

# Effect of Varying Epistasis on the Evolution of Recombination

Roger D. Kouyos,\* Sarah P. Otto<sup>†</sup> and Sebastian Bonhoeffer\*<sup>1</sup>

\*Integrative Biology, ETH Zürich, ETH Zentrum CHN, CH-8092 Zürich, Switzerland and <sup>†</sup>Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Manuscript received November 4, 2005  
Accepted for publication March 13, 2006

## ABSTRACT

Whether recombination decelerates or accelerates a population's response to selection depends, at least in part, on how fitness-determining loci interact. Realistically, all genomes likely contain fitness interactions both with positive and with negative epistasis. Therefore, it is crucial to determine the conditions under which the potential beneficial effects of recombination with negative epistasis prevail over the detrimental effects of recombination with positive epistasis. Here, we examine the simultaneous effects of diverse epistatic interactions with different strengths and signs in a simplified model system with independent pairs of interacting loci and selection acting only on the haploid phase. We find that the average form of epistasis does not predict the average amount of linkage disequilibrium generated or the impact on a recombination modifier when compared to results using the entire distribution of epistatic effects and associated single-mutant effects. Moreover, we show that epistatic interactions of a given strength can produce very different effects, having the greatest impact when selection is weak. In summary, we observe that the evolution of recombination at mutation–selection balance might be driven by a small number of interactions with weak selection rather than by the average epistasis of all interactions. We illustrate this effect with an analysis of published data of *Saccharomyces cerevisiae*. Thus to draw conclusions on the evolution of recombination from experimental data, it is necessary to consider the distribution of epistatic interactions together with the associated selection coefficients.

**P**URGING deleterious mutations is one of the main explanations for the evolution of recombination. The menace of deleterious mutations is as ubiquitous as recombination itself: Every organism experiences at any time the opposing forces of selection and mutation. Selection drives populations to the peaks of the fitness landscape, maximizing the average fitness, while mutation spreads them out across the sequence space, generating deleterious mutations. Recombination potentially increases the efficiency of selection by combining good alleles with good genetic backgrounds and bad alleles with bad genetic backgrounds. This results in an indirect selection for high recombination rates, because subpopulations with high recombination rates will be more localized around the fitness peaks. However, the only effect of recombination is to weaken statistical associations between different loci. Therefore recombination increases the chance of good alleles and good backgrounds occurring together, only if the good alleles are predominantly found within bad backgrounds. Such a statistical association is termed a negative linkage disequilibrium.

Why should linkage disequilibria be negative? In this article we focus on statistical associations between loci

due to fitness interactions. In a two-locus ( $A_1, A_2$ )/two-allele (0, 1) model fitness interactions can be measured by epistasis, which is given by

$$\epsilon = w_{A_1A_2=11}w_{A_1A_2=00} - w_{A_1A_2=10}w_{A_1A_2=01},$$

where  $w_{A_1A_2=xy}$  is the fitness of the haploid genotype with  $A_1 = x$  and  $A_2 = y$ . If epistasis is positive, the double mutant has a higher fitness than expected from the fitness of the single mutants (measured on a multiplicative scale). Conversely, if epistasis is negative the double mutant has a lower fitness than expected from the fitness of the single mutants. Therefore, as long as no other effects generate linkage disequilibria, the extreme genotypes are expected to be under- (over) represented for negative (positive) epistasis. Indeed, in an infinite panmictic population at mutation–selection balance linkage disequilibrium always has the same sign as epistasis (ESHEL and FELDMAN 1970). Thus, for two loci, negative epistasis leads to negative linkage disequilibria and hence recombination increases the efficiency of selection.

In two-locus/two-allele haploid models there is only one epistatic interaction. Thus the above reasoning is limited to cases where all epistatic interactions have the same sign. It is no surprise, however, that in real genomes both positive and negative epistatic interactions have been found (MUKAI 1964; KITAGAWA 1967; DE VISSER *et al.* 1997; ELENA and LENSKI 1997; WEST *et al.*

<sup>1</sup>Corresponding author: Theoretical Biology, ETH Zürich, ETH Zentrum CHN, Universitaetsstr. 16, CH-8092 Zürich, Switzerland.  
E-mail: seb@env.ethz.ch

1998; ELENA 1999; DE LA PEÑA *et al.* 2000; PETERS and KEIGHTLEY 2000; WLOCH *et al.* 2001; BONHOEFFER *et al.* 2004). Although it is well understood how epistasis affects the evolution of recombination in a two-locus system, the effects of several epistatic interactions with different signs and strengths remain largely unexplored. Using an analytical approximation, BARTON (1995) found that if the recombination modifier is tightly linked to loci under direct selection, a high recombination rate is favored whenever the mean epistasis is negative. Using the same method OTTO and FELDMAN (1997) showed that, due to the recombinational load, variance in epistasis reduces the parameter space in which recombination is favored. However, due to the underlying assumptions these theoretical results are limited to cases with weak selection and weak epistasis.

In this article we examine how linkage disequilibrium and a recombination modifier behave in a multilocus system with a broad range of epistasis. We assume that genomes consist of independent pairs of loci, which is the simplest possibility to take account of several different epistatic interactions. Furthermore, we assume that selection occurs only in the haploid phase. First we examine how the shape of the linkage-disequilibrium distribution relates to the shape of the underlying epistasis distribution. Possible incongruities between the epistasis and the linkage-disequilibrium distributions are crucial, because fitness interactions affect the evolution of recombination only via the linkage disequilibrium they produce. Next we determine the strength of the linkage disequilibrium and the selection on the recombination modifier produced by a single fitness interaction in a two-locus system. Considering single interactions is important for our original question because the effect of epistatic interactions in a two-locus system may indicate the strength of their effect in a multilocus system. In contrast to OTTO and FELDMAN (1997), we assume small recombination rates both between the selected loci themselves and between the modifier and the selected loci. Last, we consider a model consisting of two independent pairs of loci and a recombination modifier. This is the simplest case, where more than one fitness interaction simultaneously affect the evolution of recombination, and helps to clarify how the results based on two-locus systems can be extended to the multilocus case.

## METHODS

**Model:** Our model describes the effect of recombination in an infinite population at mutation–selection balance. The state of the population is given by the frequencies  $f_i$ , where  $i$  is a haplotype of the genome. The genome consists of  $L$  biallelic loci, with the alleles at locus  $k$  denoted by  $A_k = 0$  and  $A_k = 1$ . We assume discrete generations. In each generation there is selection, mutation, and recombination. For simplicity, we consider

selection to occur only in the haploid phase. Therefore the frequencies after selection are given by

$$f_i' = \frac{w_i}{\bar{w}} f_i,$$

where  $w_i$  is the fitness of the haplotype  $i$  and  $\bar{w}$  is the mean fitness. Unless stated otherwise, we assume that backward mutations occur at the same rate as forward mutations. Thus, the frequencies after mutation are

$$f_i'' = \sum_{j=1}^{2^L} m^{h(i,j)} (1-m)^{N-h(i,j)} f_j',$$

where  $m$  is the mutation rate and  $h(i, j)$  is the Hamming distance, which measures the number of loci at which haplotypes  $i$  and  $j$  differ. To simplify the recursion relations, we allow a maximum of one crossover event per replication. Between any pair of adjacent loci, crossover occurs with a probability  $r/(L-1)$  such that the total probability of a crossover is  $r$  (*i.e.*, crossover interference is complete). Assuming random mating, we obtain

$$f_i''' = (1-r)f_i'' + \sum_{j,l=1}^{2^L} R_{i,j,l} f_j'' f_l'',$$

where  $R_{i,j,l} = r/(L-1)$  provided that one can construct genotype  $i$  from genotypes  $j$  and  $l$  using one crossover and  $R_{i,j,l} = 0$  in all other cases. We obtain the haplotype frequencies of successive generations by recursively equating  $f_i$  to  $f_i'''$ .

Locus  $A_1$  assumes the role of a modifier of recombination rate. That is, locus  $A_1$  has no impact on the fitness-function  $w_i$  but affects recombination in the following manner: Recombination takes place with rate  $r = r_0$  when both genomes have allele 0 at the modifier locus, with rate  $r = r_0 + \delta r$  when only one has allele 1, and with rate  $r = r_0 + 2\delta r$  if both have allele 1. OTTO and FELDMAN (1997) found that for loose linkage variance in epistasis inhibits the evolution of recombination. In contrast, we focus here mainly on tight linkage. In most cases we have chosen  $r_0 = 0$  and  $\delta r$  very small, typically 0.01 or 0.001. Note that the crossing-over probability between the modifier locus  $A_1$  and its neighbor  $A_2$  is the same as between any other pair of adjacent loci. Furthermore the mutation rate for the modifier locus is zero.

Simulating the model, we typically let the system run to steady state with the modifier switched off (*i.e.*,  $A_1 = 0$  for all genomes) and then change the modifier from allele 0 to allele 1 in half of the genomes. After this, we can distinguish between three consecutive phases (see Figure 1): a rather short perturbation phase caused by the abrupt change of a large number of modifier alleles, a linear phase in which the modifier frequency increases or decreases with a constant slope, and finally, when either the 0 or the 1 allele reaches fixation, a long

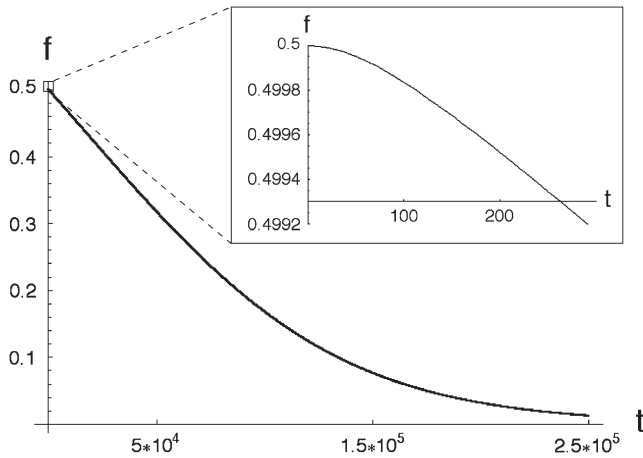


FIGURE 1.—Frequency  $f$  of allele 1 at a recombination modifier linked to a two-locus/two-allele system as a function of time. The graph displays the linear and the saturation phase; the inset displays the perturbation and the linear phase. Parameters are  $m = 0.001$ ,  $r = 0$ ,  $\delta r = 0.5 \times 10^{-2}$ , and  $w_{A_1=1,A_2=0} = w_{A_1=0,A_2=1} = w_{A_1=1,A_2=1} = 0.995$ .

saturation phase, where the steepness of the curve decreases continuously. We call the slope of the modifier frequency in the linear phase the strength of selection acting on the modifier. Similar results were obtained by calculating the leading eigenvalue of the local stability matrix using the *Mathematica* package of OTTO and FELDMAN (1997).

**Fitness landscape:** To reduce complexity, we focus here on an idealized genome structure that decomposes into  $l = (L/2)$  independent pairs of loci. This means that the genome has the form  $A_1A'_1A_2A'_2, \dots, A_lA'_l$ . The term independent refers to our assumption that there are no epistatic interactions between pairs of loci, save for the pairs of interacting loci  $A_kA'_k$  ( $k = 1, \dots, l$ ), which are always adjacent in our simulations. Hence, the fitness of the whole genome can be written as the product of the fitness values of the pairs of interacting loci:

$$w_{A_1A'_1, \dots, A_lA'_l} = \prod_{k=1}^l w_{A_kA'_k}.$$

Epistatic interactions are assumed to occur only within pairs of interacting loci, where epistasis is defined as

$$\epsilon_k = w_{A_kA'_k=11}w_{A_kA'_k=00} - w_{A_kA'_k=10}w_{A_kA'_k=01}.$$

The linkage disequilibrium between pairs of interacting loci vanishes at mutation–selection balance, because there are no epistatic interactions between them. Thus their steady-state frequencies are given by

$$\hat{f}_{A_1A'_1, \dots, A_lA'_l} = \prod_{k=1}^l \hat{f}_{A_kA'_k},$$

where  $\hat{f}_{A_kA'_k}$  are the steady-state frequencies of the two-locus system with fitness function  $w_{A_kA'_k}$  and mutation rate  $m$ . The only nonvanishing linkage disequilibria are

those within pairs of interacting loci,  $D_{A_kA'_k}$ , which are the same as the linkage disequilibria in the two-locus systems given by  $w_{A_kA'_k}$ . Hence, if we are interested only in the distribution of linkage disequilibria, we can solve the steady-state equations for  $l$  two-locus systems instead of one  $L$ -locus system. This combines the simplicity of two-locus genomes with the possibility of multilocus genomes to account for the simultaneous presence of diverse epistatic interactions (PHILLIPS *et al.* 2000).

RESULTS

**Distributions of epistasis and linkage disequilibrium:** Epistatic interactions affect the evolution of recombination by generating linkage disequilibria. Recombination reduces the magnitude of these linkage disequilibria, which, depending on the sign of linkage disequilibrium, can increase or decrease the population’s response to selection. This process will have an overall beneficial effect only if the beneficial effects of reducing negative disequilibria dominate the detrimental effects of reducing positive linkage disequilibria. Which effect prevails depends on the linkage-disequilibrium distribution.

We examined how the form of this distribution relates to the form of the epistasis distribution in systems consisting of independent pairs of loci. To this end the fitness values of each pair were drawn from a given probability distribution. We considered different distributions, but with the constraint that they produced symmetric epistasis distributions with mean zero. Examples are given in Figure 2. Note that although the epistasis distributions are symmetric with mean zero (Figure 2, A and C), the corresponding distributions of linkage disequilibria are not symmetric and their mean differs significantly from zero (Figure 2, B and D).

In Figure 2B there are far more strongly negative than strongly positive linkage disequilibria, while the converse is true for Figure 2D. The main difference between both distributions is that for Figure 2, A and B, pairs with negative epistasis are associated with weak selection while for Figure 2, C and D, this is the case for pairs with positive epistasis. Since the level of linkage disequilibrium produced by an epistatic interaction of a given strength depends inversely on the selection coefficient (see next section), this leads to skewed linkage-disequilibrium distributions.

What does this mean for recombination? Recombination increases (decreases) the response to selection whenever linkage disequilibrium is negative (positive). The magnitude of this change depends monotonically on the magnitude of the linkage disequilibrium. Hence, for the linkage disequilibrium shown in Figure 2B we expect that recombination increases the response to selection. Yet, for other fitness distributions we observed predominantly positive linkage disequilibria (see Figure 2D), in which case we expect recombination to decrease the response to selection. Thus one cannot conclude

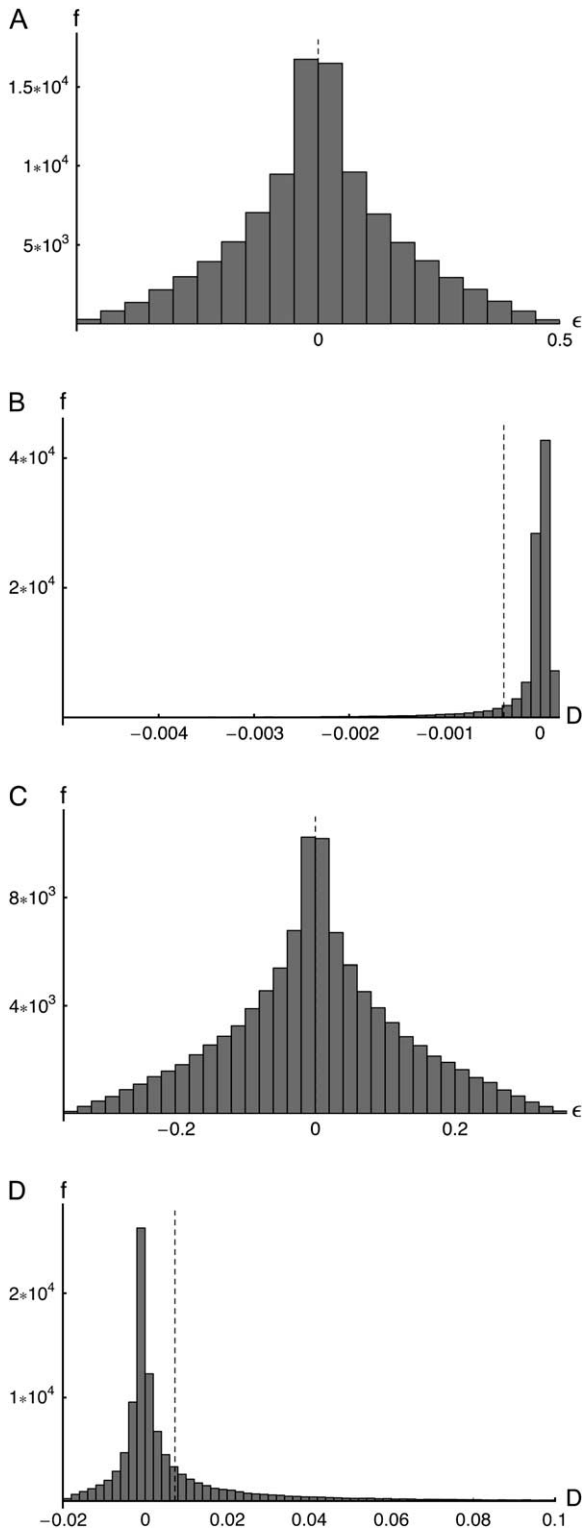


FIGURE 2.—Histogram for epistasis  $\epsilon = w_1w_4 - w_2w_3$  (A and C) and linkage disequilibrium  $D$  (B and D) for  $10^5$  independent pairs of loci, where the fitnesses  $w_i$  are drawn according to the following distributions: Dashed lines refer to the mean values of  $D$  and  $\epsilon$ , respectively. (A and B) First the double-mutant fitness  $w_4$  is drawn uniformly between 0 and 0.5. Then the epistasis  $\epsilon$  is drawn uniformly between  $-w_4$  and  $w_4$ , resulting in a single-mutant fitness of  $w_2 = w_3 = \sqrt{w_4 - \epsilon}$ . (C and D) First the single-mutant fitness  $w_2 = w_3$  is drawn uniformly

whether positive or negative linkage disequilibria have a greater impact in general. However, two important points should be noted: First, the shape of the linkage-disequilibrium distribution does not in general reflect the shape of the epistasis distribution. Second, this incongruity can crucially affect the predicted fate of a modifier for recombination.

**Effects of epistasis in two-locus systems:** To interpret the above result, we examined the effects of epistasis in a two-locus/two-allele system. First, we calculated the strength of the linkage disequilibrium from a single epistatic interaction at mutation–selection balance in the absence of recombination (*i.e.*,  $r = 0$ ). The genotype frequencies  $(\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4)$  at mutation–selection balance are given by the normalized eigenvector of the mutation–selection matrix:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & 0 & w_3 & 0 \\ 0 & 0 & 0 & w_4 \end{pmatrix} \times \begin{pmatrix} (1 - m)^2 & (1 - m)m & (1 - m)m & m^2 \\ m(1 - m) & (1 - m)^2 & m^2 & m(1 - m) \\ m(1 - m) & m^2 & (1 - m)^2 & m(1 - m) \\ m^2 & m(1 - m) & m(1 - m) & (1 - m)^2 \end{pmatrix}.$$

The corresponding linkage disequilibrium  $\hat{D} = \hat{f}_1\hat{f}_4 - \hat{f}_2\hat{f}_3$  is a complicated expression of the fitnesses  $w_i$  and the mutation rate  $m$ , but we can obtain a simple approximation by neglecting back mutations.

Assuming furthermore that the strength of selection on both single and double mutants is considerably larger than the mutation rate [*i.e.*,  $m \ll (1 - w_{2,3,4})$ ], we obtain the expression

$$D = \frac{m^2\epsilon}{(1 - w_2)(1 - w_3)(1 - w_4)} + O[m^3],$$

with the epistasis  $\epsilon = w_4 - w_2w_3$ .

To investigate the effect of recombination, we consider an approximation of the steady-state equations for small recombination rates  $r$  (see APPENDIX). From this we obtain the following expressions for the change of linkage disequilibrium,

$$\Delta D \approx -\frac{r}{1 - w_4}D,$$

and for the change of mean fitness

$$\Delta \bar{w} \approx -rD,$$

between the steady states with recombination rates  $r$  and 0, respectively. The above equations show that  $D$ ,  $\Delta D$ ,

between 0.8 and 1. Then the epistasis  $\epsilon$  is drawn uniformly between  $-(1 - w_2^2)$  and  $(1 - w_2^2)$ , resulting in a double-mutant fitness of  $w_4 = w_2^2 + \epsilon$ .

and  $\Delta\bar{w}$  are proportional to the epistasis  $\varepsilon$  and inversely proportional to the squared strength of selection on the single mutant ( $1 - w_2$ ) and the strength of selection on the double mutant ( $1 - w_4$ ). As expected, epistasis does have an effect on the resulting linkage disequilibrium. However, the strength of this effect is weighted by the strength of selection. In particular, epistasis has the strongest effect on linkage disequilibrium when selection on either the single or the double mutant becomes weak (*i.e.*, if  $w_2$  or  $w_4$  is close to 1).

Importantly, epistasis of the same level can induce very different absolute levels of linkage disequilibria. This is the reason why one can obtain a strongly skewed distribution of linkage disequilibria from a symmetric distribution of epistatic interactions, as observed in Figure 2. In Figure 2, A and B, negative epistatic interactions induce on average stronger disequilibria than positive epistatic interactions of the same absolute magnitude. As can be seen in Figure 3A the above conclusions for the linkage disequilibrium  $D$  also hold if we allow for back mutations (the corresponding plots for  $\Delta D$  and  $\Delta\bar{w}$  look similar).

Considering only fitnesses leading to large linkage disequilibria, we can distinguish three regions in the parameter space. In region 1 selection is strong on the single mutants, but weak on the double mutant (*i.e.*,  $w_2 = w_3 \ll 1$ ,  $w_4 \simeq 1$ ). In region 2 selection is weak on the single mutants, but strong on the double mutants (*i.e.*,  $w_2 = w_3 \simeq 1$ ,  $w_4 \ll 1$ ). Finally, in region 3 selection is weak on both single and double mutants (*i.e.*,  $w_2, w_3, w_4 \simeq 1$ ). We refer to these regions as compensatory mutations, synthetic lethals, and overall weak selection, although these terms have been used in the literature without this strict quantitative definition. In the remaining region the linkage disequilibrium is several orders of magnitude smaller. For a genome consisting of independent pairs of loci this implies that the bulk of epistatic interactions leads to a vanishingly small linkage disequilibrium and is therefore irrelevant for the evolution of recombination. Instead relatively few epistatic interactions may dominate the behavior of a recombination modifier.

Next we go one step further and determine the relation between epistasis and strength of selection acting on a modifier of recombination. Because we are unable to derive analytical expressions for the strength of selection on the modifier for the entire parameter space, we show numerical simulations in Figure 3B. Figure 3B shows that the conclusions are qualitatively the same for the strength of selection as for linkage disequilibrium. In particular, synthetic lethals, compensatory mutations, and overall weak selection exert a far stronger (indirect) selection on the modifier than all other fitness values. Qualitatively, these results also hold for loose linkage between selected loci, but the sensitivity of the modifier to the strength of selection on individual loci decreases rapidly as the linkage between the selected loci decreases (see Figure 4).

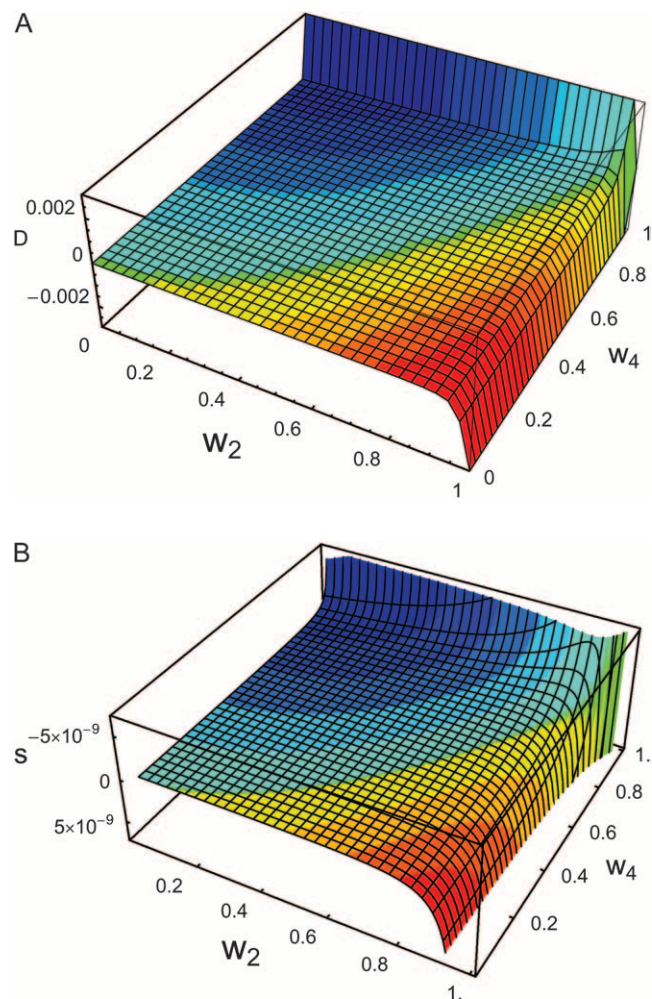


FIGURE 3.—Linkage disequilibrium (A) and strength of selection on the recombination modifier (B) in a two-locus system as a function of the single- and double-mutant fitnesses  $w_2 = w_3$  and  $w_4$ . Note that in B the z-axis has been flipped to correspond to A. Different colors indicate different levels of epistasis. Other parameters are  $m = 10^{-3}$  (backward mutation rate = forward mutation rate),  $r_0 = 0$ , and  $\delta r = 0.5 \times 10^{-3}$  (B only).

Two points are worth further consideration. First, although selection on the modifier is positive for negative epistasis that is not too strong relative to selection on the single mutant, it becomes negative as the selection on the single mutant vanishes (see Figure 5). This is in accordance with previous theory (see, for example, Equation 1, box 1 in OTTO and LENORMAND 2002).

Second, we always consider systems at mutation–selection equilibrium. This approach makes sense only if we can assume that this state is reached within a reasonable time, *i.e.*, on a timescale that is considerably shorter than the timescale of changes in the fitness landscape. If this is not the case, we expect adaptive and purging dynamics to interfere. Therefore the above selection coefficients would most probably overestimate the impact of deleterious mutations. Figure 6 shows the

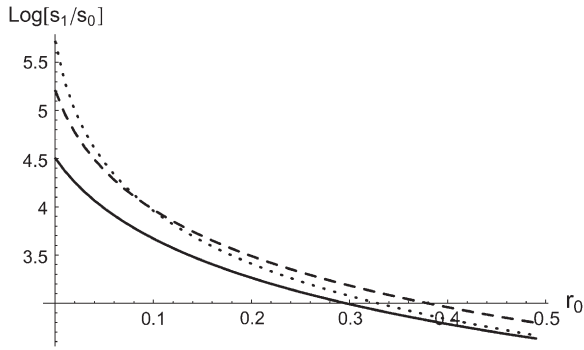


FIGURE 4.—The sensitivity of a modifier to selection coefficients is plotted as a function of the recombination rate between the selected loci  $r_0$ . The sensitivity is measured as  $\text{Log}_{10}(s_1/s_0)$ , where  $s_1$  is the selection on the modifier under weak selection ( $w_2 = w_3 = 0.99$ ) and  $s_0$  is the selection on the modifier under strong selection ( $w_2 = w_3 = 0.1$ ), holding epistasis constant ( $\epsilon = -0.01$ ). Parameters are:  $m = 0.001$ ; linkage between selected loci,  $r_0$  variable,  $\delta r = 0.5 \times 10^{-3}$ ; and linkage between the recombination modifier and the first selected locus fixed at  $0.5 \times 10^{-3}$ .

time it takes to reach 90% of the steady-state level of the linkage disequilibrium (starting from a homogeneous wild-type population) for different fitness values. As expected the equilibration time is larger the weaker selection is. Hence, the linkage disequilibrium takes a large time to establish in those regions of the fitness landscape where the strength of selection on the modifier is largest.

**Impact of two fitness interactions on a recombination modifier:** Last, we discuss five-locus systems (that is, four selected loci plus one modifier locus) to investigate how the above results for the modifier extend to systems with more than one fitness interaction. We considered two independent pairs of loci, with fitnesses  $w_1 = 1, w_2 = w_3, w_4$  for the first pair and  $\tilde{w}_1 = 1, \tilde{w}_2 = \tilde{w}_3, \tilde{w}_4$  for the

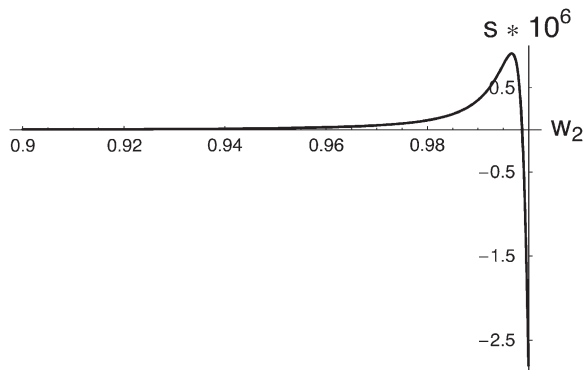


FIGURE 5.—Strength of selection on the modifier in a two-locus system as a function of single-mutant fitness  $w_2$ . The selection on the recombination modifier is positive save for very weak selection on the single mutants (*i.e.*,  $w_2 = w_3 \simeq 1$ ). The double-mutant fitness is fixed at  $w_4 = 0.5$ . Other parameters are  $m = 10^{-3}$ ,  $r_0 = 0$ , and  $\delta r = 0.5 \times 10^{-3}$ .

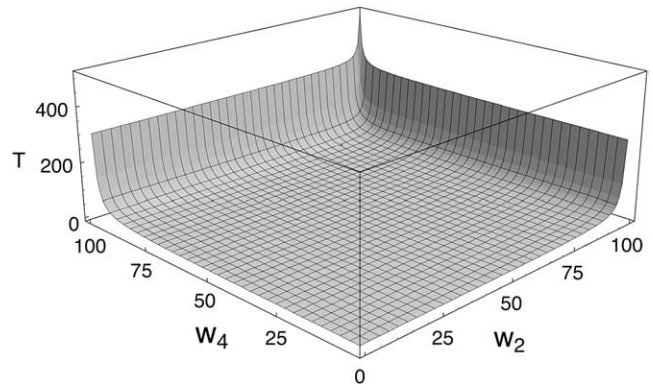


FIGURE 6.—Time until linkage disequilibrium is within 10% of its expected value ( $T$ ) as a function of  $w_2 = w_3$  and  $w_4$ . Other parameters are  $m = 10^{-3}$  and  $r = 0$ .

second pair. We compare this five-locus system with two three-locus systems that we obtain by linking a recombination modifier to each of the two pairs. In our simulations we varied the fitness values over the parameter space in steps of 0.02 and found that in all cases the strength of selection on a modifier linked to two pairs is given to a good approximation by the sum of the selection coefficients acting on modifiers linked to each of the pairs separately, in agreement with the analytical result for selection on the modifier presented in BARTON (1995). For the level of linkage used in Figure 3 the relative error was always  $< 1\%$ .

This means that, at least for two independent pairs, adding up selection coefficients is a reasonable way of bookkeeping interactions between more than two loci. Thus, it is the distribution of the two-locus selection coefficients on the modifier rather than the distribution of epistasis that determines the fate of the recombination modifier (in other words, the key question is where in Figure 3 the interacting loci lie). Therefore, except for the special case where the epistasis distribution does not depend on the strength of selection (*i.e.*, the same epistasis is measured for every selection coefficient), the epistasis distribution can lead to erroneous conclusions about the fate of a recombination modifier. To illustrate this consider the following extreme example: A genome is made up of a modifier and two independent pairs. The first pair is a synthetic lethal ( $w_1 = 1, w_2 = w_3 = 0.99, w_4 = 0.9$ ). The second pair has variable positive epistasis and strong selection on both the single and the double mutant ( $\tilde{w}_1 = 1, \tilde{w}_2 = \tilde{w}_3 = 0.9, \tilde{w}_4 = 0.9^2 + \epsilon_2, \epsilon_2 > 0$ ). From the position of the pairs in Figure 3 one expects that the negatively interacting pair (the synthetic lethal) exerts a much stronger influence than the positively interacting pair, such that a high recombination rate is selected regardless of the average epistasis. Figure 7 shows the selection coefficient on a modifier as a function of  $\epsilon_2$ . As expected, the selection coefficient on the modifier is positive (favoring recombination) even if

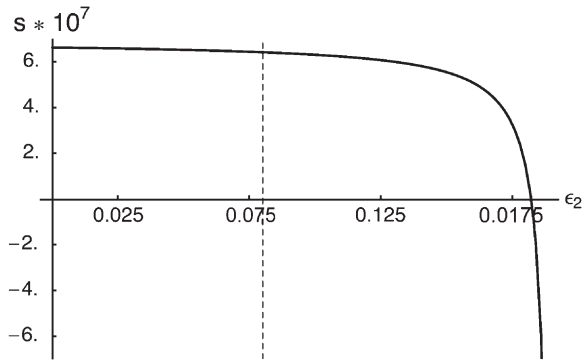


FIGURE 7.—Strength of selection on the modifier linked to two independent pairs of loci. The first pair has fitnesses  $w_1 = 1$ ,  $w_2 = w_3 = 0.99$ ,  $w_4 = 0.9$  and thus a negative epistasis  $\epsilon_1 = -0.089$ . The second pair has fitnesses  $w_2 = w_3 = 0.9$  and  $w_4 = 0.9^2 + \epsilon_2$ , where the positive epistasis  $\epsilon_2$  is plotted on the  $x$ -axis. The dashed line refers to the case where positive and negative epistases are of the same strength (*i.e.*,  $\epsilon_2 = 0.089$ ). Other parameters are  $m = 10^{-3}$ ,  $r_0 = 0$ , and  $\delta r = 0.5 \times 10^{-3}$ .

the positive epistasis of the second pair is considerably stronger than the negative epistasis of the first pair ( $\epsilon_1 = 0.9 - 0.99^2 = -0.081$  represented by the dashed line). The modifier is selected against only if the positive epistasis  $\epsilon_2$  is so large that selection on the double mutant becomes weak. Note that we could have also chosen an example where there is selection against recombination although the average epistasis is negative.

**Epistatic effects in *Saccharomyces cerevisiae*:** Our models suggest that the level of recombination might be governed by a small number of epistatic interactions and that considering epistatic interactions independently of the corresponding levels of selection can lead to erroneous conclusions. Whether this caveat is relevant for the interpretation of natural systems depends mainly on the covariance between single- and double-mutant fitnesses.

As a preliminary examination of this question, we used data on epistatic interactions among mutants in *S. cerevisiae* from SZAFRANIEC *et al.* (2003). The growth rate estimates were converted to fitness estimates of single and double mutants (Figure 8A; data kindly provided by R. Korona); we restricted our attention to the 33 pairs of mutants with fitnesses less than that of the wild type. We ignored measurement error and assumed that the heterozygous mutant fitnesses could be used as proxies for haploid mutant fitnesses. For each pair of mutants, the selection induced on a modifier of recombination was calculated numerically from the *Mathematica* package of OTTO and FELDMAN (1997). Figure 8B illustrates the results, assuming a recombination rate between loci of 0.1. A rare modifier that increases recombination by 10% is selected against according to most pairs of loci. Most importantly, selection against recombination was 16.9 times stronger when the fitness effects of each pair of mutants were considered separately and then aver-

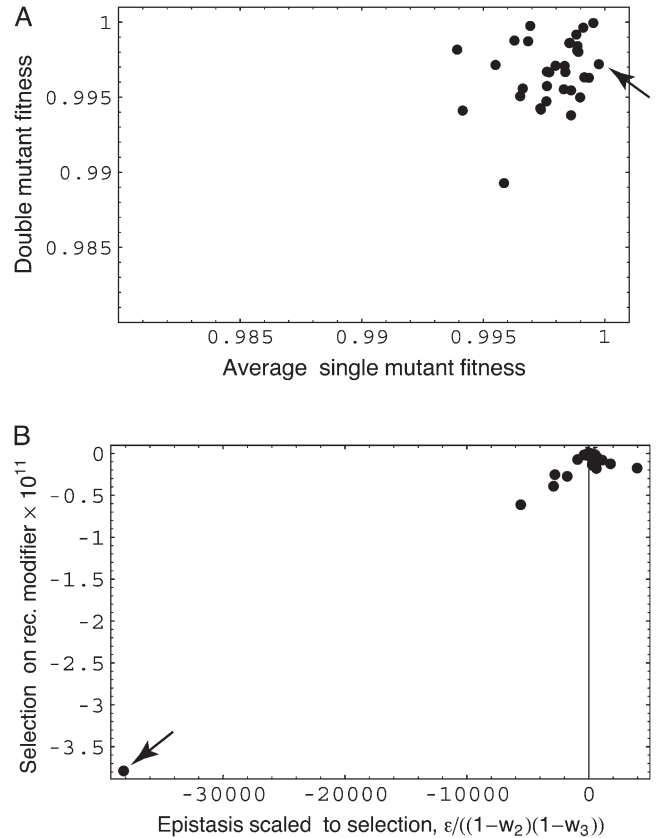


FIGURE 8.—(A) Fitness effects of single and double deleterious mutants from SZAFRANIEC *et al.* (2003). (B) The selection induced on a modifier allele that increases recombination by 10%, as estimated by the leading eigenvalue minus one in a local stability analysis (OTTO and FELDMAN 1997). Selection on a modifier is strongly influenced by a single pair of mutant alleles, with very weak effects in single mutants relative to double mutants (arrows).

aged ( $-2.0 \times 10^{-12}$ ) than when the average fitnesses were used ( $-1.2 \times 10^{-13}$ ). Similar results were obtained when recombination rates of 0.0001, 0.01, or 0.4 were used or when the mutant fitness effects were magnified by a factor of 50. This result demonstrates the importance of considering distributions of fitness effects; relying on average fitness effects generated results that were incorrect by more than an order of magnitude.

DISCUSSION

Recombination has the potential to increase genotypic variation at mutation–selection balance and hence the population’s resistance against mutational decay. However, this effect takes place only under a restricted set of conditions. For a two-locus system, linkage disequilibrium must be negative. A negative linkage disequilibrium in turn is induced by negative epistasis (in the absence of other factors such as drift). Thus, in a two-locus system and an infinite population, recombination

increases the response to selection whenever epistasis is negative. Things are more complicated for multilocus systems, particularly because both negative and positive epistases can be present simultaneously. In this article we investigated the effects of many epistatic interactions on the evolution of recombination.

We found that in general the distribution of linkage disequilibria does not reflect the distribution of epistasis. In particular, a symmetric or even positively skewed epistasis distribution can produce a clearly negatively skewed linkage-disequilibrium distribution. Because it is linkage disequilibrium and not epistasis on which recombination acts, knowledge of epistasis alone is insufficient for drawing conclusions about the nature of selection on a recombination modifier.

The incongruity between the linkage-disequilibrium distribution and the epistasis distribution is due to the fact that the epistasis does not fully determine linkage disequilibrium. To this end we determined what linkage disequilibria and selection coefficients on the modifier can result from a given epistasis in a two-locus system. We found that epistatic interactions of the same strength can lead to very different linkage disequilibria and selection strengths on the modifier. In particular, interactions with weak selection on either single mutant or on the double mutant showed much stronger effects than interactions with strong selection on all mutants. This suggests that in a multilocus system the behavior of a modifier of recombination may be dominated by a small number of interactions, whereas the majority of interactions are irrelevant. In contrast to this disproportionate importance of weakly selected pairs, we found that the time a system requires to reach mutation–selection balance increases steeply as selection diminishes. This could imply that the system is governed by interactions with weak but not too weak selection.

To apply the above findings to systems containing more than one fitness interaction, we considered the simplest such model, *i.e.*, two independent pairs linked to a modifier. We found that the selection coefficient on the modifier was given approximately by the sum of the selection coefficients for each pair if they were considered separately. Thus, in contrast to epistasis, averaging selection coefficients should provide useful information about the fate of a modifier. In particular this result implies that in systems containing two interactions, a modifier can select for increased recombination even if the average epistasis is positive.

Our results indicate that using the average level of epistasis can generate highly misleading predictions about the overall evolutionary force acting on a modifier of recombination. By summing the selection induced on a modifier of recombination over all pairs of interacting loci, we can generate more accurate predictions from empirical data on the fitness effects of single and double mutants. Whether these more accurate predictions will act in favor of or against the

evolution of recombination is an open empirical question, as it depends on the covariance between single- and double-mutant effects. Using estimates of the selection coefficients from *S. cerevisiae* confirmed that averaging epistatic effects results in inaccurate estimates of the selection on a modifier of recombination. In this example, a single pair of weakly selected loci dominated all other fitness interactions. The strong influence of pairs of weakly selected loci has an additional important consequence for the interpretation of experimental data: As can be seen in Figures 3 and 5 the selection coefficient on the modifier depends very sensitively on  $w_{2,3}$  if selection is weak. This implies that small measurement errors in the fitness values of these weakly selected loci can have a strong impact on the estimated selection on the recombination modifier.

A large number of studies have attempted to draw conclusions on the evolution of recombination based on the distribution of epistasis (WEST *et al.* 1998; ELENA 1999; DE LA PEÑA *et al.* 2000; PETERS and KEIGHTLEY 2000; BONHOEFFER *et al.* 2004). However, we have shown that to interpret experimental data in the context of the evolution of recombination, it is necessary to consider not only the distribution of epistatic interactions, but also the associated selection coefficients.

We thank Lucy Crooks, Marcel Salathe, and Martin Ackermann for helpful discussions. R.D.K. and S.B. gratefully acknowledge support from the Swiss National Foundation.

#### LITERATURE CITED

- BARTON, N., 1995 A general model for the evolution of recombination. *Genet. Res.* **65**: 123–145.
- BONHOEFFER, S., C. CHAPPEY, N. PARKIN, J. WHITCOMB and C. PETROPOULOS, 2004 Evidence for positive epistasis in HIV-1. *Science* **306**: 1547–1550.
- DE LA PEÑA, M., S. ELENA and A. MOYA, 2000 Effect of deleterious mutation-accumulation on the fitness of RNA bacteriophage MS2. *Evol. Int. J. Org. Evol.* **54**: 686–691.
- DE VISSER, J., R. HOEKSTRA and H. VAN DEN ENDE, 1997 An experimental test for synergistic epistasis and its application in *Chlamydomonas*. *Genetics* **145**: 815–819.
- ELENA, S., 1999 Little evidence for synergism among deleterious mutations in a nonsegmented RNA virus. *J. Mol. Evol.* **49**: 703–707.
- ELENA, S., and R. LENSKI, 1997 Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**: 395–398.
- ESHEL, I., and M. FELDMAN, 1970 On the evolutionary effect of recombination. *Theor. Popul. Biol.* **1**: 88–100.
- KITAGAWA, O., 1967 Interaction in fitness between lethal genes in heterozygous condition in *Drosophila melanogaster*. *Genetics* **57**: 809–820.
- MUKAI, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**: 1–19.
- OTTO, S., and M. FELDMAN, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* **51**: 134–147.
- OTTO, S., and T. LENORMAND, 2002 Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**: 252–261.
- PETERS, A., and P. KEIGHTLEY, 2000 A test for epistasis among induced mutations in *Caenorhabditis elegans*. *Genetics* **156**: 1635–1647.
- PHILLIPS, P., S. OTTO and M. WHITLOCK, 2000 Beyond the average: the evolutionary importance of gene interactions and variability



of epistatic effects, pp. 20–38 in *Epistasis and the Evolutionary Process*, edited by J. WOLF, E. I. BRODIE and M. WADE. Oxford University Press, Oxford.

SZAFRANIEC, K., D. WŁOCH, P. SLIWA, R. BORTS and R. KORONA, 2003 Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast *Saccharomyces cerevisiae*. *Genet. Res.* **82**: 19–31.

WEST, S., A. PETERS and N. BARTON, 1998 Testing for epistasis between deleterious mutations. *Genetics* **149**: 435–444.

WŁOCH, D., R. BORTS and R. KORONA, 2001 Epistatic interactions of spontaneous mutations in haploid strains of the yeast *Saccharomyces cerevisiae*. *J. Evol. Biol.* **14**: 310–316.

Communicating editor: H. G. SPENCER

APPENDIX: EFFECTS OF A SMALL RECOMBINATION RATE

For a two-locus/two-allele system with fitnesses ( $w_1 = 1, w_2, w_3, w_4$ ), recombination rate  $r$ , mutation rate  $m$ , and no back mutation the recursion relations read

$$\begin{aligned} f_1'' &= (1 - m)^2 \frac{1}{\bar{w}} f_1 \\ f_2'' &= (1 - m) \frac{w_2}{\bar{w}} f_2 + m(1 - m) \frac{1}{\bar{w}} f_1 \\ f_3'' &= (1 - m) \frac{w_3}{\bar{w}} f_3 + m(1 - m) \frac{1}{\bar{w}} f_1 \\ f_4'' &= \frac{w_4}{\bar{w}} f_4 + m \frac{w_2}{\bar{w}} f_2 + m \frac{w_3}{\bar{w}} f_3 + m^2 \frac{1}{\bar{w}} f_1 \\ f_1''' &= f_1'' - r[f_1'' f_4'' - f_2'' f_3''] \\ f_2''' &= f_2'' + r[f_1'' f_4'' - f_2'' f_3''] \\ f_3''' &= f_3'' + r[f_1'' f_4'' - f_2'' f_3''] \\ f_4''' &= f_4'' - r[f_1'' f_4'' - f_2'' f_3'']. \end{aligned}$$

(see *Model* section). We obtain the steady-state equations by setting  $f_i''' = f_i''$  in the recursion relations.

For  $r = 0$  the steady-state equations are simple to solve (see *Effects of epistasis in two-locus systems*) and yield the frequencies (for  $m < 1 - w_2$ )

$$\begin{aligned} f_1^0 &= \frac{1}{\mathcal{N}} (1 - m - w_2)(1 - m - w_3)((1 - m)^2 - w_4) \\ f_2^0 &= \frac{1}{\mathcal{N}} m(1 - m - w_3)((1 - m)^2 - w_4) \\ f_3^0 &= \frac{1}{\mathcal{N}} m(1 - m - w_2)((1 - m)^2 - w_4) \\ f_4^0 &= \frac{1}{\mathcal{N}} m^2((1 - m)^2 - w_2 w_3), \end{aligned}$$

with  $\mathcal{N} = 2m(1 - w_2)(1 - w_3) - m^2(1 - w_2 - w_3 + w_4) - (1 - w_2)(1 - w_3)(1 - w_4)$ .

For  $r \neq 0$  the steady-state frequencies take the form

$$f = f^0 + rx,$$

with  $x = O[1]$ . For small  $r$ , we insert this expression in the steady-state equations and keep only terms linear in  $r$  (there are no terms independent of  $r$ , because  $f^0$  solves the steady-state equations for  $r = 0$ ). Because  $x$  occurs in the equations only as  $x \times r$ , this results in a linear equation for  $x$ , which can easily be solved.

The change in linkage disequilibrium then reads

$$\begin{aligned} \Delta D &= [(f_1^0 + rx_1)(f_4^0 + rx_4) - (f_2^0 + rx_2)(f_3^0 + rx_3)] \\ &\quad - [f_1^0 f_4^0 - f_2^0 f_3^0] \\ &= r(f_1^0 x_4 + f_4^0 x_1 - f_2^0 x_3 - f_3^0 x_2) + O[r^2] \end{aligned}$$

and the change in mean fitness is

$$\begin{aligned} \Delta \bar{w} &= \sum_{i=1}^4 w_i (f_i^0 + rx_i) - \sum_{i=1}^4 w_i f_i^0 \\ &= r(w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4). \end{aligned}$$

If we insert the solution for  $x$ ,  $\Delta D$  and  $\Delta \bar{w}$  become complicated functions of  $w_i$  and  $m$ . However, by keeping only lowest-order terms in  $m$ , we obtain the expressions for  $\Delta D$  and  $\Delta w$  referred to in *Effects of epistasis in two-locus systems*.