allele m caused by the cost of recombination

is given by:

$$\Delta_{\text{cost}}\, p_m = \frac{\delta R}{2}\, \frac{d \ln W_c}{dR}\, p_m q_m = -\frac{\delta R\, c}{2}\, p_m q_m \tag{16}$$

to the first order in δR (e.g., [5]). From the previous results, the strength on indirect selection is given by:

$$\langle \Delta_{\text{ind}}\, p_m \rangle \approx -s\, h \sum_i \langle D_{mi} \rangle \tag{17}$$

where the sum is over all selected loci i, and where $\langle D_{mi} \rangle$ is given by equation 13, replacing A by i and B by j, and summing over all j. Neglecting the effects of dominance (terms in d_a, d_b in the equations above), and after replacing $\tilde{p}_i$, $\tilde{p}_j$ by $u/(s\,h)$, this yields an expression of the form:

$$\langle \Delta_{\text{ind}}\, p_m \rangle \approx \frac{1}{N\,(s\,h)^3} \sum_{i,j} \delta r_{ij}\, g(\rho_{mi}, \rho_{mj}, \rho_{ij})\, u^2\, p_m q_m \tag{18}$$

where $\rho_{mi} = r_{mi}/(s\,h)$, $\rho_{mj} = r_{mj}/(s\,h)$, $\rho_{ij} = r_{ij}/(s\,h)$, and where the function g can be found in the *Mathematica* notebook available as Supplementary Material. Because indirect selection mostly stems from tightly linked loci, recombination rates may be approximated by genetic distances between loci, and δr_{ij} by $\delta R\,(r_{ij}/R)$. In the case of a continuous genome, the sum in equation 18 becomes an integral, yielding:

$$\langle \Delta_{\text{ind}}\, p_m \rangle \approx \frac{\delta R\, s_{\text{ind}}}{2}\, p_m q_m \tag{19}$$

with:

$$\begin{aligned} s_{\text{ind}} = \frac{4U^2}{NR^3} \Bigg[& \int_0^{\frac{R}{2sh}} \int_0^{\frac{R}{2sh}} (x+y)\, g(x, y, x+y)\, dx\, dy \\ & + \int_0^{\frac{R}{2sh}} \int_0^{\frac{R}{2sh}} |x-y|\, g(x, y, |x-y|)\, dx\, dy \Bigg]. \end{aligned} \tag{20}$$

The first double integral in equation 20 corresponds to the overall effect of pairs of selected loci located on opposite sides of the modifier locus on the chromosome, and the second to the overall effect of pairs of loci located on the same side of the modifier

locus. These integrals can be evaluated numerically using the NIntegrate function of *Mathematica*, in order to compute s_{ind} for a range of values of R: s_{ind} is typically very small when R is large, and increases as R tends to zero. From equations 16 and 19, the evolutionarily stable map length corresponds to the value of R for which $s_{\text{ind}} = c$, which can be obtained by interpolation (see Supplementary Material). The terms in d_a, d_b apearing in equations 1 – 13 (effects of dominance) can be treated similarly, generating an extra term that takes the same form as equation 20 (with a different function of scaled recombination rates in the integrand) multiplied by $(1-2h)/h$ (see Supplementary Material). Although this term was included in the analyses, its effect is minor in most cases, and the curves appearing on Figures 2 – 4, S2 – S4 stay nearly unchanged when it is neglected.

When $R/(sh)$ is large, indirect selection mostly stems from the effect of loci located in the chromosomal vicinity of the modifier, and the integrals in equation 20 may be approximated by the same integrals taken between zero and infinity, yielding:

$$s_{\text{ind}} \approx 1.8 \frac{(NU)^2}{(NR)^3}. \tag{21}$$

More accurate results are obtained by taking into account the fact that the parameter N entering the equations above should be the effective population size N_{e} (determining the strength of drift in the population), which is reduced by the presence of segregating deleterious alleles (background selection, [6]). Although N_{e} varies along the chromosome, this variation should stay relatively minor as long as $R \gg sh$ (so that the reduction of N_{e} at a given locus is mostly due to mutations occurring in the chromosomal vicinity of this locus), and one may thus approximate N_{e} for all loci by

its value at the mid-point of the chromosome, given by equation 8 in [7]:

$$N_{\mathrm{e}} \approx N \exp\left[-\frac{2U}{R + 2s\,h}\right] \tag{22}$$

(note that U refers to the haploid chromosomal mutation rate in the present paper, and to the diploid mutation rate in [7], explaining the extra factor 2). From equations 20 – 22, one can notice that s_{ind} does not depend on N as long as the products NU, NR and Ns stay constant: one thus predicts that for a given value of c (the direct cost of recombination), the evolutionarily stable value of NR should be independent of N as long as NU and Ns stay constant. This prediction is confirmed by simulations (Figure S2).

The analysis above can be extended to multiple chromosomes. In the case of a local modifier solely affecting the map length of its own chromosome, the other chromosomes will only affect s_{ind} by reducing N_{e}, each additional chromosome introducing an extra e^{-8shU} factor to the background selection effect — where U is still the deleterious mutation rate per chromosome [8, 9]. In the case of a global modifier affecting the map length of all chromosomes, the extra component of indirect selection stemming from the effect of the modifier on each additional chromosome can be obtained by replacing r_{mi} and r_{mj} by 1/2 in the expressions given above. Although more accurate expressions may be obtained by repeating the previous analysis without the assumption that r_{mi} and r_{mj} are small, numerical results show that indirect selection caused by the effect of the modifier on other chromosomes is typically much weaker than indirect selection caused by its effect on its local chromosome, and may thus be neglected (see *Mathematica* notebook). Given that the reduction in N_{e} caused by other chromosomes is also usually much weaker than the effect of linked selected loci,

the overall strength of selection for recombination is generally well predicted by the single-chromosome model (see also [10]). This is confirmed by the simulation results shown on Figure S2.

Epistasis

The effect of negative epistasis between deleterious alleles can be included as follows. Assuming pairwise epistasis among mutations, the fitness of an individual may be written as:

$$W = (1 - sh)^{n_{\text{het}}} (1 - s)^{n_{\text{hom}}} (1 + e)^{n_{\text{pairs}}} \tag{23}$$

where e is epistasis, n_{het} and n_{hom} are the number of heterozygous and homozygous deleterious alleles in the genome of the individual, while n_{pairs} is the number of pairwise interactions between deleterious alleles at different loci, given by:

$$n_{\text{pairs}} = \frac{1}{2} n_{\text{het}} (n_{\text{het}} - 1) + 2 n_{\text{het}} n_{\text{hom}} + 2 n_{\text{hom}} (n_{\text{hom}} - 1) \tag{24}$$

(indeed, two pairwise interactions occur between a heterozygous locus and a homozygous locus for the deleterious allele, while four pairwise interactions occur between two homozygous mutations). Equation 23 neglects the potential effects of additive-by-dominance and dominance-by-dominance epistasis (e.g., [11, 12]), but these should stay minor as long as mating is random, so that deleterious alleles are mostly present in the heterozygous state.

Barton showed that indirect selection on a recombination modifier caused by epistasis can be expressed in terms of coefficients a_i and e_{ij}, representing the net strength of selection at locus i and the effect of (multiplicative) epistasis between loci

i and j [13]. Using the fitness function given by equation 23, these are approximately (e.g., [11, 12]):

$$a_i \approx -sh + 2e\sum_{j\neq i} p_j, \quad e_{ij} \approx e. \tag{25}$$

Extending Barton's analysis to the case of deleterious alleles at mutation-selection balance under weak recombination, linkage disequilibria generated by epistasis are given by:

$$D_{ij} \approx \frac{e_{ij}\,\tilde{p}_i\tilde{p}_j}{r_{ij} - a_i - a_j} \tag{26}$$

while D_{mij}, D_{mi} and the change in frequency of the modifier are given by:

$$D_{mij} \approx \frac{-\delta r_{ij}\,(h_m + d_m p_m)\,D_{ij}}{r_{mij} - a_i - a_j}\,p_m q_m, \quad D_{mi} \approx \sum_{j\neq i}\frac{a_j\,D_{mij}}{r_{mi} - a_i}, \tag{27}$$

$$\Delta p_m \approx \sum_i a_i\,D_{mi} + \sum_{i<j}(a_i a_j + e_{ij})\,D_{mij}. \tag{28}$$

Equation 28 can be integrated over the genetic map as we have seen previously, in order to quantify the overall effect of epistatic interactions on indirect selection acting on the recombination modifier (see *Mathematica* notebook).

In Figure 4, the effective strength of selection against deleterious alleles ($a_i < 0$, the same for all loci) is kept constant as epistasis varies (in order to maintain a constant number of deleterious alleles per genome and constant additive variance in fitness). From equation 25, we have $a_i \approx -sh + 2e\overline{n}$, where $\overline{n} = \sum_i p_i$ is the mean number of deleterious alleles per chromosome. Furthermore, the change in p_i due to selection is $\Delta_{\rm sel}p_i = a_i\,p_i q_i \approx a_i\,p_i$, so that $\Delta_{\rm sel}\overline{n} \approx a_i\,\overline{n}$. Given that the change in $\overline{n}$ due to mutation is U, the value of $\overline{n}$ at mutation – selection balance is obtained by solving $-sh\overline{n} + 2e\overline{n}^2 = -U$, yielding

$$\overline{n} \approx \frac{1}{4e}\left[sh - \sqrt{(sh)^2 - 8Ue}\right], \quad a_i \approx -\frac{1}{2}\left[sh + \sqrt{(sh)^2 - 8Ue}\right]. \tag{29}$$

For a given effective strength of selection a_i, the minimal possible value of epistasis is thus $-{a_i}^2/(2U)$, while sh is given by $-(a_i + 2Ue/a_i)$, varying between 0 (when $e = -{a_i}^2/(2U)$ and selection is entirely due to epistatic interactions) and $-a_i$ (when $e = 0$). Finally, from equation 23 and neglecting the effect of linkage disequilibria between selected loci, one obtains that mean fitness at mutation – selection balance is approximately:

$$\overline{W} \approx \exp\left[-2sh\overline{n} + 2e\overline{n}^2\right] \approx \exp\left[-2U\left(1 + \frac{Ue}{{a_i}^2}\right)\right] \tag{30}$$

varying between $\exp(-U)$ (when e takes its minimal value of $-{a_i}^2/(2U)$ for a given effective strength of selection a_i) and $\exp(-2U)$ (when $e = 0$).

Simulation programs

Two-locus model. Two-locus simulations were used to check the analytical prediction for the average linkage disequilibrium between deleterious alleles $\langle D_{ab} \rangle$, given by equation 5 (Figure 1A). For this, the program (written in C++) represents the effects of mutation (also including back mutation at a rate $v = u/10$), drift, selection and recombination on two-locus genotype frequencies over a large number of generations (10^9). D_{ab} among gametes and $p_a q_a p_b q_b$ were measured every 10 generations to obtain averages over 10^8 data points, and the results were averaged over 10 replicate simulations.

Baseline multilocus model. The multilocus simulation program represents a population of N individuals carrying two copies of a linear chromosome. Each generation, the number of new deleterious mutations per chromosome is drawn from a

Poisson distribution with parameter U, while the position of each new mutation on the chromosome is drawn from a uniform distribution between 0 and 1 (the number of loci at which mutations can occur is thus effectively infinite). In practice, each chromosome is represented by a table of values representing the positions at which deleterious alleles are present. The fitness of each individual is computed as $W = (1 - sh)^{n_{\mathrm{he}}} (1 - s)^{n_{\mathrm{ho}}} \exp(-cR)$, where n_{he} and n_{ho} are the numbers of heterozygous and homozygous mutations present in its genome, and R the chromosome map length coded by its recombination modifier locus (see below). To form each new individual of the next generation, two parents are sampled according to the following procedure: an individual is sampled at random among the N potential parents; if a random number (drawn from a uniform distribution between 0 and 1) is lower than its fitness (divided by the maximum fitness of all potential parents), the individual is retained and produces a recombinant gamete, otherwise another individual is sampled until the test is satisfied (by doing so, the expected number of offspring of an individual is $W/\overline{W}$, where $\overline{W}$ is the average fitness of the population). Gametes are produced by recombining the two chromosomes of the parent, the number of crossovers being drawn from a Poisson distribution with parameter R (the chromosome map length of the parent), while the position of each crossover along the chromosome is drawn from a uniform distribution. Map length R is determined by a modifier locus located at the mid-point of the chromosome, with an infinite number of possible alleles coding for different values of R (if the individual is heterozygous at the modifier locus, R is given by the average between the values coded by its two alleles). Mutation occurs at the modifier locus at a rate μ per generation (generally set to 10^{-4}); when a mutation occurs, with probability 0.95 the value of the allele is multiplied by a random number

drawn from a Gaussian distribution with average 1 and variance ${\sigma_{\rm m}}^2$ (generally set to 0.04), while with probability 0.05 a number drawn from a uniform distribution between -1 and 1 is added to the value of the allele (to allow for large effect mutations), the new value being set to zero if it is negative. During the first 20,000 generations, map length does not evolve and is fixed to $R = 1$; mutations are then introduced at the modifier locus and the population is let to evolve (generally during 5×10^6 generations, the value of the average map length usually reaching an equilibrium during the first 5×10^5 generations). The average map length, average fitness, average number of deleterious mutations per chromosome and number of fixed mutations are recorded every 500 generations (fixed mutations are removed from the population is order to minimize execution speed). Error bars in the figures are obtained by splitting the simulation results into batches of 5×10^5 generations (removing the first batch during which the average map length reaches its equilibrium) and computing the variance of batch averages (error bars correspond to ± 1.96 S.E.). Different modifications and extensions of the program were considered, as described below.

Effective population size and sum of pairwise LD. In the simulation results shown in Figure 1 (B, C, D), the modifier locus was replaced by a neutral locus (with an infinite number of possible alleles, and mutation rate $\mu = 0.001$) in order to estimate the effective population size, $N_{\rm e}$ being estimated by $\pi / [4\mu (1 - \pi)]$, where π is the expected heterozygosity at the neutral locus measured over 10^6 generations, with one point every 100 generations. The sum of all pairwise linkage disequilibria between deleterious alleles (shown in Figure 1C, D) is obtained from the frequencies of those alleles in the population and from the variance in the number of mutations

per gamete $\mathrm{Var}(n)$. Indeed, we have:

$$\mathrm{Var}(n) = \sum_i p_i q_i + \sum_{i \neq j} D_{ij} \tag{31}$$

where the first sum is over all loci segregating for deleterious alleles, and the second sum over all pairs of such loci, so that $\sum_{i \neq j} \langle D_{ij} \rangle$ is given by $\langle \mathrm{Var}(n) \rangle - \sum_i \langle p_i q_i \rangle$ (the last sum is approximately equal to the mean number of mutations per chromosome, but was computed exactly in order to obtain exact measures in regimes where deleterious alleles may reach high frequencies). In Figure 1C, D, $\sum_{i \neq j} \langle D_{ij} \rangle$ is compared with the analytical prediction obtained by integrating equation 5 over the chromosome. Assuming that $\sum_{i \neq j} \langle D_{ij} \rangle$ is mostly generated by pair of loci at small genetic distances (so that recombination rates can be approximated by genetic distances), and after some rearranging, one obtains:

$$\sum_{i \neq j} \langle D_{ij} \rangle \approx \frac{U^2 (1-4h)}{N_\mathrm{e} R^2 s h^2} \int_0^{\frac{R}{sh}} \frac{\frac{R}{sh} - x}{(x+2)^2 (x+3)} dx \tag{32}$$

with $N_\mathrm{e} \approx N \exp[-2U/(R+2sh)]$. When $R \gg sh$, the integral in equation 32 may be approximated by $\frac{R}{sh} \int_0^\infty dx / \left[(x+2)^2 (x+3)\right] \approx 0.095 R/(sh)$, yielding:

$$\sum_{i \neq j} \langle D_{ij} \rangle \approx \frac{0.095}{N_\mathrm{e} R} \frac{1-4h}{h} \overline{n}^2 \tag{33}$$

with $N_\mathrm{e} \approx N \exp[-2U/R]$, and where $\overline{n} = U/(sh)$ is the mean number of mutations per chromosome. Equations 32 and 33 yield nearly undistinguishable curves on Figures 1C and 1D (not shown).

Distribution of fitness effects of deleterious alleles. The effect of variable selection coefficients of deleterious alleles (Figure S2 C, D) was explored by modifying the program in order to associate a value of s drawn from a log-normal distribution

to each new mutation: the value of $\ln s$ is drawn from a Gaussian distribution with variance σ^2 and average equal to $\ln \overline{s} - \sigma^2/2$ (so that the average selection coefficient stays equal to $\overline{s}$, set to 0.05). The dominance coefficient of deleterious alleles stayed fixed at $h = 0.2$ in these simulations.

Multiple modifier loci. The baseline model was extended to an arbitrary number of modifier loci n_{m} affecting map length, evenly spaced along the chromosome. The effects of the different modifier loci on R are assumed additive (R being set to zero when the sum is negative). At the start of the simulation the allelic value at each modifier locus is fixed at $R_{\mathrm{init}}/n_{\mathrm{m}}$, with $R_{\mathrm{init}} = 1$. In order to maintain the same mutational variance on R independently of the number of modifier loci, the total mutation rate at modifier loci is fixed at $\mu = 10^{-4}$, while each mutation adds a term RX to the allelic value coded by the allele before mutation, where R is the genetically encoded map length (before mutation) and X a random number drawn from a Gaussian distribution with average 0 and variance ${\sigma_{\mathrm{m}}}^2$ (set to 0.04).

Multiple chromosomes. The standard model was also extended to the more realistic case of a genome made of several chromosomes (Figure S2 E, F), considering either a single global modifier affecting the map length of all chromosomes (located at the mid-point of one of the chromosomes) or local modifiers affecting the map length of their own chromosome (as is the single-chromosome program). In both cases, the fitness of an individual is given by $W = (1 - sh)^{n_{\mathrm{he}}} (1 - s)^{n_{\mathrm{ho}}} \exp(-cR_{\mathrm{tot}})$, where n_{he} and n_{ho} are the numbers of heterozygous and homozygous mutations present in its genome, while R_{tot} corresponds to its total genome map length (the sum of all chromosome

map lengths).

Beneficial mutations. Beneficial alleles were introduced in the standard model in order to explore the effect of the interaction between beneficial and deleterious mutations on the evolution of recombination (Figures 3 and S3). In that case, beneficial mutations with selection and dominance coefficients s_{ben} and h_{ben} (and with multiplicative effects across loci) occur at a rate U_{ben} per chromosome per generation (an additional table is associated to each chromosome, containing the positions of the different beneficial alleles present on the chromosome). Once a beneficial allele has reached fixation, it is removed from the population in order to minimize execution speed.

Epistasis. Epistasis is introduced into the baseline program by implementing the fitness function given by equation 23.

REFERENCES

[1] Slatkin, M. (1972). On treating the chromosome as the unit of selection. Genetics *72*, 157–168.

[2] Roze, D. (2014). Selection for sex in finite populations. J. Evol. Biol. *27*, 1304–1322.

[3] Roze, D. (2016). Background selection in partially selfing populations. Genetics *203*, 937–957.

[4] Barton, N.H. and Otto, S.P. (2005). Evolution of recombination due to random drift. Genetics *169*, 2353–2370.

[5] Gervais, C. and Roze, D. (2017). Mutation rate evolution in partially selfing and partially asexual organisms. Genetics *207*, 1561–1575.

[6] Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. Genetics *134*, 1289–1303.

[7] Hudson, R.R. and Kaplan, N.L. (1995). Deleterious background selection with recombination. Genetics *141*, 1605–1617.

[8] Robertson, A. (1961). Inbreeding in artificial selection programmes. Genet. Res. *2*, 189–194.

[9] Charlesworth, B. (2012). The effects of deleterious mutations on evolution at linked sites. Genetics *190*, 5–22.

[10] Otto, S.P. and Barton, N.H. (1997). The evolution of recombination: removing the limits to natural selection. Genetics *147*, 879–906.

[11] Roze, D. (2009). Diploidy, population structure and the evolution of recombination. Am. Nat. *174*, S79–S94.

[12] Abu Awad, D. and Roze, D. (2020). Epistasis, inbreeding depression and the evolution of self-fertilization. Evolution *74*, 1301–1320.

[13] Barton, N.H. (1995). A general model for the evolution of recombination. Genet. Res. *65*, 123–144.

SUPPLEMENTARY FIGURES

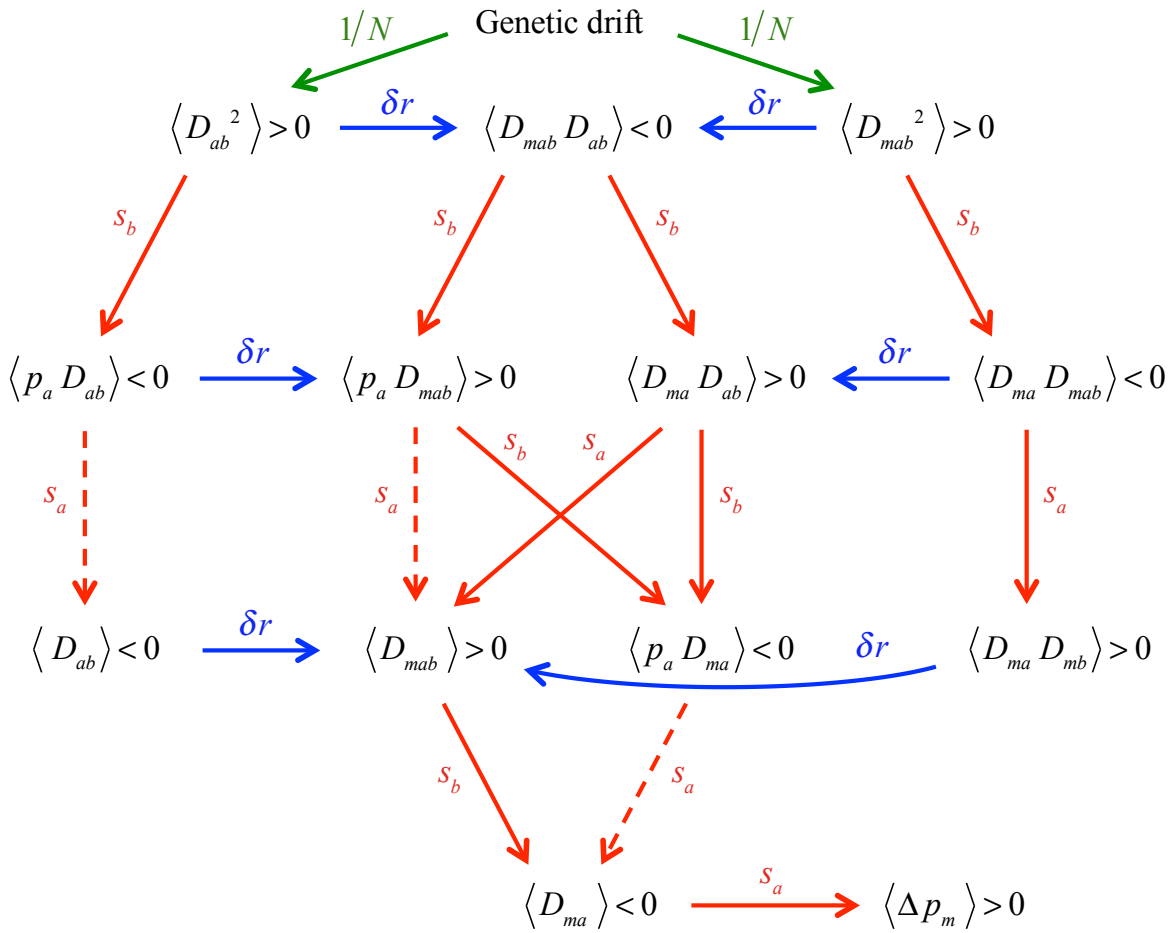


Figure S1. Summary of the different effects generating indirect selection for recombination due to interference between selected loci (three-locus model). Green arrows correspond to the effects of drift, red arrows to the effect of selection against deleterious alleles, and blue arrows to the effect of the recombination modifier. Note that symmetric moments (swapping a and b indices) are generated by the same processes, generating $\langle D_{mb} \rangle < 0$. The signs of the different moments are given in the case where the dominance coefficient of allele a (h_a) is greater than 0.25: when $h_a < 0.25$, the

contributions of dashed arrows reverses, *i.e.*, $\langle p_a D_{ab} \rangle$ tends to produce positive $\langle D_{ab} \rangle$, while $\langle p_a D_{mab} \rangle$ tends to produce negative $\langle D_{mab} \rangle$, and $\langle p_a D_{ma} \rangle$ tends to produce positive $\langle D_{ma} \rangle$. When allele b is partially recessive, the moment $\langle p_b D_{mab} \rangle$ also contributes to producing negative $\langle D_{ma} \rangle$ (not shown here). When $h_m \neq 1/2$ (non-additive modifier), $\langle D_{mab} \rangle$ is also affected by the moment $\langle p_m D_{mab} \rangle$ (not shown).

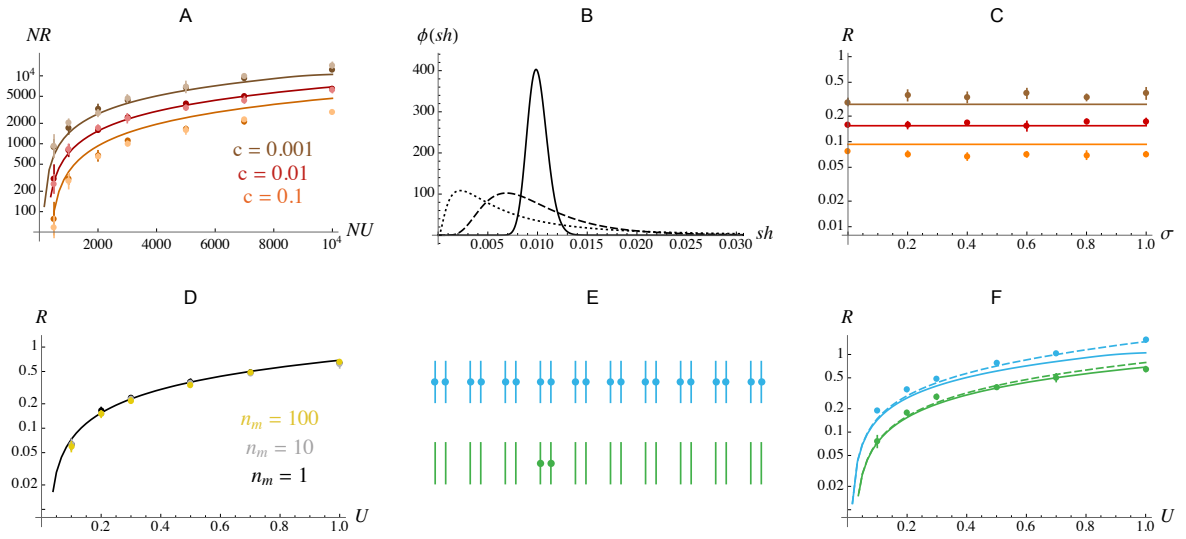
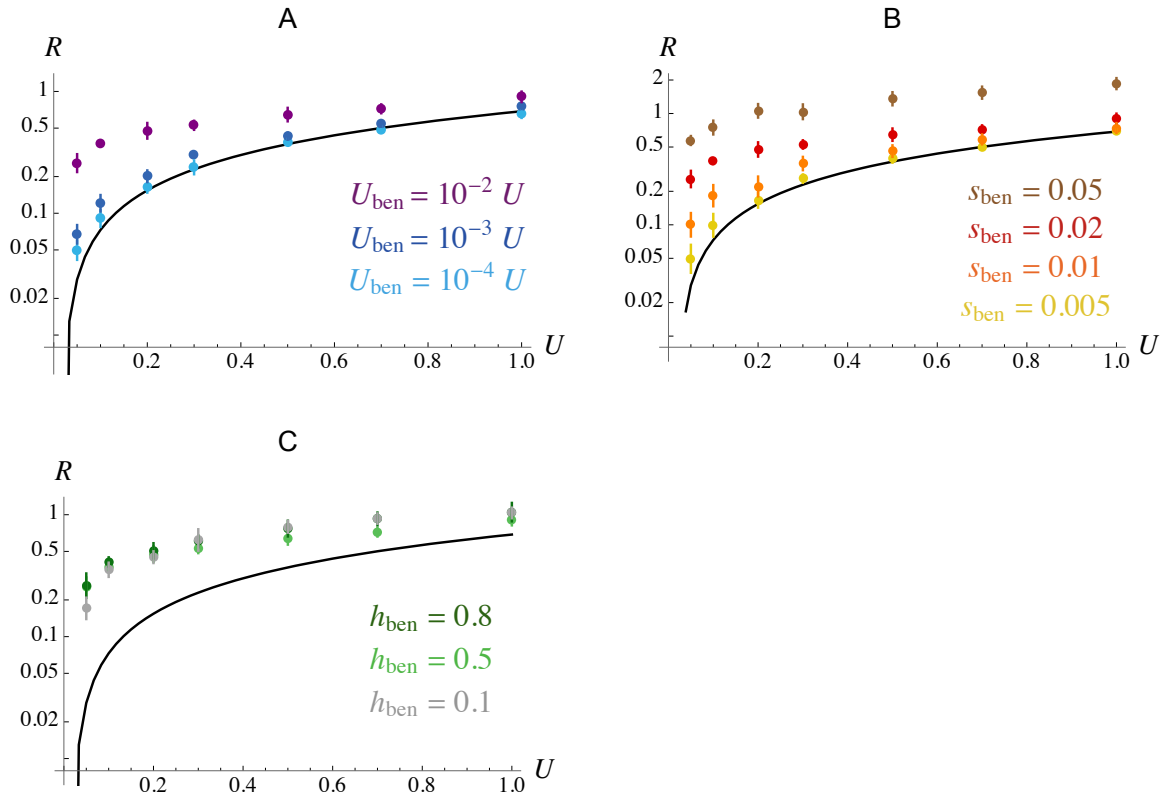


Figure S2. A: scaling with population size: NR at equilibrium as a function of NU, for $Ns = 500$, $h = 0.2$ and different values of the cost of recombination c. Curves correspond to analytical predictions, dots to simulation results with $N = 10^4$, and lighter dots to simulation results with $N = 10^5$ (keeping NU and Ns constant). B, C: distribution of fitness effects of deleterious alleles: B shows the p.d.f. of sh for three values of σ (the standard deviation of $\ln s$, see Methods): $\sigma = 0.1$ (plain), 0.5 (dashed) and 1 (dotted); C shows the equilibrium chromosome map length R as a function of σ for different values of the cost of recombination c (parameter values as in Figure 2). D: Increasing the number of recombination modifier loci does not affect the equilibrium map length: dots show simulation results with different numbers n_{m} of modifier loci (with additive effects, see Supplementary Material), for $c = 0.01$ and other parameter values as in Figure 2. E, F: extension to 10 chromosomes: blue dots in F correspond to simulations in which each chromosome carries a local modifier affecting the map length of its own chromosome (as illustrated in E), and green dots to simulations in which one global modifier affects the map length of all chromosomes. Parameter values are the

same as in Figure 2, with $c = 0.001$. Because c is multiplied by the total map length of the genome in the fitness function (see Supplementary Material), the strength of direct selection acting on local modifiers is c, but $n_{\mathrm{chr}}\, c$ in the case of a global modifier, where n_{chr} is the number of chromosomes (here 10). Solid curves show predictions from the single-chromosome model with $c = 0.001$ (blue) and $c = 0.01$ (green); dashed curves show predictions from the 10 chromosomes model with $c = 0.001$, in the case of one local modifier per chromosome (blue) and one global modifier (green). The small increase in the strength of indirect selection (compared with the single-chromosome model) is caused by the decrease in N_{e} due to extra chromosomes, and to the effect of the modifier on other chromosomes in the case of a global modifier.

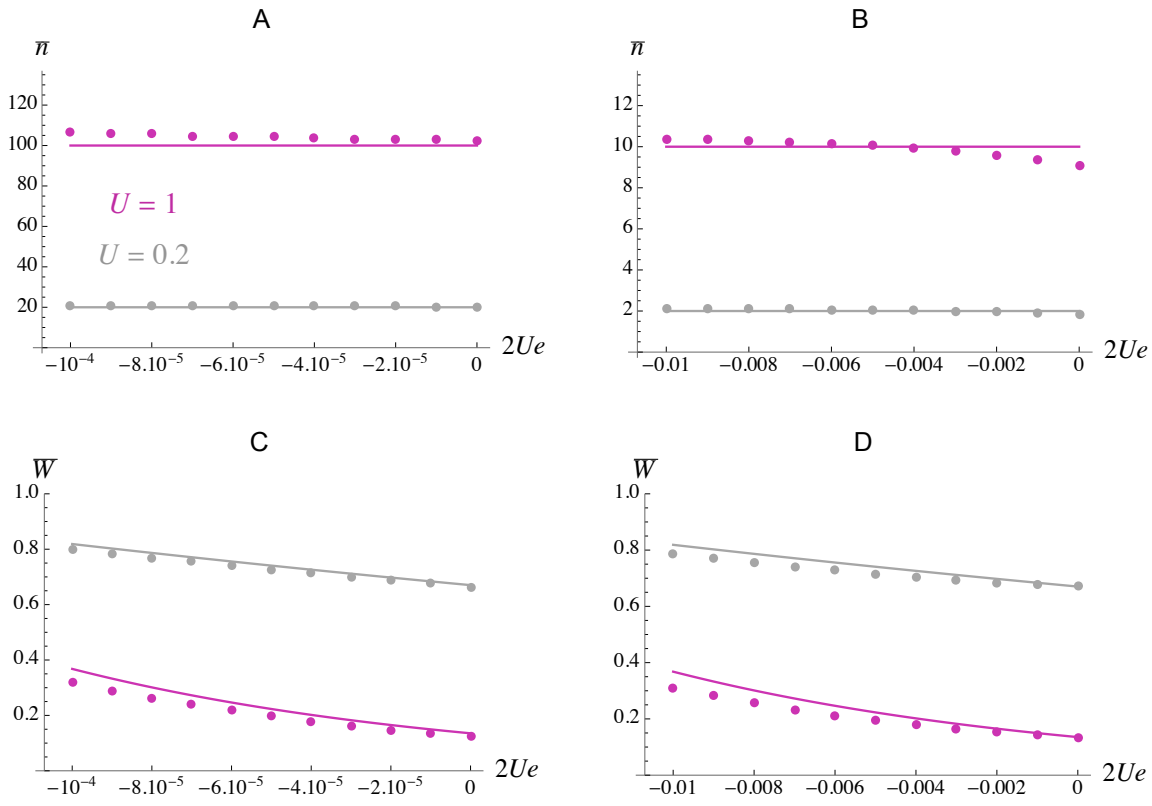


41

42 **Figure S3.** Same as Figure 3A, 3B, 3C when the beneficial mutation rate U_{ben} is

43 proportional to the deleterious mutation rate U. Parameter values are as in Figure 3,

44 $U_{\text{ben}} = 10^{-2}\,U$ in B, C.



45

46 **Figure S4.** Mean number of deleterious mutations per chromosome $\overline{n}$ (A, B) and mean
47 fitness $\overline{W}$ (C, D) as a function of the coefficient of epistasis between deleterious alleles
48 (e) multiplied by $2U$, for the same parameter values as in Figure 4 (the overall strength
49 of selection against heterozygous mutations is $-a_i = 0.01$ in A, C, and $-a_i = 0.1$ in
50 B, D). Dots correspond to simulation results, and lines to $U/(-a_i)$ in A, B, and to
51 equation 30 from the Supplementary Material in C, D.

52

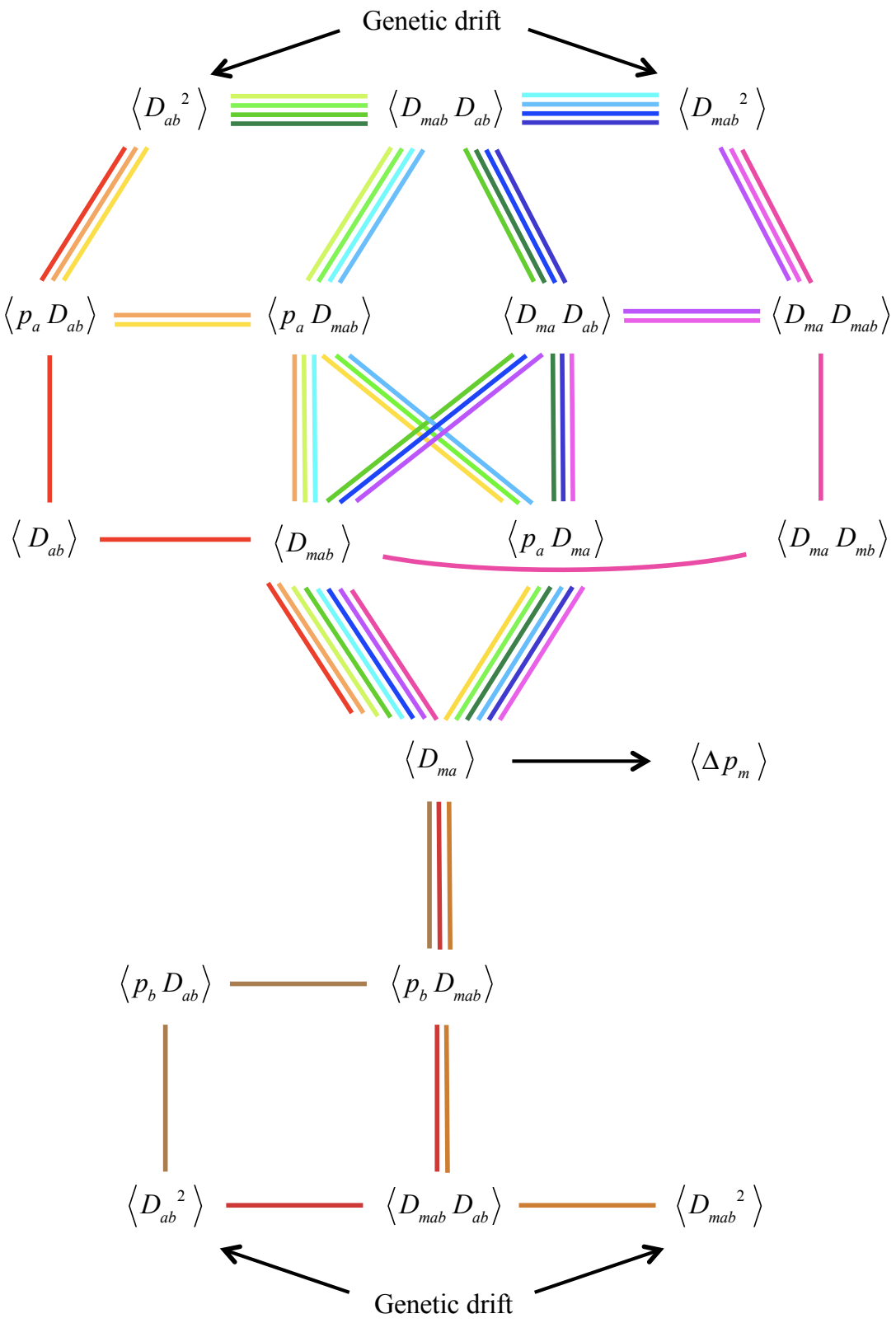

Genetic drift
$\langle D_{ab}^{\,2} \rangle$
$\langle D_{mab} D_{ab} \rangle$
$\langle D_{mab}^{\,2} \rangle$
$\langle p_a D_{ab} \rangle$
$\langle p_a D_{mab} \rangle$
$\langle D_{ma} D_{ab} \rangle$
$\langle D_{ma} D_{mab} \rangle$
$\langle D_{ab} \rangle$
$\langle D_{mab} \rangle$
$\langle p_a D_{ma} \rangle$
$\langle D_{ma} D_{mb} \rangle$
$\langle D_{ma} \rangle$
$\langle \Delta p_m \rangle$
$\langle p_b D_{ab} \rangle$
$\langle p_b D_{mab} \rangle$
$\langle D_{ab}^{\,2} \rangle$
$\langle D_{mab} D_{ab} \rangle$
$\langle D_{mab}^{\,2} \rangle$
Genetic drift

Figure S5 (previous page). The different paths generating indirect selection on the recombination modifier (through $\langle D_{ma} \rangle$), shown by different colors (same color code as in Figure S6). The effect of the moment $\langle p_b D_{mab} \rangle$ involving dominance at locus b (see equation 13 in the Supplementary Material), which was not shown on Figure S1, is now represented by the three brown paths at the bottom.

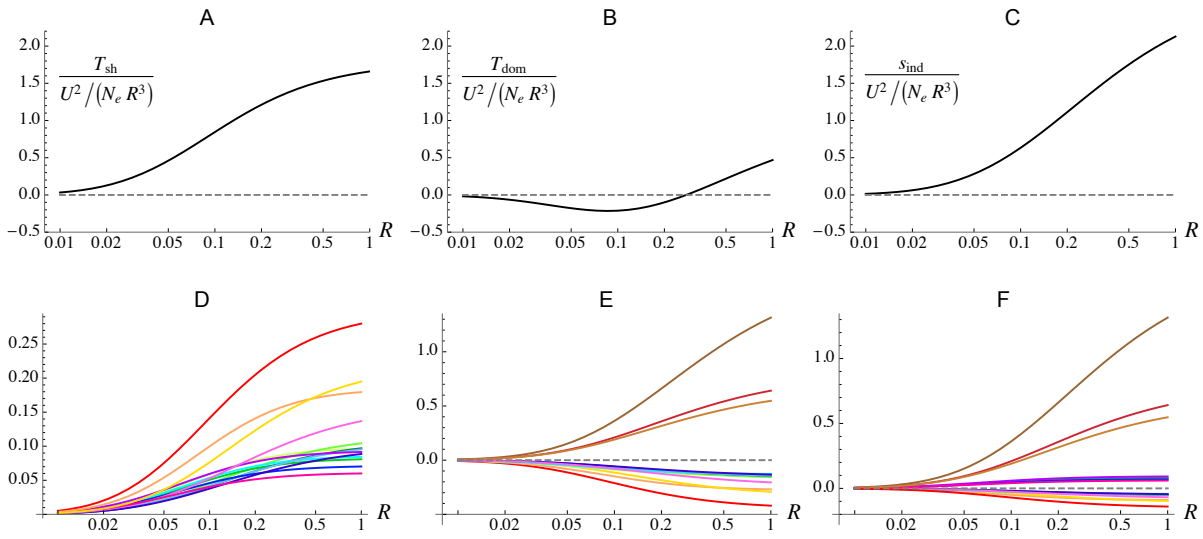


Figure S6. A: general contribution of terms in sh (T_{sh}) to indirect selection for recombination, divided by $U^2/\left(N_{\mathrm{e}}R^3\right)$. B: general contribution of terms generated by dominance (T_{dom}), corresponding to terms in $s\,(1-2h)$. C shows the overall strength of indirect selection ($s_{\mathrm{ind}} = T_{\mathrm{sh}} + T_{\mathrm{dom}}$). D, E and F show the contributions of the different paths highlighted in Figure S5 to T_{sh}, T_{dom} and s_{ind}, respectively (same color code as in Figure S5). Parameter values: $s = 0.05$, $h = 0.2$. Note that in the absence of dominance but for the same value of sh (*i.e.*, for $s = 0.02$, $h = 0.5$) A and D would stay unchanged, while the curves in B and E would vanish. For $h = 0.2$, the net effect of the path involving $\langle D_{ab} \rangle$ is to disfavor recombination due to dominance effects (red curves in E, F), but this path makes the strongest contribution to T_{sh} (A). Finally, note that the fact that indirect selection seems to vanish for low R is due to the scaling in $1/R^3$ (without the scaling, results for high R are difficult to see).