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

Denis Roze[a,b,1]

[a]International Research Laboratory 3614, CNRS, 29680 Roscoff, France; and [b]Station Biologique de Roscoff, Sorbonne Université, 29680 Roscoff, France

Edited by Nicholas H. Barton, Institute of Science and Technology, Klosterneuburg, Austria, and accepted by Editorial Board Member Anne M. Villeneuve March 26, 2021 (received for review November 10, 2020)

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 cross-overs at meiosis) remains difficult. This article derives a simple 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 $(N_eU)^2/(N_eR)^3$, where N_e is the effective population size, U is the deleterious mutation rate per chromosome, and R is 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.

evolution of recombination | genetic architecture | genetic interference | meiosis | multilocus population genetics

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 a long time 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 cross-overs. In particular, it is usually thought that, in most species, at least one cross-over 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 cross-overs 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 (14, 24). In particular, recombination increases the efficiency of natural selection in the presence of negative linkage disequilibria (LDs) 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., refs. 25, 26 and 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 than when loci are loosely linked). 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 three, four, 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

Significance

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 mathematical quantification of the selective advantage of a mutation increasing the genetic map length (average number of cross-overs 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.

Author contributions: D.R. designed research, performed research, analyzed data, and wrote the paper.

The author declares no competing interest.

This article is a PNAS Direct Submission. N.H.B. is a guest editor invited by the Editorial Board.

[1]Email: roze@sb-roscoff.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2022805118/-/DCSupplemental.

Published May 3, 2021.

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 of $1-hs$ 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 *Materials and Methods* and in *SI Appendix*). In the first step, the strength of indirect selection acting at the recombination modifier locus due to interference between two deleterious alleles (labeled a and b) at different loci is quantified (the expression obtained staying valid even when selected loci are tightly linked). In the 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, multilocus 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 *Materials and Methods*. A direct fitness cost associated with recombination is introduced in the simulation program, by assuming that the fitness of individuals is proportional to $\exp(-c\,R)$ (c may thus be considered as the fitness cost per cross-over). 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 LD 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 (Fig. 1*A*). As explained in *SI Appendix*, 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 LDs 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., ref. 41 and Fig. 1*B*). Fig. 1 *C* and *D* shows 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 LDs 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)

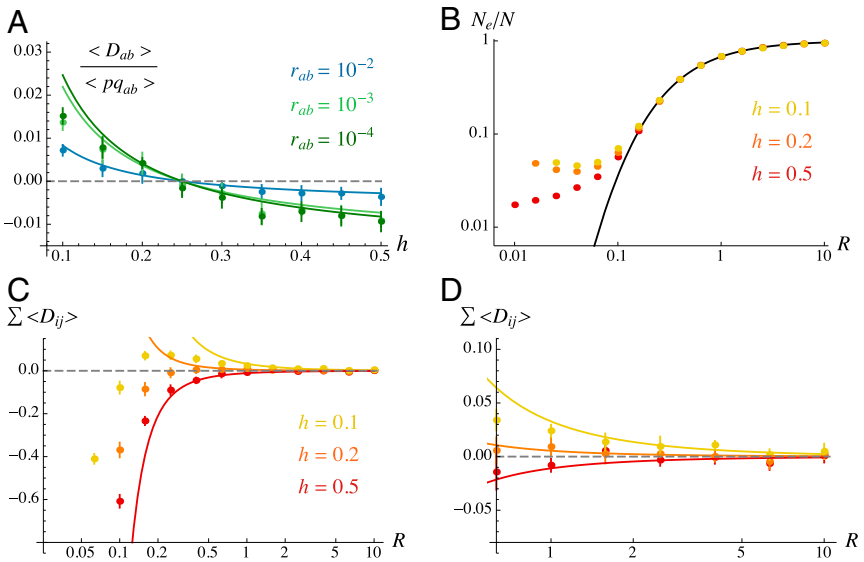

Fig. 1. (*A*) Average LD between two deleterious alleles at mutation–selection–drift balance (scaled by $\langle p_a q_a p_b q_b\rangle$) as a function of their dominance coefficient *h*, for different recombination rates r_{ab} between deleterious alleles (population size $N=1{,}000$, heterozygous effect of mutations *sh* kept constant at 0.01). Dots correspond to two-locus simulation results (see *SI Appendix*), and curves correspond to the analytical prediction $s^2h\,(1-4h)/\left[2N\,(r_{ab}+2sh)^2\,(r_{ab}+3sh)\right]$ (from *SI Appendix*, Eq. **S5**). (*B*) Effective population size N_e divided by the census size *N* (on log scale) at the midpoint of a linear chromosome, as a function of the chromosome map length *R* (on log scale), and for different values of the dominance coefficient of deleterious alleles *h* (which occur at a rate $U=0.2$ per chromosome). The *sh* product is kept constant at 0.01. Curve: prediction from *SI Appendix*, Eq. **S22**; dots: multilocus simulation results (see *Materials and Methods*) with $N=10^4$. (*C* and *D*) Sum of all pairwise LDs between deleterious alleles as a function of the chromosome map length *R*, and for different values of *h*. Dots correspond to simulation results (same simulations as in *B*), and curves correspond to the analytical prediction $0.095\,(1-4h)\bar{n}^2/\,(N_e R\,h)$, where $\bar{n}=U/\,(sh)$ is the mean number of deleterious alleles per chromosome (*SI Appendix*, Eq. **S33**). *D* shows a magnification of the right part of *C* (higher values of *R*). In this and the following figures, error bars measure ± 1.96 SE.

affecting pairwise LD, which 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 *SI Appendix*, Fig. 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 generates 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 *SI Appendix*, Figs. S5 and 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 (as long as h is not too small). 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 on Dryad (42)]. 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 $\delta R\, s_{\mathrm{ind}}$, with

$$s_{\mathrm{ind}} \approx 1.8\, \frac{(N_e U)^2}{(N_e R)^3} \qquad [1]$$

independently of s and h, and with $N_e \approx N \exp(-2U/R)$ 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 *Materials and Methods* and *SI Appendix*).

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, $s_{\mathrm{ind}} = c$. Fig. 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 *SI Appendix*, 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 *SI Appendix*, Fig. S2*A*. Fig. 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 *SI Appendix*, Fig. S2*C*.

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, Eq. **1** should still hold when map length is a polygenic trait coded by several loci located at various positions along the chromosome. Indeed, *SI Appendix*, Fig. S2*D* 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 (*SI Appendix*, Fig. S2 *E* and *F*). Indeed, the evolution of a local recombination modifier affecting the map length of its own chromosome is not affected much

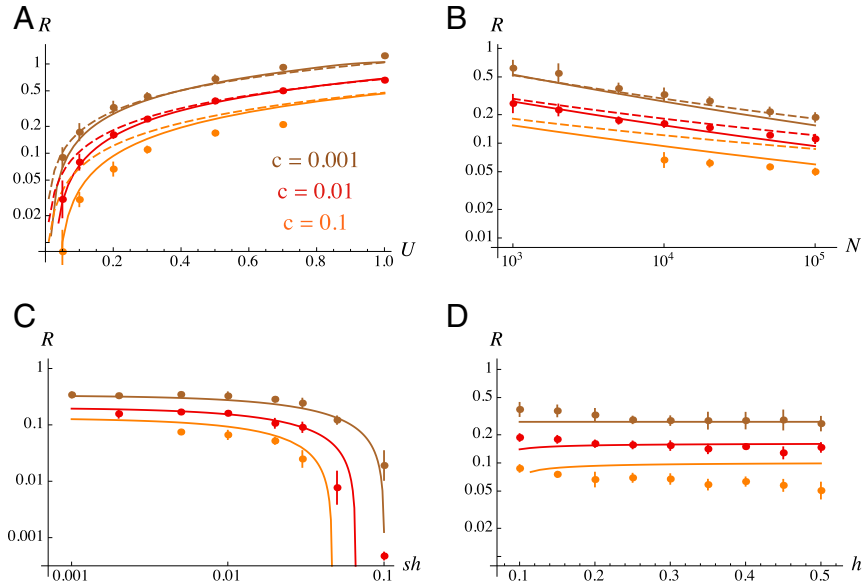

Fig. 2. Equilibrium chromosome map length R (on log scale) for different values of the cost of recombination c, as a function of (*A*) the deleterious mutation rate per haploid chromosome U, (*B*) population size N (on log scale), (*C*) fitness effect of heterozygous mutations sh (on log scale), and (*D*) dominance coefficient h of deleterious alleles. 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 *Materials and Methods*, while dashed curves in *A* and *B* correspond to the predictions from Eq. **1**, also corresponding to the limits of the curves in *C* for low sh); dots correspond to simulation results (see *Materials and Methods*). Default parameter values are $N = 10^4$, $U = 0.2$, $s = 0.05$, and $h = 0.2$. 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*).

by the presence of other chromosomes (as their only effect is to cause a modest reduction in N_e, by a factor of $\sim\exp(-8sh\,U)$ 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 Eq. **1**.

Including Beneficial Mutations. Obtaining analytical predictions for the equilibrium map length when beneficial and deleterious mutations cooccur 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 toward beneficial alleles, all with the same selection and dominance coefficients s_{ben} and h_{ben}).

As shown by Fig. 3, higher rates of recombination evolve when beneficial mutations cooccur 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 (*SI Appendix*, Fig. 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 $s_{\text{ben}}h_{\text{ben}}$ (Fig. 3*B*), while their dominance coefficient has only a little effect as long as $s_{\text{ben}}h_{\text{ben}}$ stays constant (Fig. 3*C*). 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} (Fig. 3*D*).

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 of $1+e$ (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 Fig. 4 (see *Materials and 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 Fig. 4. As can be seen from Fig. 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. *SI Appendix*, Fig. 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 (43). As shown by Fig. 4*B*, 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 the Introduction, mechanistic constraints

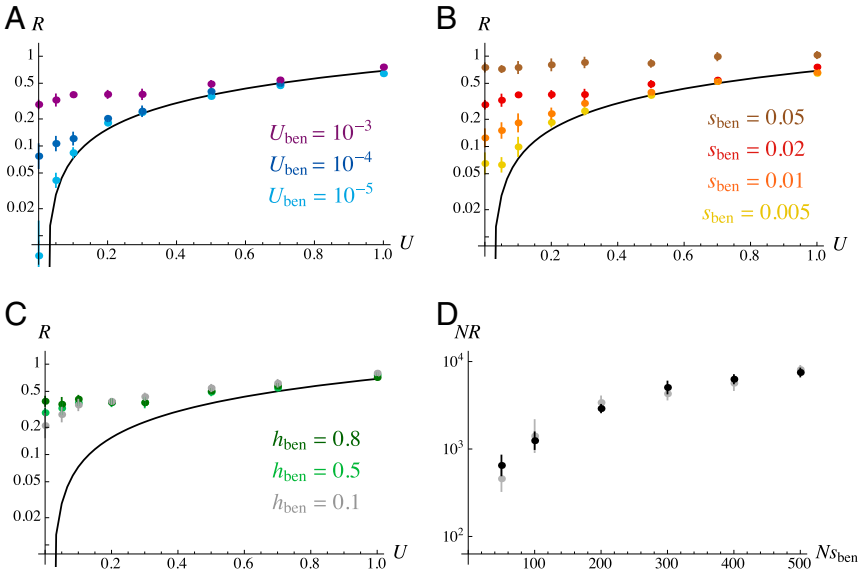

Fig. 3. (*A–C*) Equilibrium chromosome map length R (on log scale) as a function of the deleterious mutation rate per haploid chromosome U, for different values of the rate of beneficial mutation U_{ben} (*A*), fitness effect s_{ben} (*B*), and dominance coefficient h_{ben} (*C*) of beneficial alleles.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

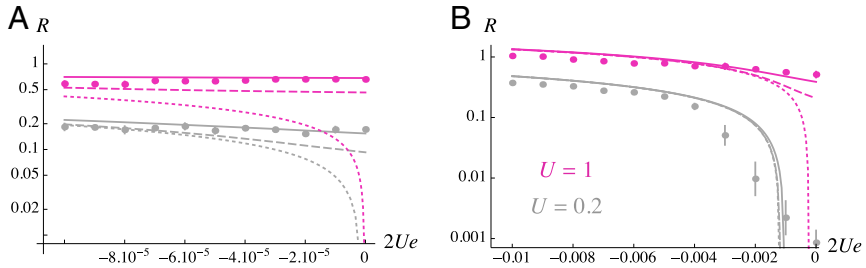

Fig. 4. Effect of negative epistasis: equilibrium chromosome map length R (on log scale) as a function of the coefficient of epistasis between deleterious alleles (e) multiplied by $2U$, for $U = 0.2$ (gray) and $U = 1$ (magenta). The overall strength of selection against heterozygous mutations is kept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

associated with chromosomal segregation probably generate stabilizing selection around an optimal number of cross-overs per bivalent (16, 44), 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 ref. 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 three-locus modifier models extend to more realistic situations involving many genes. The results presented in this article show that extrapolations from three-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 Figs. 2 and 4). Therefore, the approximations presented here may not accurately quantify selection for recombination in populations with very low (or no) recombination, but they 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 $sh \ll 1/N_e$ (N_e being greatly reduced by background selection when $U > R$, as shown by Fig. 1*B*). While an analytical description of this regime remains challenging (e.g., ref. 45), 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 (46), 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., ref. 47). 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 $N_e R$ (Eq. **1**). This stands in contrast to 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 ref. 48 showed that obligate asexuality is often favored when $h \leq 0.25$ (see figure 7 in ref. 48). This difference probably stems from the fact that, unlike recombination modifiers, sex modifiers have a direct effect on heterozygosity among offspring (see also ref. 49). 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 ref. 26) even when population size is large (up to 10^5 in Fig. 4*A*). 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., ref. 50, p. 426). This seems to contradict the simulation results obtained by Keightley and 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 and 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 and 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 1*D* in ref. 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 (47, 51), although these values are associated with considerable uncertainty. According to the present results (Eq. **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 cross-over per bivalent). For example, Fig. 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 Fig. 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

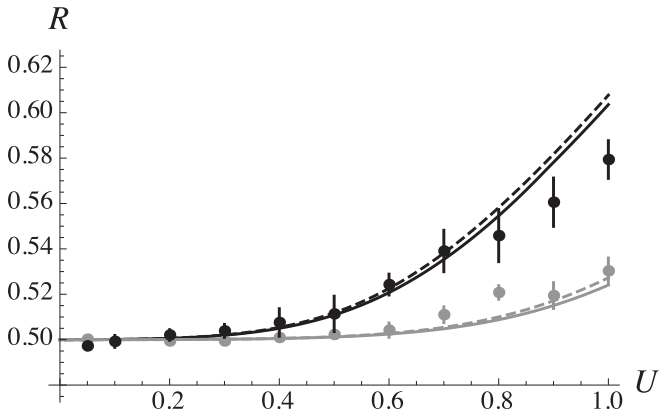

Fig. 5. Equilibrium chromosome map length R as a function of the deleterious mutation rate per haploid chromosome U, under direct stabilizing selection around $R = 0.5$ (of the form $W_c = e^{-c(R-0.5)^2}$, with $c = 0.1$). Dashed curves correspond to the predictions obtained by solving $-2c\,(R-0.5)+1.8\,(N_e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

factors may increase the strength of indirect selection. A first is that cross-overs 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 (52). While gene density is also higher at the chromosome peripheries in plants, this is not particularly the case in animals (52), 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 Fig. 3 and *SI Appendix*, Fig. S3 indicate that the effect of beneficial alleles stays negligible when the beneficial mutation rate is very small relative to U ($U_{\mathrm{ben}} < 10^{-3}\,U$), 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 (53, 54). 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.

Materials and 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 by $1 - s_i$ 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 toward 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 *SI Appendix*, 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 ref. 31, the general principle of the method consists in deriving expressions for different moments of allele frequencies and LDs. 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 refs. 40 and 55). In the case of an additive recombination modifier ($h_m = 1/2$), the expected change in frequency of the modifier takes the form

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

where f is a function of recombination rates, selection, and dominance coefficients, and where $\tilde{p}_a$ and $\tilde{p}_b$ correspond to the frequencies of deleterious alleles at mutation–selection balance [see *SI Appendix* and Mathematica notebook (42) 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 midpoint of the chromosome, that the density of mutations and cross-overs 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

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

where $g(\rho_{ma}, \rho_{mb}, \rho_{ab})$ is a function of scaled recombination rates $\rho_{ma} = r_{ma}/\,(sh)$, $\rho_{mb} = r_{mb}/\,(sh)$, and $\rho_{ab} = r_{ab}/\,(sh)$ that can be found in the Mathematica notebook available on Dryad (42). The first double integral in Eq. **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 corresponds 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 is given by $N_e \approx N \exp(-2U/R)$ (56). 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 Eq. **3** may be approximated by the same integrals taken between zero and infinity, which yields Eq. **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 refs. 40 and 55). 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 $\sim dU/\,(sh)$) behaves similarly to a locus in the three-locus model.

Epistasis. The analysis of ref. 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 *SI Appendix* 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

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

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 two- and three-locus LDs are given by

$$D_{ij} \approx \frac{e\,p_i\,p_j}{r_{ij} - a_i - a_j}, \quad [5]$$

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

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 Fig. 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. The calculations detailed in *SI Appendix* 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 zero (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 zero and one (the number of loci at which mutations can occur is thus effectively infinite). The fitness of each individual is computed as $W = (1 - sh)^{n_{he}}\,(1 - s)^{n_{ho}}\exp(-cR)$, where n_{he} and n_{ho} are the numbers of heterozygous and homozygous mutations present in its genome, and R is the chromosome map length coded by its recombination modifier locus. Gametes are produced by recombining the two chromosomes of the parent, the number of cross-overs being drawn from a Poisson distribution with parameter R (the chromosome map length of the parent), while the position of each cross-over along the chromosome is drawn from a uniform distribution (no interference). Map length R is determined by a modifier locus located at the midpoint 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 one and variance ${\sigma_m}^2$ (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 in 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 *SI Appendix*.

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 (DOI:10.5061/dryad.5mkkwh754) (42).

ACKNOWLEDGMENTS. 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).

1. G. Coop, X. Wen, C. Ober, J. K. Pritchard, M. Przeworski, High resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**, 1395–1398 (2008).
2. J. M. Comeron, R. Ratnappan, S. Bailin, The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1002905 (2012).
3. A. Kong *et al.*, Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
4. A. Kong *et al.*, Common and low-frequency variants associated with genome-wide recombination rate. *Nat. Genet.* **46**, 11–18 (2014).
5. S. E. Johnston, C. Bérénos, J. Slate, J. M. Pemberton, Conserved genetic architecture underlying individual recombination rate variation in a wild population of soay sheep (*Ovis aries*). *Genetics* **203**, 583–598 (2016).
6. K. Samuk, B. Manzano-Winkler, K. R. Ritz, M. A. F. Noor, Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. *Curr. Biol.* **30**, 1517–1528 (2020).
7. J. R. True, J. M. Mercer, C. C. Laurie, Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142**, 507–523 (1996).
8. S. E. Ptak *et al.*, Fine-scale recombination patterns differ between chimpanzees and humans. *Nat. Genet.* **37**, 429–434 (2005).
9. W. Winckler *et al.*, Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308**, 107–111 (2005).
10. C. S. Smukowski, M. A. F. Noor, Recombination rate variation in closely related species. *Heredity* **107**, 496–508 (2011).
11. C. L. Brand, M. V. Cattani, S. B. Kingan, E. L. Landeen, D. C. Presgraves, Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. *Curr. Biol.* **28**, 1289–1295 (2018).
12. B. L. Dumont, B. A. Payseur, Evolution of the genomic rate of recombination in mammals. *Evolution* **62**, 276–294 (2007).
13. J. Stapley, P. G. D. Feulner, S. E. Johnston, A. W. Santure, C. M. Smadja, Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Phil. Trans. Roy. Soc. Lond. B* **372**, 20160455 (2017).
14. S. P. Otto, T. Lenormand, Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**, 252–261 (2002).
15. A. L. Dapper, B. A. Payseur, Connecting theory and data to understand recombination rate evolution. *Phil. Trans. Roy. Soc. Lond. B* **372**, 20160469 (2017).
16. K. R. Ritz, M. A. F. Noor, N. D. Singh, Variation in recombination rate: Adaptive or not? *Trends Genet.* **33**, 364–374 (2017).
17. J. Gonsalves *et al.*, Defective recombination in infertile men. *Hum. Mol. Genet.* **13**, 2875–2883 (2004).
18. A. Kong *et al.*, Recombination rate and reproductive success in humans. *Nat. Genet.* **36**, 1203–1206 (2004).
19. K. A. Ferguson, E. Chan Wong, V. Chow, M. Nigro, S. Ma, Abnormal meiotic recombination in infertile men and its association with sperm aneuploidy. *Hum. Mol. Genet.* **16**, 2870–2879 (2007).
20. A. Fledel-Alon *et al.*, Broad-scale recombination patterns underlying proper disjunction in humans. *PLoS Genet.* **5**, e1000658 (2009).
21. C. S. Ottolini *et al.*, Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. *Nat. Genet.* **47**, 727–737 (2015).
22. K. E. Koehler, R. Scott Hawley, S. Sherman, T. Hassold, Recombination and nondisjunction in human and flies. *Hum. Mol. Genet.* **5**, 1495–1504 (1996).
23. B. Arbeithuber, A. J. Betancourt, T. Ebner, I. Tiemann-Bogge, Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2109–2114 (2015).
24. S. P. Otto, Y. Michalakis, The evolution of recombination in changing environments. *Trends Ecol. Evol.* **13**, 145–151 (1998).
25. B. Charlesworth, Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221 (1990).
26. N. H. Barton, A general model for the evolution of recombination. *Genet. Res.* **65**, 123–144 (1995).
27. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
28. J. Felsenstein, The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
29. S. P. Otto, N. H. Barton, The evolution of recombination: Removing the limits to natural selection. *Genetics* **147**, 879–906 (1997).
30. S. P. Otto, N. H. Barton, Selection for recombination in small populations. *Evolution* **55**, 1921–1931 (2001).
31. N. H. Barton, S. P. Otto, Evolution of recombination due to random drift. *Genetics* **169**, 2353–2370 (2005).
32. D. Roze, N. H. Barton, The Hill–Robertson effect and the evolution of recombination. *Genetics* **173**, 1793–1811 (2006).
33. M. Nei, Modification of linkage intensity by natural selection. *Genetics* **57**, 625–641 (1967).
34. M. W. Feldman, F. B. Christiansen, L. D. Brooks, Evolution of recombination in a constant environment. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4838–4841 (1980).
35. B. Charlesworth, Directional selection and the evolution of sex and recombination. *Genet. Res.* **61**, 205–224 (1993).
36. M. M. Iles, K. Walters, C. Cannings, Recombination can evolve in large finite populations given selection on sufficient loci. *Genetics* **165**, 2249–2258 (2003).
37. P. D. Keightley, S. P. Otto, Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
38. I. Gordo, P. R. A. Campos, Sex and deleterious mutations. *Genetics* **179**, 621–626 (2008).
39. M. Hartfield, S. P. Otto, P. D. Keightley, The role of advantageous mutations in enhancing the evolution of a recombination modifier. *Genetics* **184**, 1153–1164 (2010).

40. D. Roze, Selection for sex in finite populations. *J. Evol. Biol.* **27**, 1304–1322 (2014).
41. B. Charlesworth, M. T. Morgan, D. Charlesworth, The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
42. D. Roze, A simple expression for the strength of selection on recombination generated by interference among mutations. Dryad. https://datadryad.org/stash/dataset/doi:10.5061/dryad.5mkkwh754. Deposited 2 April 2021.
43. M. Kimura, T. Maruyama, The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351 (1966).
44. S. P. Otto, B. A. Payseur, Crossover interference: Shedding light on the evolution of recombination. *Annu. Rev. Genet.* **53**, 19–44 (2019).
45. B. H. Good, A. M. Walczak, R. A. Neher, M. M. Desai, Genetic diversity in the interference selection limit. *PLoS Genet.* **10**, e1004222 (2014).
46. A. Poon, S. P. Otto, Compensating for our load of mutations: Freezing the meltdown of small populations. *Evolution* **54**, 1467–1479 (2000).
47. B. Charlesworth, Causes of natural variation in fitness: Evidence from studies of *Drosophila* populations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1662–1669 (2015).
48. D. Roze, R. E. Michod, Deleterious mutations and selection for sex in finite, diploid populations. *Genetics* **184**, 1095–1112 (2010).
49. S. P. Otto, The advantages of segregation and the evolution of sex. *Genetics* **164**, 1099–1118 (2003).
50. J. F. Crow, M. Kimura, *An Introduction to Population Genetics Theory* (Harper & Row, New York, NY, 1970).
51. P. D. Keightley, Rates and fitness consequences of new mutations in humans. *Genetics* **190**, 295–304 (2012).
52. Q. Haenel, T. G. Laurentino, M. Roesti, D. Berner, Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol. Ecol.* **27**, 2477–2497 (2018).
53. G. Martin, S. P. Otto, T. Lenormand, Selection for recombination in structured populations. *Genetics* **172**, 593–609 (2006).
54. D. Roze, Diploidy, population structure and the evolution of recombination. *Am. Nat.* **174**, S79–S94 (2009).
55. D. Roze, Background selection in partially selfing populations. *Genetics* **203**, 937–957 (2016).
56. R. R. Hudson, N. L. Kaplan, Deleterious background selection with recombination. *Genetics* **141**, 1605–1617 (1995).