



**HAL**  
open science

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

Denis Roze

► **To cite this version:**

Denis Roze. A simple expression for the strength of selection on recombination generated by interference among mutations. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118 (19), pp.e2022805118. 10.1073/pnas.2022805118 . hal-03185704

**HAL Id: hal-03185704**

**<https://hal.sorbonne-universite.fr/hal-03185704>**

Submitted on 30 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

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

Denis Roze<sup>\*,†</sup>

\* 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

1  
2 One of the most widely cited hypotheses to explain the evolutionary main-  
3 tenance of genetic recombination states that the reshuffling of genotypes at meiosis  
4 increases the efficiency of natural selection by reducing interference among selected  
5 loci. However, and despite several decades of theoretical work, a quantitative estima-  
6 tion of the possible selective advantage of a mutant allele increasing chromosomal map  
7 length (the average number of crossovers at meiosis) remains difficult. This article de-  
8 rives a simple and accurate expression for the strength of selection acting on a modifier  
9 gene affecting the genetic map length of a whole chromosome or genome undergoing  
10 recurrent mutation. In particular, it shows that indirect selection for recombination  
11 caused by interference among mutations is proportional to  $(N_e U)^2 / (N_e R)^3$ , where  $N_e$   
12 is the effective population size,  $U$  the deleterious mutation rate per chromosome and  
13  $R$  the chromosome map length. Indirect selection is relatively insensitive to the fit-  
14 ness effects of deleterious alleles, epistasis, or the genetic architecture of recombination  
15 rate variation, and may compensate for substantial costs associated with recombina-  
16 tion when linkage is tight. However, its effect generally stays weak in large, highly  
17 recombining populations.

19 Genetic variation for rates of crossing over at meiosis has been reported in  
20 several species [1–6], showing that recombination landscapes may evolve by selection  
21 or drift: accordingly, differences in recombination rates have been observed between  
22 closely related species [7–11] and over broader taxonomic scales [12, 13]. It has been  
23 recognized for long that both direct and indirect selective forces may drive the evo-  
24 lution of recombination [14–16]. Direct selection stems in particular from molecular  
25 constraints acting on the number of crossovers: in particular, it is usually thought  
26 that in most species, at least one crossover per bivalent is required to ensure proper  
27 chromosomal disjunction and segregation at meiosis; for example, in humans low re-  
28 combination is associated with the production of aneuploid gametes and infertility  
29 [17–21]. Too many crossovers may also be detrimental, as it may lead to disjunction  
30 failure during the first meiotic division [22] and to elevated mutation rates [23]. Indi-  
31 rect selection corresponds to the potential benefits associated with the production of  
32 novel genotypes by recombination [14, 24]. In particular, recombination increases the  
33 efficiency of natural selection in the presence of negative linkage disequilibria (LD) be-  
34 tween selected loci, that is, when beneficial alleles tend to be associated with deleterious  
35 alleles at other loci. Negative LD may be the consequence of epistatic interactions (on  
36 fitness) among loci [25, 26], but is also predicted to arise in any finite population under  
37 selection (a phenomenon known as the Hill-Robertson effect, or selective interference)  
38 [27–32].

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

41 rate of recombination between two selected loci (e.g., [25, 26, 29–35]). In general,  
42 these models show that indirect selection on a recombination modifier should mostly  
43 stem from its effect on selected loci to which it is tightly linked (as the modifier  
44 remains longer associated with the beneficial combinations it contributed to create).  
45 However, evaluating the overall strength of indirect selection on a modifier affecting  
46 the genetic map length of a whole genome or chromosome remains challenging. This  
47 is partly due to the fact that the contribution of higher-order disequilibria between  
48 selected loci (associations between 3, 4 or more loci) is difficult to assess, and also to  
49 the fact that the mathematical approximations used often break down in the case of  
50 tightly linked loci (corresponding to the situation in which indirect selection should  
51 be strongest). Multilocus simulation models have offered important insights [30, 36–  
52 39], showing that indirect selection caused by selective interference among many loci  
53 may be rather strong when linkage is tight. However, these simulations are necessarily  
54 restricted to limited ranges of parameters (in particular, they often focus on situations  
55 in which recombination is very rare) and therefore, how the strength of selection for  
56 recombination scales with the different parameters describing mutation and selection  
57 remains unclear. Another limitation of current theory is that most models on selective  
58 interference consider haploid organisms, while many eukaryotic species are diploid. As  
59 a consequence, we are still lacking general expressions quantifying the possible strength  
60 of selection for recombination at the level of a whole genome, and applicable to most  
61 extent species.

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

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

**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



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

113

114 **The strength of selection for increased map length.** Although the positive LD  
115 observed for intermediate values of  $R$  and  $h < 0.25$  tends to disfavor recombination  
116 (as breaking positive LD decreases the variance in fitness and reduces the efficiency  
117 of selection), the mathematical analysis of the three-locus model shows that indirect  
118 selection on recombination involves at least 14 different mechanisms (corresponding to  
119 the different paths generating  $\langle D_{ma} \rangle$  on Figure S1), of which only one involves  $\langle D_{ab} \rangle$ .  
120 All of these mechanisms favor recombination in the absence of dominance at the se-  
121 lected loci ( $h = 0.5$ ), while dominance generate effects that disfavor recombination  
122 (for example, through its effect on  $\langle D_{ab} \rangle$  just discussed) and other effects that favor  
123 recombination. Interestingly, these different effects of dominance tend to compensate  
124 each other (as shown by Figures S5, S6), so that the net effect of interference favors  
125 increased map length for most parameter values, and is often well approximated by  
126 ignoring the terms generated by dominance (at least as long as  $h \geq 0.2$ ). In that case,  
127 the strength of indirect selection becomes equivalent as in a haploid population of size  
128  $2N$  in which mutations have an effect  $sh$  on fitness (results for haploids are derived in a  
129 *Mathematica* notebook available as Supplementary Material). Furthermore, when the  
130 fitness effect of deleterious alleles is sufficiently weak ( $sh \ll R$ ) selection for recombina-  
131 tion is mostly caused by segregating mutations located in the chromosomal vicinity  
132 of the recombination modifier. In that case, the strength of indirect selection on an  
133 additive modifier increasing map length by an amount  $\delta R$  is found to be approximately

134  $\delta R s_{\text{ind}}$ , with

$$s_{\text{ind}} \approx 1.8 \frac{(N_e U)^2}{(N_e R)^3} \quad (1)$$

135 independently of  $s$  and  $h$ , and with  $N_e \approx N \exp(-2U/R)$  under the model's assump-  
136 tions (a more accurate result for higher values of  $sh$  or lower values of  $R$  can be  
137 obtained by numerical integration over the genetic map, as explained in the Methods  
138 and Supplementary Material).

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

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

157 equation 1 should still hold when map length is a polygenic trait coded by several loci  
158 located at various positions along the chromosome. Indeed, Figure S2D confirms that  
159 the same equilibrium map length is reached when  $R$  is coded by a single locus or by  
160 100 loci with additive effects (adjusting parameters so that the mutational variance  
161 on  $R$  stays the same). The results also extend to the case of a genome consisting of  
162 multiple chromosomes (Figures S2E, S2F). Indeed, the evolution of a local recombina-  
163 tion modifier affecting the map length of its own chromosome is not affected much by  
164 the presence of other chromosomes (as their only effect is to cause a modest reduction  
165 in  $N_e$ , by a factor  $\sim \exp(-8shU)$  per extra chromosome), while indirect selection on  
166 a global modifier affecting the map length of all chromosomes mostly stems from its  
167 local effect, and is thus still approximately given by equation 1.

168

169 **Including beneficial mutations.** Obtaining analytical predictions for the equi-  
170 librium map length when beneficial and deleterious mutations co-occur remains chal-  
171 lenging. Approximations for the strength of selection for recombination generated by  
172 interference between two beneficial alleles have been derived for the case of haploid  
173 populations, but in many cases, accurate predictions can only be obtained numeri-  
174 cally [29, 32]. Furthermore, no simple expression exists for the effective population  
175 size and for the probability of fixation of beneficial mutations when both beneficial  
176 and deleterious alleles segregate at many loci. Therefore, the extra effect of beneficial  
177 mutations on selection for recombination was only explored by simulation (assuming  
178 a constant rate  $U_{\text{ben}}$  of mutation towards beneficial alleles, all with the same selection  
179 and dominance coefficients  $s_{\text{ben}}, h_{\text{ben}}$ ).

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

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

190

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

205 corresponds to the lowest values on the  $x$ -axes of Figure 4 (see Methods). Because  
206 selection for recombination due to interference depends on the effective strength of  
207 selection against deleterious alleles, it is predicted to stay constant along each curve of  
208 Figure 4. As can be seen from Figure 4, the effect of negative epistasis on selection for  
209 recombination often remains small relative to the effect of interference (as the equi-  
210 librium map length is not affected much by  $e$ ), even for population sizes as large as  
211  $10^5$ . Figure S4 confirms that the average number of deleterious alleles per chromosome  
212 stays approximately constant in the simulations as  $e$  varies (due to the scaling of  $s$ ),  
213 while mean fitness increases as epistasis becomes more negative [42]. As shown by  
214 Figure 4B, the effect of epistasis on the ES value of  $R$  becomes more important for  
215 high effective strengths of selection against deleterious alleles.

## 216 DISCUSSION

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

227 increased as a correlated response (e.g., Table 1 in [30]). Although a large body of  
228 theoretical work has explored the possible selective advantages of recombination, as-  
229 sessing the plausible order of magnitude of indirect selection acting on chromosomal  
230 map length stays difficult, as it is generally not obvious how mathematical results from  
231 3-locus modifier models extend to more realistic situations involving many genes. The  
232 results presented in this article show that extrapolations from 3-locus models accu-  
233 rately predict the overall strength of indirect selection acting on a modifier affecting  
234 the map length of a chromosome in finite diploid populations, as long as map length  
235 is not too small relative to the chromosomal mutation rate (roughly, when  $U < R$ ).  
236 Under tight linkage ( $U > R$ ), the analytical model tends to overestimate the strength  
237 of indirect selection (as can be seen from Figures 2 and 4): therefore, the approx-  
238 imations presented here may not accurately quantify selection for recombination in  
239 populations with very low (or no) recombination, but provide correct predictions in  
240 situations where recombination is already frequent, as in most sexual species. The fact  
241 that the model performs poorly when  $U > R$  may be caused by higher-order interac-  
242 tions among selected loci, and also by the assumption that deleterious alleles stay near  
243 mutation–selection balance, which does not hold when  $sh \ll 1/N_e$  ( $N_e$  being greatly  
244 reduced by background selection when  $U > R$ , as shown by Figure 1B). While an ana-  
245 lytical description of this regime remains challenging (e.g., [44]), simulation approaches  
246 are also problematic as mutations may accumulate at a high rate when selection is in-  
247 effective, and the equilibrium map length of a population whose mean fitness declines  
248 rapidly is probably not biologically meaningful. Possible compensatory effects among  
249 mutations should be taken into account when dealing with such situations [45], which  
250 would imply extending the model to incorporate distributions of epistasis.

251 Current estimates of the distribution of fitness effects of mutations indicate that  
252 most deleterious alleles have weak fitness effects (e.g., [46]). Interestingly, the model  
253 shows that in this regime (and as long as  $sh > 1/N_e$  for most mutations), the strength  
254 of indirect selection generated by interference among mutations does not depend much  
255 on the details of the genetic architecture of fitness (selection and dominance coefficients  
256 of deleterious alleles), and can be approximated by a simple expression of  $N_e U$  and  $N_e R$   
257 (equation 1). This stands in contrast with the evolution of sex modifiers (affecting the  
258 rate of sex in partially clonal organisms) which is more dependent on dominance: in  
259 particular, the simulation results of [47] showed that obligate asexuality is often favored  
260 when  $h \leq 0.25$  (see Figure 7 in [47]). This difference probably stems from the fact that,  
261 unlike recombination modifiers, sex modifiers have a direct effect on heterozygosity  
262 among offspring (see also [48]). In agreement with previous results [30, 37], the effect of  
263 epistasis among mutations stays relatively small (and is well predicted by an extension  
264 of the model presented in [26]) even when population size is large (up to  $10^5$  in Figure  
265 4A). Approximation 1 also shows that the  $N_e s_{\text{ind}}$  product (determining to what extent  
266 indirect selection is efficient relative to drift) does not depend on  $N_e$ . From classical  
267 diffusion results, one thus predicts that the fixation probability of a recombination  
268 modifier (relative to the fixation probability of a neutral allele) should not depend on  
269  $N_e$ , since this relative fixation probability is approximately  $2N_e s_{\text{ind}}$  (e.g., p. 426 in [49]).  
270 This seems to contradict the simulation results obtained by Keightley & Otto [37],  
271 showing that the relative fixation probability of a recombination modifier increases with  
272 population size. This discrepancy is probably due to the fact that Keightley & Otto  
273 mostly considered situations in which  $U \gg R$ , while the present approximations break  
274 down in this regime (and also possibly from the fact that the classical diffusion result

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

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



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

## 312 METHODS

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

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

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

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

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

345

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

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

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

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

373

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

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

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

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

383

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

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

393

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

408 for different values of  $R$  (if the individual is heterozygous at the modifier locus,  $R$  is  
409 given by the average between the values coded by its two alleles). Mutation occurs at  
410 the modifier locus at a rate  $\mu$  per generation (generally set to  $10^{-4}$ ); when a mutation  
411 occurs, with probability 0.95 the value of the allele is multiplied by a random number  
412 drawn from a Gaussian distribution with average 1 and variance  $\sigma_m^2$  (generally set to  
413 0.04), while with probability 0.05 a number drawn from a uniform distribution between  
414 -1 and 1 is added to the value of the allele (to allow for large effect mutations), the  
415 new value being set to zero if it is negative. During the first 20,000 generations, map  
416 length does not evolve and is fixed to  $R = 1$ ; mutations are then introduced at the  
417 modifier locus and the population is let to evolve (generally during  $5 \times 10^6$  generations,  
418 the value of the average map length usually reaching an equilibrium during the first  
419  $5 \times 10^5$  generations). The average map length, average fitness, average number of dele-  
420 terious mutations per chromosome and number of fixed mutations are recorded every  
421 500 generations (fixed mutations are removed from the population in order to minimize  
422 execution speed). Different modifications and extensions of the program were consid-  
423 ered (including multiple modifier loci, multiple chromosomes, beneficial mutations and  
424 epistasis) and are described in the Supplementary Material.

425

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

429

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

432 service of Roscoff's Biological Station (Abims platform) for computing time, and the  
433 Agence Nationale pour la Recherche for funding (GenAsex project: ANR-17-CE02-  
434 0016-01, and SelfRecomb project: ANR-18-CE02-0017-02).

- 436 [1] Coop, G., Wen, X., Ober, C., Pritchard, J.K., and Przeworski, M. (2008). High  
437 resolution mapping of crossovers reveals extensive variation in fine-scale recombina-  
438 tion patterns among humans. *Science* *319*, 1395–1398.
- 439 [2] Comeron, J.M., Ratnappan, R., and Bailin, S. (2012). The many landscapes of  
440 recombination in *Drosophila melanogaster*. *PLoS Genet.* *8*, e1002905.
- 441 [3] Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A.,  
442 Jonasdottir, A., Bragi Walters, G., Jonasdottir, A., Gylfason, A., Kristinsson,  
443 K.T., et al. (2010). Fine-scale recombination rate differences between sexes, pop-  
444 ulations and individuals. *Nature* *467*, 1099–1103.
- 445 [4] Kong, A., Thorleifsson, G., Frigge, M.L., Masson, G., Gudbjartsson, D.F., Ville-  
446 moes, R., Magnusdottir, E., Olafsdottir, S.B., Thorsteinsdottir, U., and Stefans-  
447 son, K. (2014). Common and low-frequency variants associated with genome-wide  
448 recombination rate. *Nat. Genet.* *46*, 11–18.
- 449 [5] Johnston, S.E., Béréños, C., Slate, J., and Pemberton, J.M. (2016). Conserved  
450 genetic architecture underlying individual recombination rate variation in a wild  
451 population of Soay sheep (*Ovis aries*). *Genetics* *203*, 583–598.
- 452 [6] Samuk, K., Manzano-Winkler, B., Ritz, K.R., and Noor, M.A.F. (2020). Nat-  
453 ural selection shapes variation in genome-wide recombination rate in *Drosophila*  
454 *pseudoobscura*. *Curr. Biol.* *30*, 1517–1528.



- 455 [7] True, J.R., Mercer, J.M., and Laurie, C.C. (1996). Differences in crossover fre-  
456 quency and distribution among three sibling species of *Drosophila*. *Genetics* *142*,  
457 507–523.
- 458 [8] Ptak, S.E., Hinds, D.A., Koehler, K., Nickel, B., Patil, N., Ballinger, D.G., Prze-  
459 worski, M., Frazer, K.A., and Pääbo, S. (2005). Fine-scale recombination patterns  
460 differ between chimpanzees and humans. *Nat. Genet.* *37*, 429–434.
- 461 [9] Winckler, W., Myers, S.R., Richter, D.J., Onofrio, R.C., McDonald, G.J., Bon-  
462 trop, R.E., McVean, G.A.T., Gabriel, S.B., Reich, D., Donnelly, P., et al. (2005).  
463 Comparison of fine-scale recombination rates in humans and chimpanzees. *Science*  
464 *308*, 107–111.
- 465 [10] Smukowski, C.S. and Noor, M.A.F. (2011). Recombination rate variation in  
466 closely related species. *Heredity* *107*, 496–508.
- 467 [11] Brand, C.L., Cattani, M.V., Kingan, S.B., Landeen, E.L., and Presgraves, D.C.  
468 (2018). Molecular evolution at a meiosis gene mediates species differences in the  
469 rate and patterning of recombination. *Curr. Biol.* *28*, 1289–1295.
- 470 [12] Dumont, B.L. and Payseur, B.A. (2007). Evolution of the genomic rate of recom-  
471 bination in mammals. *Evolution* *62*, 276–294.
- 472 [13] Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., and Smadja, C.M.  
473 (2017). Variation in recombination frequency and distribution across eukaryotes:  
474 patterns and processes. *Phil. Trans. Roy. Soc. (Lond.) B* *372*, 20160455.
- 475 [14] Otto, S.P. and Lenormand, T. (2002). Resolving the paradox of sex and recom-  
476 bination. *Nat. Rev. Genet.* *3*, 252–261.

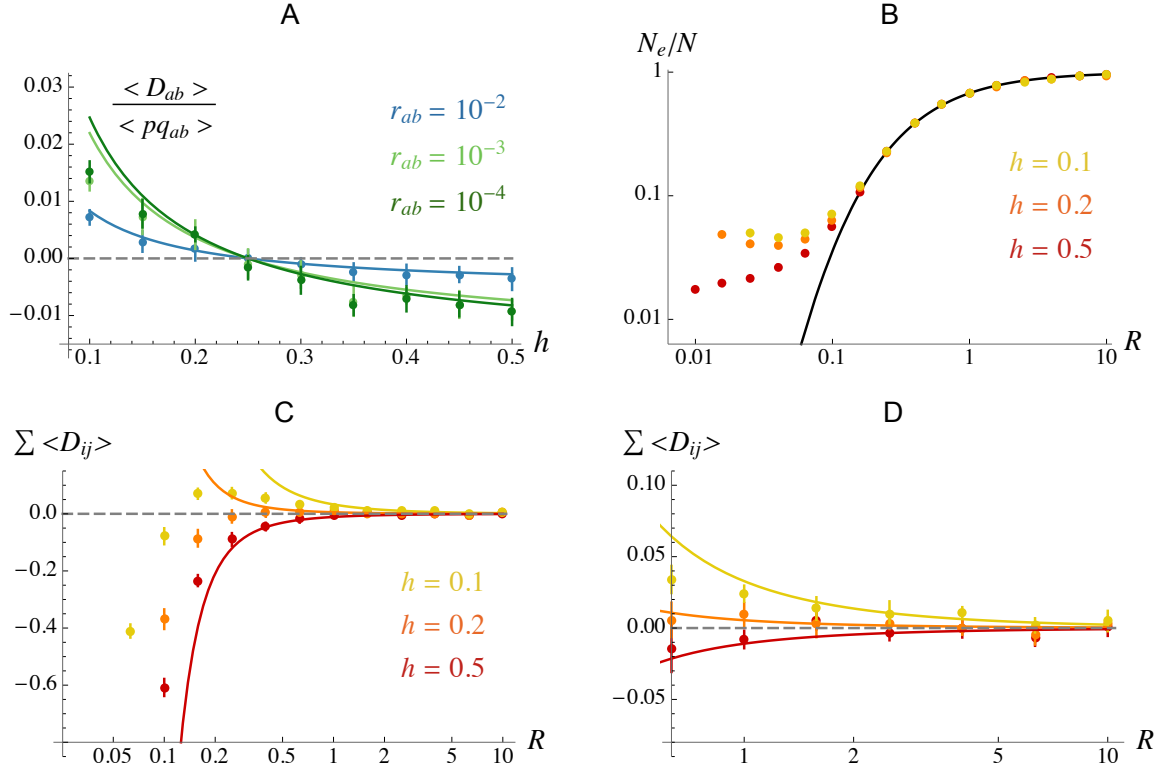
- 477 [15] Dapper, A.L. and Payseur, B.A. (2017). Connecting theory and data to un-  
478 derstand recombination rate evolution. *Phil. Trans. Roy. Soc. (Lond.) B* *372*,  
479 20160469.
- 480 [16] Ritz, K.R., Noor, M.A.F., and Singh, N.D. (2017). Variation in recombination  
481 rate: adaptive or not? *Trends Genet.* *33*, 364–374.
- 482 [17] Gonsalves, J., Sun, F. Schlegel, P.N., Turek, P.J., Hopps, C.V., Greene, C., Mar-  
483 tin, R.H., and Reijo Pera, R.A. (2004). Defective recombination in infertile men.  
484 *Hum. Mol. Genet.* *13*, 2875–2883.
- 485 [18] Kong, A., Barnard, J., Gudbjartsson, D.F., Thorleifsson, G., Jonsdottir, G., Sig-  
486 urdardottir, G., Richardsson, B., Jonsdottir, J., Thorgeirsson, T., Frigge, M.L.,  
487 et al. (2004). Recombination rate and reproductive success in humans. *Nat.*  
488 *Genet.* *36*, 1203–1206.
- 489 [19] Ferguson, K.A., Chan Wong, E., Chow, V., Nigro, M., and Ma, S. (2007). Ab-  
490 normal meiotic recombination in infertile men and its association with sperm  
491 aneuploidy. *Hum. Mol. Genet.* *16*, 2870–2879.
- 492 [20] Fledel-Alon, A., Wilson, D.J., Broman, K., Wen, X., Ober, C., Coop, G., and  
493 Przeworski, M. (2009). Broad-scale recombination patterns underlying proper  
494 disjunction in humans. *PLoS Genet.* *5*, e1000658.
- 495 [21] Ottolini, C.S., Newnham, L.J., Capalbo, A., Natesan, S.A., Joshi, H.A.,  
496 Cimadomo, D., Griffin, D.K., Sage, K., Summers, M.C., Thornhill, A.R., et al.  
497 (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.

- 540 [41] Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of  
541 deleterious mutations on neutral molecular variation. *Genetics* *134*, 1289–1303.
- 542 [42] Kimura, M. and Maruyama, T. (1966). The mutational load with epistatic gene  
543 interactions in fitness. *Genetics* *54*, 1337–1351.
- 544 [43] Otto, S.P. and Payseur, B.A. (2019). Crossover interference: shedding light on  
545 the evolution of recombination. *Ann. Rev. Gen.* *53*, 19–44.
- 546 [44] Good, B.H., Walczak, A.M., Neher, R.A., and Desai, M.M. (2014). Genetic  
547 diversity in the interference selection limit. *PLoS Genetics* *10*, e1004222.
- 548 [45] Poon, A. and Otto, S.P. (2000). Compensating for our load of mutations: freezing  
549 the meltdown of small populations. *Evolution* *54*, 1467–1479.
- 550 [46] Charlesworth, B. (2015). Causes of natural variation in fitness: Evidence from  
551 studies of *Drosophila* populations. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 1662–1669.
- 552 [47] Roze, D. and Michod, R.E. (2010). Deleterious mutations and selection for sex in  
553 finite, diploid populations. *Genetics* *184*, 1095–1112.
- 554 [48] Otto, S.P. (2003). The advantages of segregation and the evolution of sex. *Genetics*  
555 *164*, 1099–1118.
- 556 [49] Crow, J.F. and Kimura, M. (1970). *An Introduction to Population Genetics*  
557 *Theory* (New York: Harper and Row).
- 558 [50] Keightley, P.D. (2012). Rates and fitness consequences of new mutations in hu-  
559 mans. *Genetics* *190*, 295–304.

- 560 [51] Haenel, Q., Laurentino, T.G., Roesti, M., and Berner, D. (2018). Meta-analysis  
561 of chromosome-scale crossover rate variation in eukaryotes and its significance to  
562 evolutionary genomics. *Mol. Ecol.* *27*, 2477–2497.
- 563 [52] Martin, G., Otto, S.P., and Lenormand, T. (2006). Selection for recombination  
564 in structured populations. *Genetics* *172*, 593–609.
- 565 [53] Roze, D. (2009). Diploidy, population structure and the evolution of recombina-  
566 tion. *Am. Nat.* *174*, S79–S94.
- 567 [54] Roze, D. (2016). Background selection in partially selfing populations. *Genetics*  
568 *203*, 937–957.
- 569 [55] Hudson, R.R. and Kaplan, N.L. (1995). Deleterious background selection with  
570 recombination. *Genetics* *141*, 1605–1617.

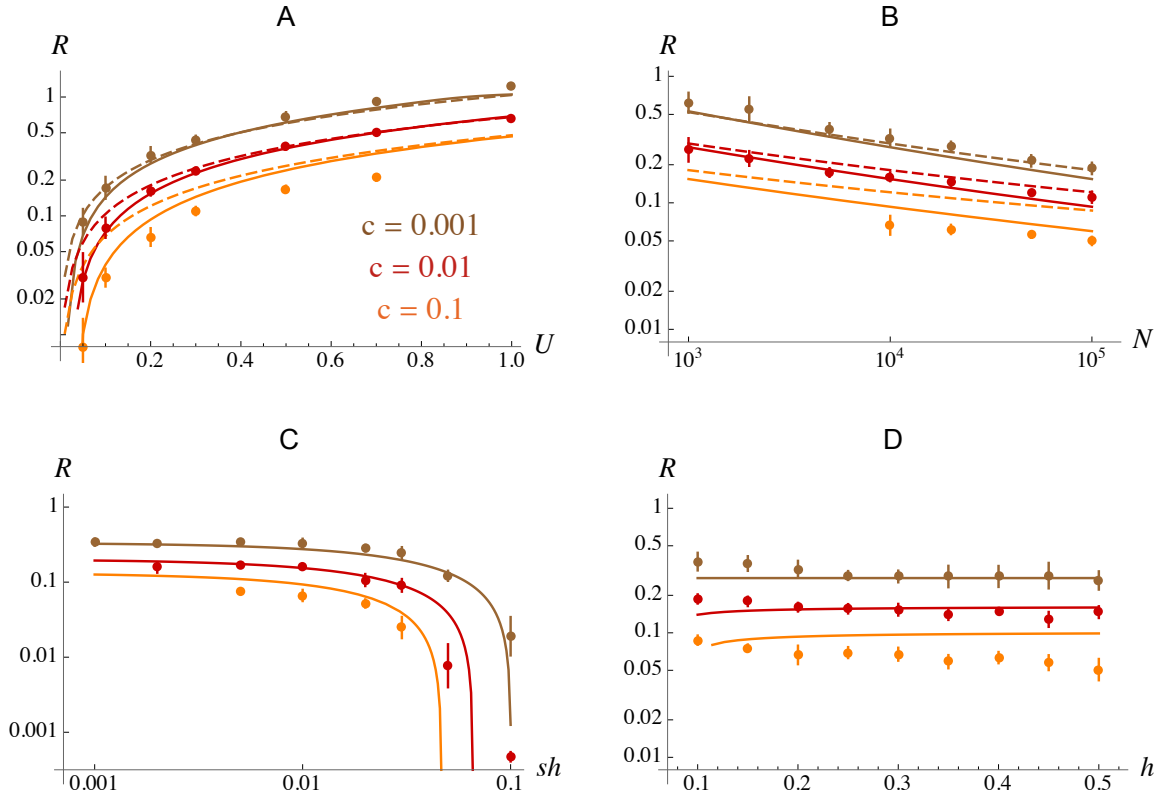


571

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  $s^2 h (1 - 4h) / [2N (r_{ab} + 2sh)^2 (r_{ab} + 3sh)]$  (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  $0.095(1 - 4h)\bar{n}^2 / (N_e R h)$ , where  $\bar{n} = U / (sh)$  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  $R$ ).

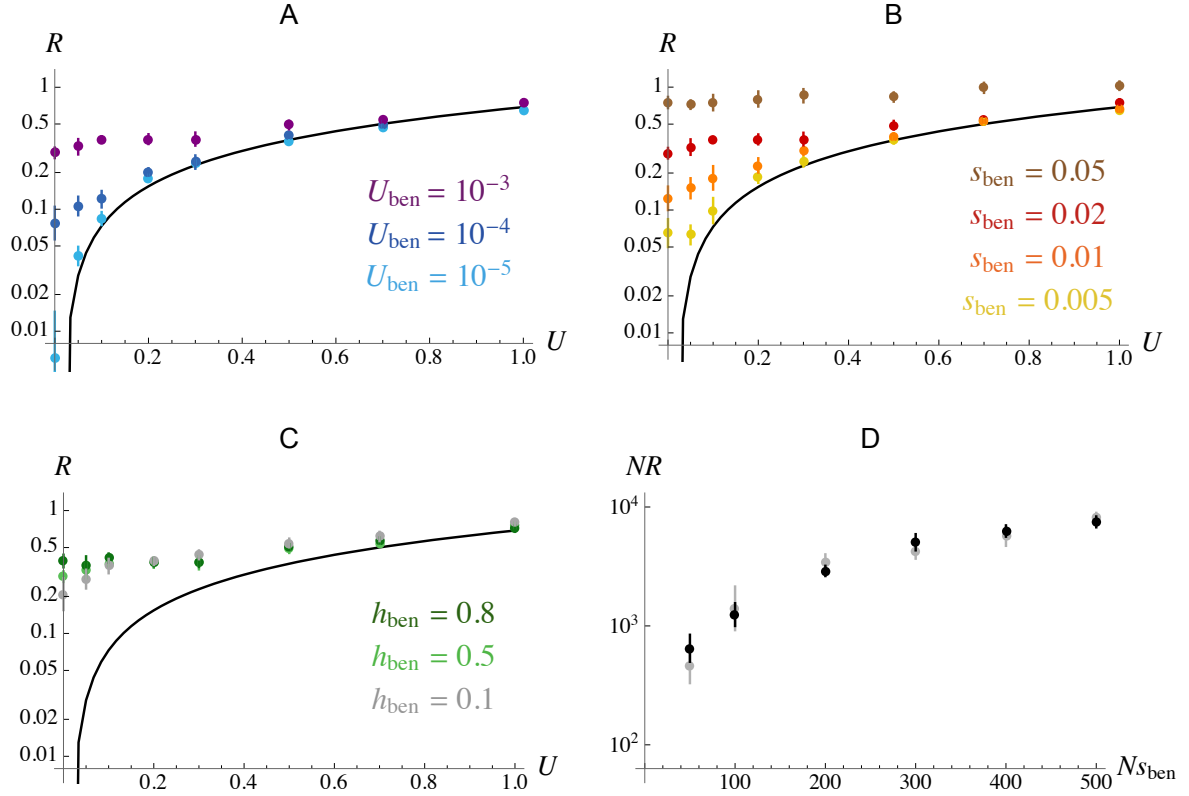




590

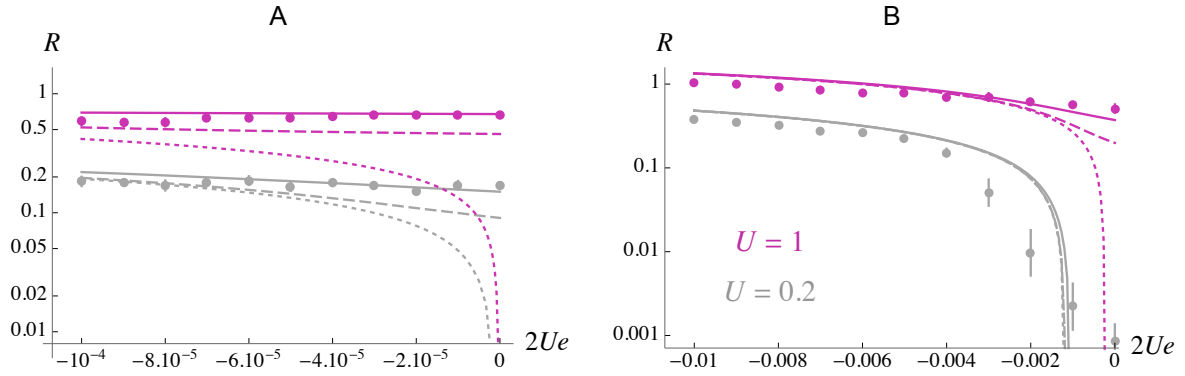
591 **Figure 2.** Equilibrium chromosome map length  $R$  (on log scale) for different values of  
 592 the cost of recombination  $c$ , as a function of the deleterious mutation rate per haploid  
 593 chromosome  $U$  (A), population size  $N$  (on log scale, B), fitness effect of heterozygous  
 594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

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



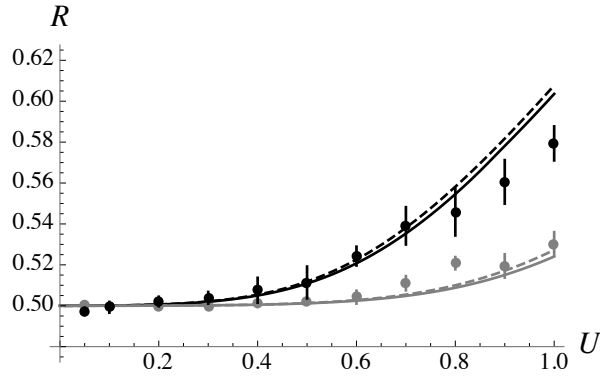
605

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  $U_{\text{ben}}$  (A), fitness effect  $s_{\text{ben}}$  (B) and dominance coefficient  
 609  $h_{\text{ben}}$  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  $h_{\text{ben}}$  varies (by adjusting  $s_{\text{ben}}$ ). D: scaling with population  
 614 size:  $NR$  at equilibrium as a function of  $Ns_{\text{ben}}$ , 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  $-2c(R - 0.5) + 1.8(N_e U)^2 / (N_e R)^3 = 0$  with  
632  $N_e = N e^{-2U/R}$ , 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).

2 **Analytical three-locus model**

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

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

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

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

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



70  $M$  in the population ( $p_m$  and  $q_m$ ), and of the equilibrium frequencies of deleterious  
 71 alleles  $\tilde{p}_a$  and  $\tilde{p}_b$ . A similar method was used by Barton and Otto to compute the  
 72 strength of indirect selection acting on the recombination modifier in a haploid model  
 73 [4]; however, their derivations assume that selection is much weaker than recombina-  
 74 tion ( $s_i \ll r_{jk}$  for all  $i, j, k$ ), a necessary assumption for the QLE to hold in the case  
 75 where beneficial alleles at loci  $A$  and  $B$  are sweeping through the population. The  
 76 results shown below thus take similar forms as equations B3 and S2.3 in [4], except  
 77 that they do not diverge when recombination rates tend to zero. As explained below,  
 78 selection for recombination is generated by a variety of effects involving selection and  
 79 drift, which are summarized in Figure S1.

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

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

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

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

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

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

$$\langle p_a D_{ab} \rangle \approx -\frac{s_b h_b}{r_{ab} + 2s_a h_a + s_b h_b} \langle D_{ab}^2 \rangle \quad (3)$$

$$\langle D_{ma} D_{mab} \rangle \approx -\frac{s_b h_b}{r_{ma} + r_{mab} + 2s_a h_a + s_b h_b} \langle D_{mab}^2 \rangle \quad (4)$$

95  
 96  $\langle p_b D_{ab} \rangle$  and  $\langle D_{mb} D_{mab} \rangle$  being given by symmetric expressions. The negative value  
 97 of  $\langle p_a D_{ab} \rangle$  corresponds to the fact that when  $D_{ab}$  is positive, allele  $a$  is associated  
 98 with the deleterious allele  $b$ , and thus tends to decrease in frequency ( $p_a$  decreases);  
 99 conversely when  $D_{ab} < 0$ , allele  $a$  is associated with the better allele  $B$ , causing  $p_a$   
 100 to increase. Negative values of  $\langle D_{ma} D_{mab} \rangle$ ,  $\langle D_{mb} D_{mab} \rangle$  stem from the fact that when  
 101  $D_{mab}$  is positive, selection against deleterious alleles is more efficient in the background  
 102 of allele  $m$  than in the background of allele  $M$  (because the variance in fitness is  
 103 higher in the background of allele  $m$ ), causing lower frequencies of deleterious alleles  
 104 in the background of allele  $m$  (that is,  $D_{ma}$ ,  $D_{mb} < 0$ ). Conversely when  $D_{mab}$  is  
 105 negative, selection against deleterious alleles is less efficient in the background of allele  
 106  $m$ , generating positive associations  $D_{ma}$ ,  $D_{mb}$ .

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

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

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

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

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

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

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

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

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

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

142 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  
143 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  
144 equation 4. We have:

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

145  $\langle D_{mb} D_{ab} \rangle$  being given by a symmetric expression. Equation 4 above shows that  
146  $\langle D_{ma} D_{mab} \rangle$  is negative: when  $D_{mab}$  is negative,  $D_{ma}$  tends to be positive. As we  
147 have just seen, a negative  $D_{mab}$  leads to positive  $D_{ab}$  in the population (when allele  
148  $m$  increases recombination), generating a positive covariance between  $D_{ma}$  and  $D_{ab}$ .  
149 Given that a negative  $D_{mab}$  leads to a positive  $D_{ma}$ , the negative moment  $\langle D_{ab} D_{mab} \rangle$   
150 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  
151 given by:

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

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

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

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

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

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

184 The moments  $\langle p_a D_{mab} \rangle$  and  $\langle D_{ma} D_{ab} \rangle$  also generate a negative covariance be-  
185 tween  $p_a$  and  $D_{ma}$ :

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

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

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

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

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

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

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

203 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  
204 parameters of the model from equations 1 – 13 above. Note that all moments gener-  
205 ated by the modifier effect are of order  $\delta r_{ab} \tilde{p}_a \tilde{p}_b / (N\epsilon^2)$ , so that the expected change  
206 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  
207 recombination modifier ( $h_m = 1/2$ ), it takes the form:

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

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

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

## 222 **Multilocus extrapolation**

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



234 is given by:

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

235 to the first order in  $\delta R$  (e.g., [5]). From the previous results, the strength on indirect  
236 selection is given by:

$$\langle \Delta_{\text{ind}} p_m \rangle \approx -sh \sum_i \langle D_{mi} \rangle \quad (17)$$

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

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

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

$$\langle \Delta_{\text{ind}} p_m \rangle \approx \frac{\delta R s_{\text{ind}}}{2} p_m q_m \quad (19)$$

246 with:

$$s_{\text{ind}} = \frac{4U^2}{NR^3} \left[ \int_0^{\frac{R}{2sh}} \int_0^{\frac{R}{2sh}} (x+y) g(x, y, x+y) dx dy \right. \\ \left. + \int_0^{\frac{R}{2sh}} \int_0^{\frac{R}{2sh}} |x-y| g(x, y, |x-y|) dx dy \right]. \quad (20)$$

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

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

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

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

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

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

$$N_e \approx N \exp \left[ -\frac{2U}{R + 2sh} \right] \quad (22)$$

272 (note that  $U$  refers to the haploid chromosomal mutation rate in the present paper,  
273 and to the diploid mutation rate in [7], explaining the extra factor 2). From equations  
274 20 – 22, one can notice that  $s_{\text{ind}}$  does not depend on  $N$  as long as the products  $NU$ ,  
275  $NR$  and  $Ns$  stay constant: one thus predicts that for a given value of  $c$  (the direct  
276 cost of recombination), the evolutionarily stable value of  $NR$  should be independent of  
277  $N$  as long as  $NU$  and  $Ns$  stay constant. This prediction is confirmed by simulations  
278 (Figure S2).

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

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

## 296 **Epistasis**

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

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

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

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

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

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

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

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

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

$$D_{ij} \approx \frac{e_{ij} \tilde{p}_i \tilde{p}_j}{r_{ij} - a_i - a_j} \quad (26)$$

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

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

318

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

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

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

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

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

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

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

## 338 Simulation programs

339 **Two-locus model.** Two-locus simulations were used to check the analytical predic-  
340 tion for the average linkage disequilibrium between deleterious alleles  $\langle D_{ab} \rangle$ , given by  
341 equation 5 (Figure 1A). For this, the program (written in C++) represents the effects  
342 of mutation (also including back mutation at a rate  $v = u/10$ ), drift, selection and  
343 recombination on two-locus genotype frequencies over a large number of generations  
344 ( $10^9$ ).  $D_{ab}$  among gametes and  $p_aq_ap_bq_b$  were measured every 10 generations to ob-  
345 tain averages over  $10^8$  data points, and the results were averaged over 10 replicate  
346 simulations.

347 **Baseline multilocus model.** The multilocus simulation program represents a pop-  
348 ulation of  $N$  individuals carrying two copies of a linear chromosome. Each gener-  
349 ation, the number of new deleterious mutations per chromosome is drawn from a

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

374 drawn from a Gaussian distribution with average 1 and variance  $\sigma_m^2$  (generally set to  
375 0.04), while with probability 0.05 a number drawn from a uniform distribution between  
376 -1 and 1 is added to the value of the allele (to allow for large effect mutations), the  
377 new value being set to zero if it is negative. During the first 20,000 generations, map  
378 length does not evolve and is fixed to  $R = 1$ ; mutations are then introduced at the  
379 modifier locus and the population is let to evolve (generally during  $5 \times 10^6$  genera-  
380 tions, the value of the average map length usually reaching an equilibrium during the  
381 first  $5 \times 10^5$  generations). The average map length, average fitness, average number  
382 of deleterious mutations per chromosome and number of fixed mutations are recorded  
383 every 500 generations (fixed mutations are removed from the population in order to  
384 minimize execution speed). Error bars in the figures are obtained by splitting the  
385 simulation results into batches of  $5 \times 10^5$  generations (removing the first batch during  
386 which the average map length reaches its equilibrium) and computing the variance  
387 of batch averages (error bars correspond to  $\pm 1.96$  S.E.). Different modifications and  
388 extensions of the program were considered, as described below.

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



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

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

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

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

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

$$\sum_{i \neq j} \langle D_{ij} \rangle \approx \frac{0.095}{N_e R} \frac{1 - 4h}{h} \bar{n}^2 \quad (33)$$

409 with  $N_e \approx N \exp[-2U/R]$ , and where  $\bar{n} = U/(sh)$  is the mean number of mutations  
 410 per chromosome. Equations 32 and 33 yield nearly undistinguishable curves on Figures  
 411 1C and 1D (not shown).

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

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

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

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

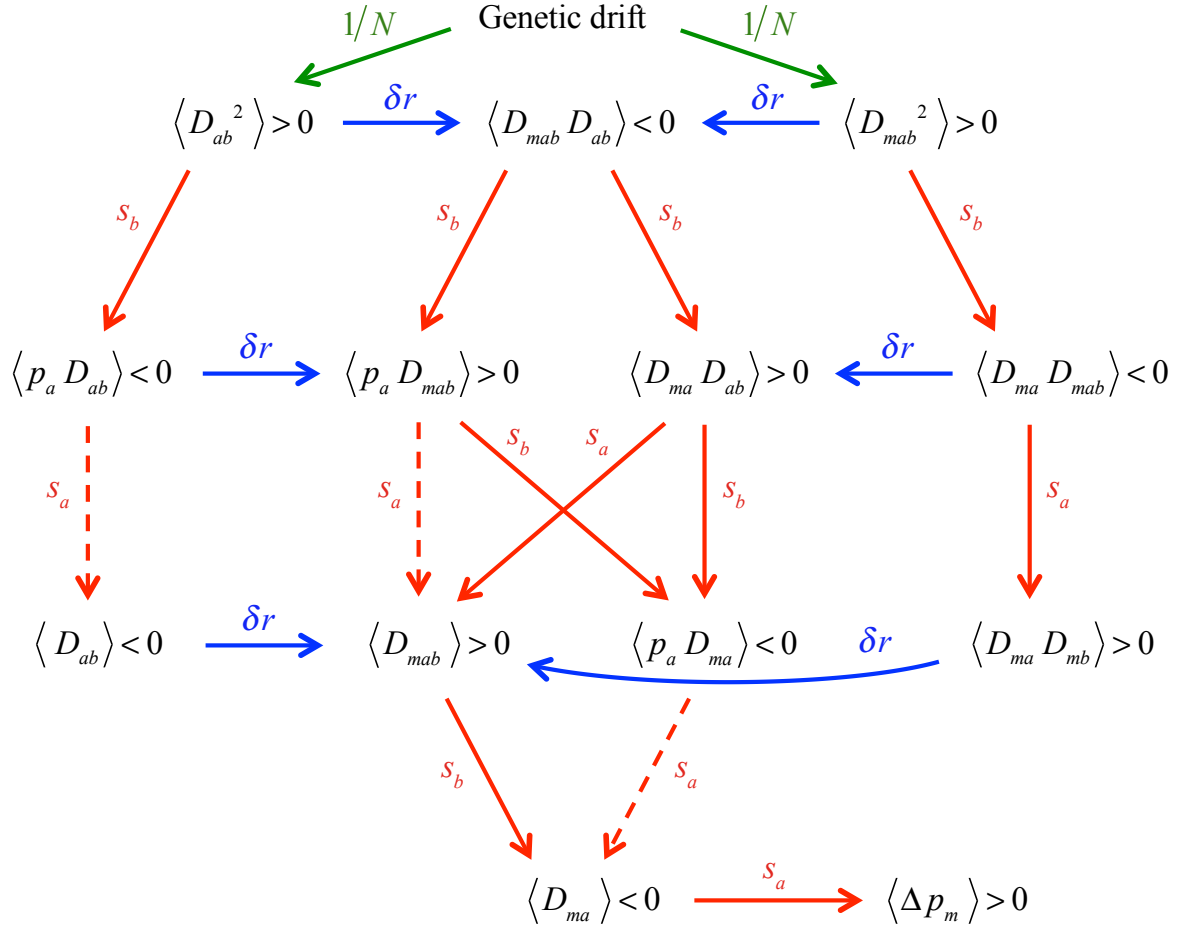
437 map lengths).

438 **Beneficial mutations.** Beneficial alleles were introduced in the standard model in or-  
439 der to explore the effect of the interaction between beneficial and deleterious mutations  
440 on the evolution of recombination (Figures 3 and S3). In that case, beneficial muta-  
441 tions with selection and dominance coefficients  $s_{\text{ben}}$  and  $h_{\text{ben}}$  (and with multiplicative  
442 effects across loci) occur at a rate  $U_{\text{ben}}$  per chromosome per generation (an additional  
443 table is associated to each chromosome, containing the positions of the different bene-  
444 ficial alleles present on the chromosome). Once a beneficial allele has reached fixation,  
445 it is removed from the population in order to minimize execution speed.

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

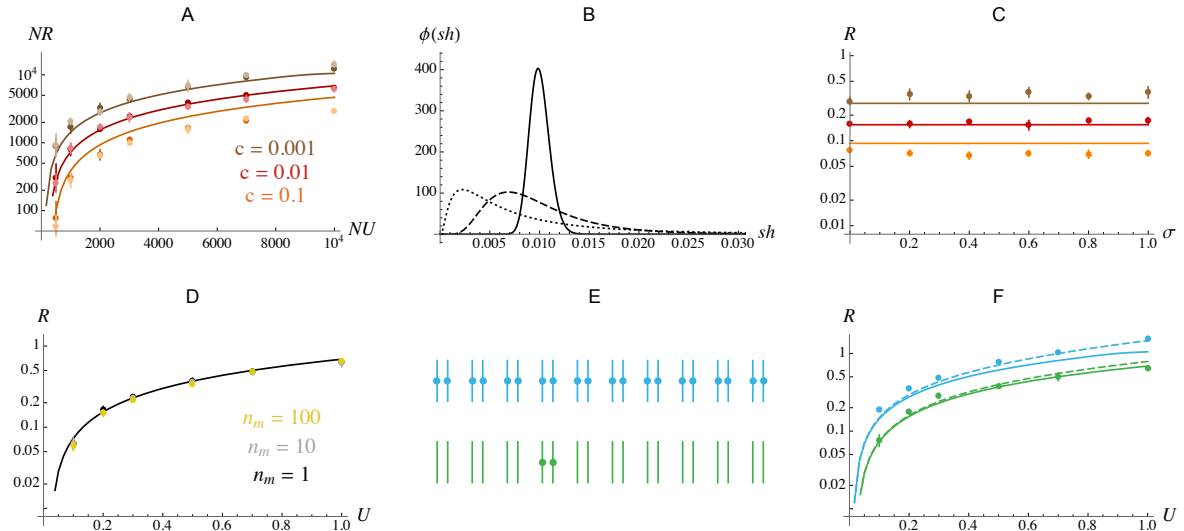
- 449 [1] Slatkin, M. (1972). On treating the chromosome as the unit of selection. *Genetics*  
450 *72*, 157–168.
- 451 [2] Roze, D. (2014). Selection for sex in finite populations. *J. Evol. Biol.* *27*, 1304–  
452 1322.
- 453 [3] Roze, D. (2016). Background selection in partially selfing populations. *Genetics*  
454 *203*, 937–957.
- 455 [4] Barton, N.H. and Otto, S.P. (2005). Evolution of recombination due to random  
456 drift. *Genetics* *169*, 2353–2370.
- 457 [5] Gervais, C. and Roze, D. (2017). Mutation rate evolution in partially selfing and  
458 partially asexual organisms. *Genetics* *207*, 1561–1575.
- 459 [6] Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of  
460 deleterious mutations on neutral molecular variation. *Genetics* *134*, 1289–1303.
- 461 [7] Hudson, R.R. and Kaplan, N.L. (1995). Deleterious background selection with  
462 recombination. *Genetics* *141*, 1605–1617.
- 463 [8] Robertson, A. (1961). Inbreeding in artificial selection programmes. *Genet. Res.*  
464 *2*, 189–194.
- 465 [9] Charlesworth, B. (2012). The effects of deleterious mutations on evolution at  
466 linked sites. *Genetics* *190*, 5–22.

- 467 [10] Otto, S.P. and Barton, N.H. (1997). The evolution of recombination: removing  
468 the limits to natural selection. *Genetics* *147*, 879–906.
- 469 [11] Roze, D. (2009). Diploidy, population structure and the evolution of recombina-  
470 tion. *Am. Nat.* *174*, S79–S94.
- 471 [12] Abu Awad, D. and Roze, D. (2020). Epistasis, inbreeding depression and the  
472 evolution of self-fertilization. *Evolution* *74*, 1301–1320.
- 473 [13] Barton, N.H. (1995). A general model for the evolution of recombination. *Genet.*  
474 *Res.* *65*, 123–144.



**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

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

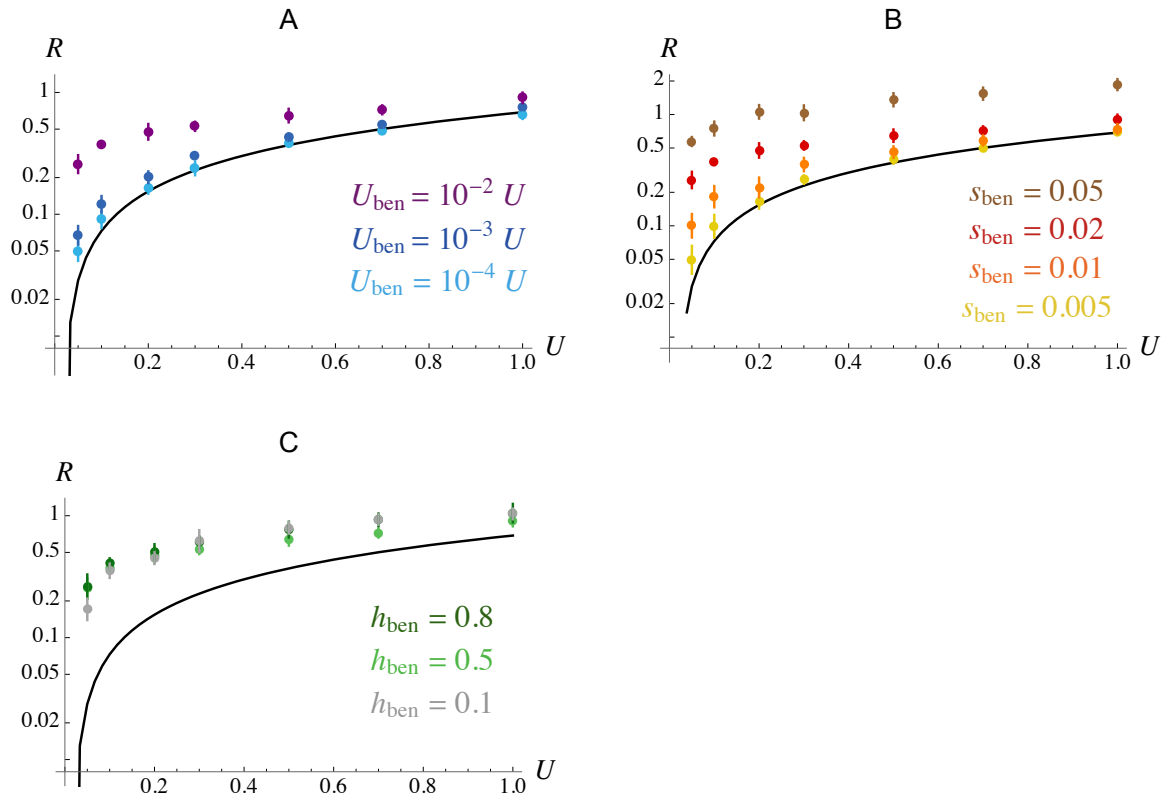


15

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

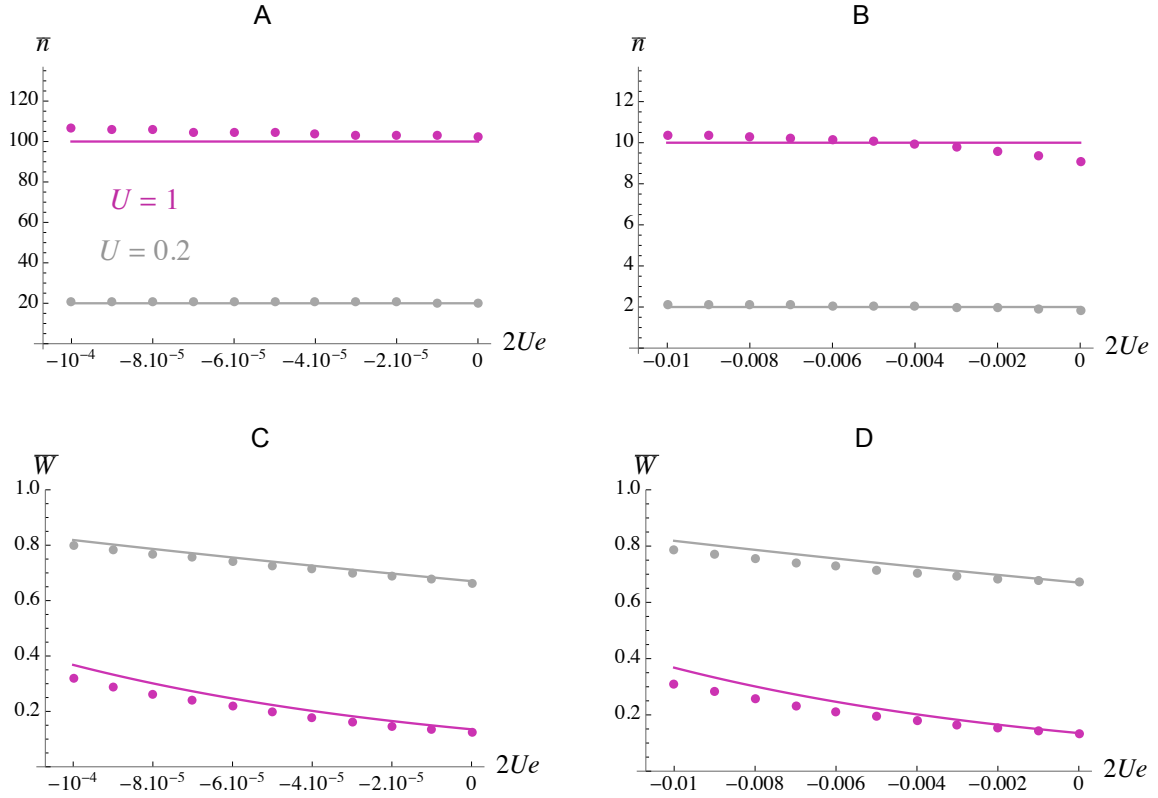


31 same as in Figure 2, with  $c = 0.001$ . Because  $c$  is multiplied by the total map length  
32 of the genome in the fitness function (see Supplementary Material), the strength of  
33 direct selection acting on local modifiers is  $c$ , but  $n_{\text{chr}} c$  in the case of a global modifier,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



41

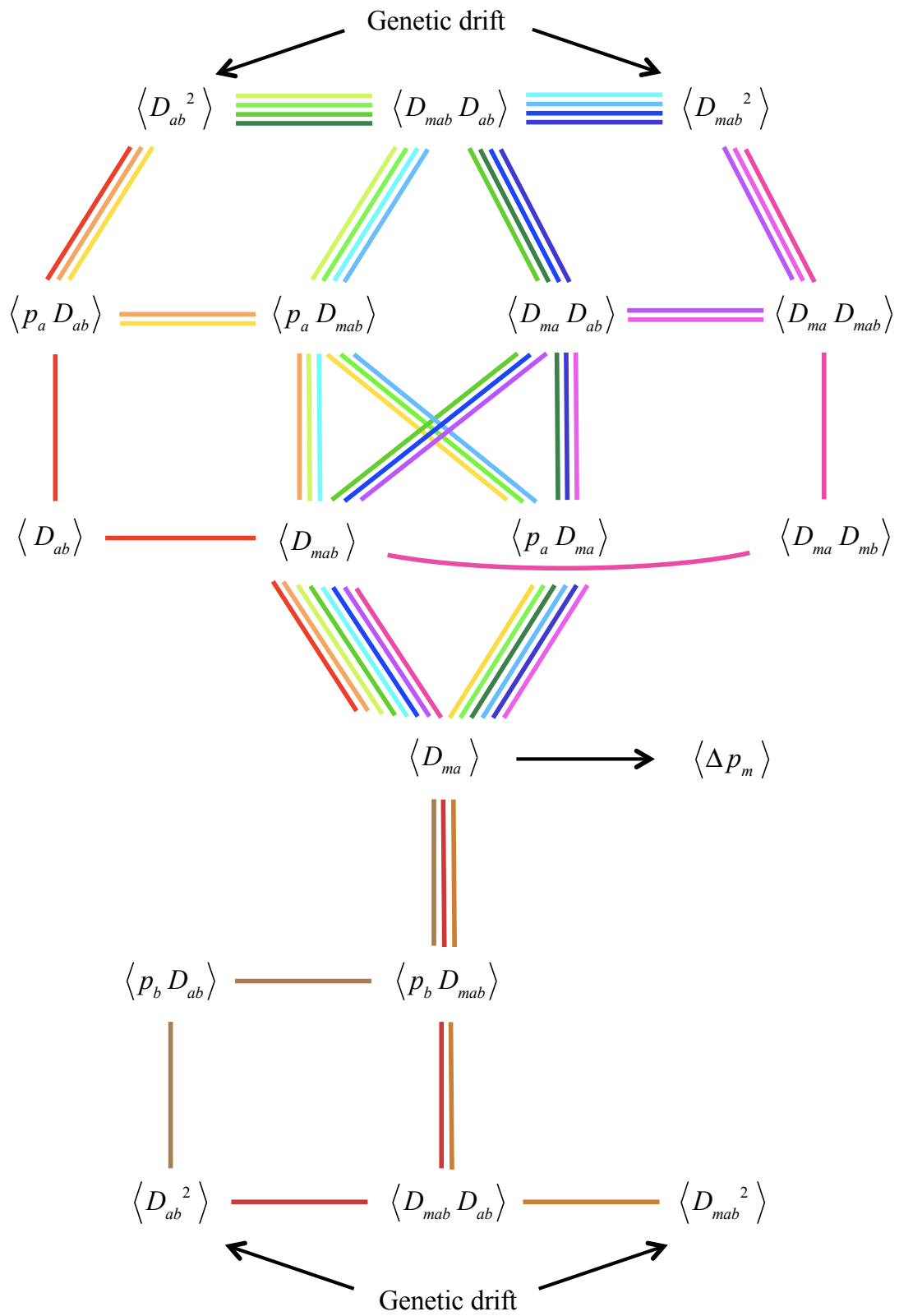
42 **Figure S3.** Same as Figure 3A, 3B, 3C when the beneficial mutation rate  $U_{\text{ben}}$  is  
 43 proportional to the deleterious mutation rate  $U$ . Parameter values are as in Figure 3,  
 44  $U_{\text{ben}} = 10^{-2} U$  in B, C.



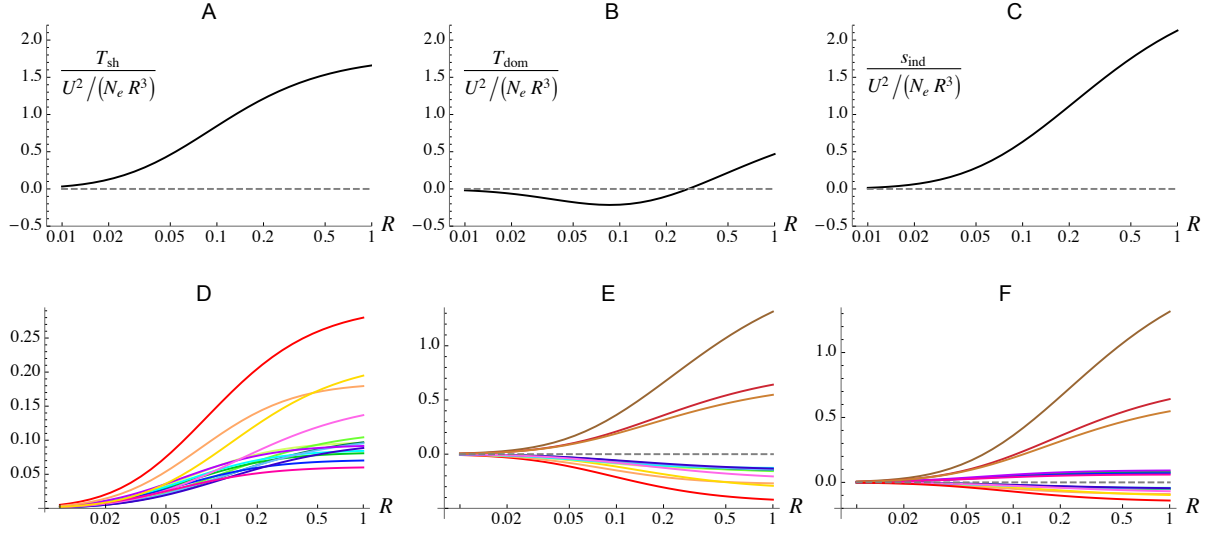
45

46 **Figure S4.** Mean number of deleterious mutations per chromosome  $\bar{n}$  (A, B) and mean  
 47 fitness  $\bar{W}$  (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



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



58

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