

# Testing for Epistasis Between Deleterious Mutations

S. A. West,\* A. D. Peters<sup>†</sup> and N. H. Barton\*

\**Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom*  
and <sup>†</sup>*Department of Biology, Indiana University, Bloomington, Indiana, 47405*

Manuscript received September 18, 1997

Accepted for publication February 4, 1998

## ABSTRACT

Determining the way in which deleterious mutations interact in their effects on fitness is crucial to numerous areas in population genetics and evolutionary biology. For example, if each additional mutation leads to a greater decrease in log fitness than the last (synergistic epistasis), then the evolution of sex and recombination may be favored to facilitate the elimination of deleterious mutations. However, there is a severe shortage of relevant data. Three relatively simple experimental methods to test for epistasis between deleterious mutations in haploid species have recently been proposed. These methods involve crossing individuals and examining the mean and/or skew in log fitness of the offspring and parents. The main aim of this paper is to formalize these methods, and determine the most effective way in which tests for epistasis could be carried out. We show that only one of these methods is likely to give useful results: crossing individuals that have very different numbers of deleterious mutations, and comparing the mean log fitness of the parents with that of their offspring. We also reconsider experimental data collected on *Chlamydomonas moewussi* using two of the three methods. Finally, we suggest how the test could be applied to diploid species.

**D**ETERMINING the way in which deleterious mutations interact in their effects on fitness is crucial to numerous areas in population genetics and evolutionary biology (reviewed by Kondrashov 1993). These interactions may take three forms: (1) independently (multiplicative selection), (2) each additional mutation leading to a greater decrease in log fitness than the last (synergistic epistasis), or (3) each additional mutation leading to a smaller decrease in log fitness than the last (antagonistic epistasis). It is convenient to work in log fitness because on this scale the three possibilities can be distinguished by the different relations that they predict with the number of deleterious mutations (Charlesworth 1990; Figure 1). Multiplicative selection leads to log fitness decreasing linearly with increasing number of mutations. In contrast, synergistic and antagonistic epistases lead to nonlinear curves, with the slope declining more (concave) and less (convex) steeply as the number of deleterious mutations increases, respectively.

Many theoretical studies have relied on the assumption that deleterious mutations interact with synergistic epistasis. For example, if deleterious mutations interact with synergistic epistasis then sexual reproduction and recombination provide an advantage over asexual reproduction because they enable individuals to better eliminate deleterious mutations (the Mutational Deterministic hypothesis; Kondrashov 1982, 1984; Charlesworth 1990). Synergistic epistasis leads to this predic-

tion because a relatively large number of deleterious mutations will be eliminated from sexual populations in the low quality offspring that have particularly high numbers of deleterious mutations. Given a sufficient mutation rate, this advantage can be more than sufficient to balance the two-fold cost of sex. However, evidence for synergistic epistasis comes primarily from the non-linear fitness decline in a mutation accumulation experiment with *Drosophila melanogaster* (Mukai 1969), and several possible problems with such experiments have been pointed out (Keightley 1996).

de Visser *et al.* (1996, 1997a) have recently suggested that tests for epistasis between deleterious mutations can be conducted in haploid species by crossing two individuals and examining the mean and/or skew in log fitness of their offspring (before selection has taken place). They used their techniques to test for synergistic epistasis in *Chlamydomonas moewussi*, a haploid unicellular algae. This novel approach is particularly important because it provides a relatively simple way to experimentally test for synergistic epistasis. Here, we extend the underlying theory, and determine the most effective way in which tests for epistasis could be carried out. We also reconsider the experimental data on *C. moewussi* that de Visser *et al.* (1996, 1997a) collected using two of these methods. Finally, we suggest how the test could be applied to diploid species.

## MEAN LOG FITNESS

de Visser *et al.* (1996) pointed out that, in a haploid species, comparing the mean log fitness of sexually produced offspring with that of their two parents (or asexu-

Corresponding author: S. A. West, Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.  
E-mail: stu.west@ed.ac.uk

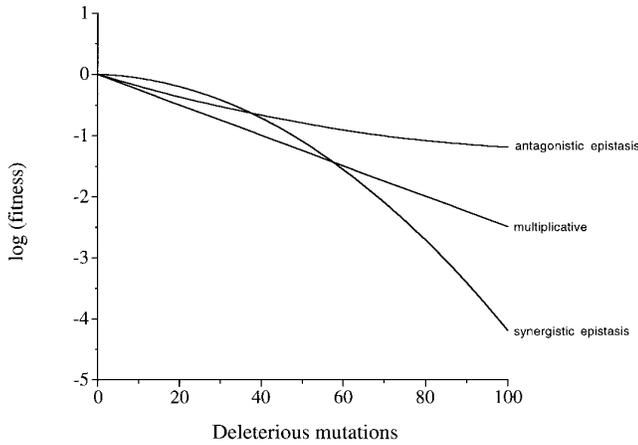


Figure 1.—The relationship between log fitness and number of deleterious mutations. The lines represent the three ways in which deleterious mutations may interact in their effects on fitness: antagonistic epistasis ( $\alpha = 0.02$ ,  $\beta = -0.00008$ ); multiplicative selection ( $\alpha = 0.025$ ,  $\beta = 0$ ), and synergistic epistasis ( $\alpha = 0.002$ ,  $\beta = 0.0008$ ).

ally produced offspring) can provide information about the way in which deleterious mutations interact. The sexually produced offspring of a cross between two individuals will, due to random segregation and recombination, have a symmetrical (*i.e.*, binomial) distribution of mutations per individual, with the mean equal to that of their parents. If deleterious mutations interact multiplicatively, then the mean offspring log fitness will equal the mean log fitness of the parent lines. However, if deleterious mutations show epistasis, then the mean log fitness of the offspring lines can differ from that of their parents. If the parents have identical or very similar numbers of deleterious mutations, then the mean offspring log fitness is predicted to be lower (synergistic epistasis) or greater (antagonistic epistasis) than that of their parents. In contrast, if the parents have very different numbers of deleterious mutations, then the mean offspring log fitness is predicted to be greater (synergistic epistasis) or lower (antagonistic epistasis) than that of their parents.

These predictions can be understood intuitively by comparing the variance in the number of deleterious mutations amongst sexually produced offspring with that amongst their two parents. For example, with synergistic epistasis, increasing the variance in the number of deleterious mutations decreases the mean log fitness because of the particularly low fitness experienced by individuals with high numbers of deleterious mutations (or relatively high fitness of individuals with intermediate numbers of deleterious mutations). If the two parents have equal (or very similar) numbers of deleterious mutations then they will have zero (or very small) variance in the number of deleterious mutations. Consequently, the variance in the number of deleterious mutations amongst the offspring will be greater and so synergistic epistasis would lead to the offspring having

a lower mean log fitness. In contrast, if the two parents have very different numbers of deleterious mutations, then their variance in numbers of deleterious mutations will be greater than that amongst their offspring, and so synergistic epistasis would lead to the offspring having a greater mean log fitness.

We now formalize and quantify the above argument. Consider two haploid parents which have exactly  $n_1$  and  $n_2$  deleterious mutations, respectively. The mean ( $M_p$ ) and variance ( $V_p$ ) in the number of deleterious mutations per individual in the parents (or their combined asexual offspring) are  $M_p = (n_1 + n_2)/2$ , and  $V_p = (n_1 - n_2)^2/4$ , respectively. In contrast, assuming random segregation and free recombination, the mean ( $M_o$ ) and variance ( $V_o$ ) in the number of deleterious mutations carried by the offspring can be obtained by summing two binomial distributions, and are given by  $M_o = (n_1 + n_2)/2$ , and  $V_o = (n_1 + n_2)/4$ , respectively. Notice that, although the means are identical, the variances differ. Two important limiting cases, which confirm the verbal arguments given above, are: (1) if  $n_1 = n_2$  then  $V_p = 0$ , and so  $V_o > V_p$  (assuming  $n_1 > 0$ ), and (2) if  $n_1 \gg n_2$  then  $V_p > V_o$  ( $V_p \approx n_1^2/4$ ;  $V_o \approx n_1/4$ ).

In order to determine the difference in mean log fitness between offspring and parents, it is necessary to assume a relationship between the number of deleterious mutations and fitness. Following Charlesworth (1990) we assume that log fitness ( $w$ ) follows a quadratic function with the number of deleterious mutations ( $n$ ):

$$\log(w) = -\left(\alpha n + \frac{\beta}{2}n^2\right). \quad (1)$$

The various ways in which deleterious mutations interact are then represented by:  $\beta > 0$  (synergistic epistasis);  $\beta = 0$  (multiplicative selection), and  $\beta < 0$  (antagonistic epistasis). The ratio  $\beta/\alpha$  measures the degree of epistasis (Charlesworth *et al.* 1990). This quadratic relationship was chosen to facilitate comparison with previous studies, and because it allows for a wide variety of forms of epistasis.

We present results for a wide range of values of  $\alpha$  and  $\beta$ , representing a variety of forms of epistasis. Mukai's (1969) mutation accumulation experiment with *D. melanogaster* suggested that the values of  $\alpha$  and  $\beta$  are: 0.002 and 0.0008, respectively, when deleterious mutations are heterozygous (assuming that the coefficient of dominance is 0.2), and 0.014 and 0.011, respectively, when deleterious mutations are homozygous (Crow 1970; Charlesworth 1990). The homozygous coefficients may be more appropriate for haploids. However, in the discussion we suggest that these methods can also be applied to diploids, and so present results over parameter values that include both of these possibilities. In addition, it has been suggested that Mukai's experiment may have overestimated the extent of epistasis (Keightley 1996; Charlesworth 1998). An important general

point to note here is that different ranges in the numbers of deleterious mutations per parent should be considered for different values of  $\alpha$  and  $\beta$ . For example, when  $\alpha$  and  $\beta$  are 0.002 and 0.0008, respectively, 75 deleterious mutations are required for the fitness of an individual to fall below 0.01. In contrast, with values of 0.014 and 0.011 for  $\alpha$  and  $\beta$ , only 20 deleterious mutations are required for such a decline in fitness.

In order to use Equation 1 to calculate the mean log fitness of parents and their offspring, we must calculate the mean fitness of a group of individuals with a given mean ( $\bar{n}$ ) and variance ( $V_n$ ) in the number of deleterious mutations per individual. The mean of the number of mutations squared will be given by the equation  $\overline{n^2} = \bar{n}^2 + V_n$ , and so the mean log fitness will be

$$\overline{\log(w)} = -\left(\alpha\bar{n} + \frac{\beta}{2}(\bar{n}^2 + V_n)\right) \quad (2)$$

(Charlesworth and Barton 1996). Note that, with synergistic epistasis, increasing the variance in deleterious mutations leads to a decrease in log fitness. The mean log fitness of the parents ( $w_p$ ) and their offspring ( $w_o$ ) can be calculated by substituting  $M_p$ ,  $V_p$ ,  $M_o$  and  $V_o$  into Equation 2. The difference in mean log fitness between the offspring and their parents [ $\Delta\overline{\log(w)} = \overline{\log(w_o)} - \overline{\log(w_p)}$ ] is given by

$$\begin{aligned} \Delta\overline{\log(w)} &= \frac{\beta}{8} (n_1^2 + n_2^2 - 2n_1n_2 - n_1 - n_2) \\ &= \frac{\beta}{2}(V_p - V_o) . \end{aligned} \quad (3)$$

The difference in mean log fitness is linearly related to the extent of epistasis ( $\beta$ ), and the difference in variance of numbers of deleterious mutations per individual between parents and their offspring ( $V_p - V_o$ ). The value of  $\alpha$  does not enter into Equation 3, and so has no effect on the predicted difference. Note that, for small changes in fitness,  $\log(w_o) - \log(w_p) \approx w_o - w_p$  (Burt 1995).

Let us first consider the case where the two parents have equal numbers of deleterious mutations ( $n_1 = n_2 = M_p$ ). Equation 3 then simplifies to

$$\Delta\overline{\log(w)} = -\frac{\beta M_p^2}{4}$$

[Figure 2; following Charlesworth (1990), only a narrow range of negative  $\beta$  values are shown]. Variation in  $\alpha$  has no effect. Larger differences in log fitness are predicted with greater numbers of deleterious mutations. So, if one were testing for epistasis it would be better to do so with lines that had accumulated large numbers of deleterious mutations. However, it should also be noted that, even with large numbers of deleterious mutations, the predicted differences in log fitness,

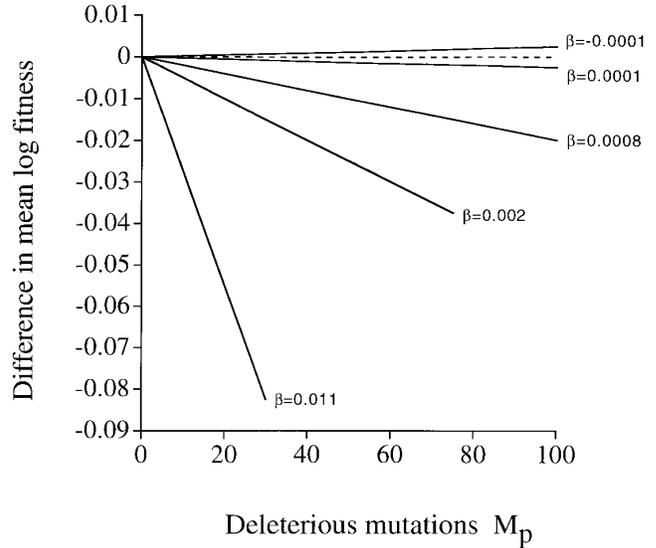


Figure 2.—The predicted difference in mean log fitness between offspring and their parents, when the parents have equal numbers of deleterious mutations. The lines represent different values of  $\beta$ ; variation in  $\alpha$  has no effect. The lines in which  $\beta = 0.011$  and  $\beta = 0.002$  are only continued up until when each parent has 30 and 75 deleterious mutations, respectively, because above this their fitness would be incredibly low ( $<0.00001$  even if  $\alpha = 0$ ).

for realistic values of  $\beta$ , are extremely small and would be very hard to detect experimentally.

We shall now consider the case where two parents have different numbers of deleterious mutations. Some examples are plotted in Figure 3 and illustrate two points. First, the number of deleterious mutations in the two parents need only differ by a small amount to predict a difference in mean log fitness equal to zero or in the opposite direction to that predicted with exactly equal numbers of deleterious mutations. This will be a problem if the number of deleterious mutations in each parent cannot be directly measured and so must be inferred. Consequently, when crossing two lines with approximately equal numbers of deleterious mutations, all possible results could be explained by synergistic or antagonistic epistasis!

The second point illustrated by Figure 3 is that crossing lines with very different numbers of deleterious mutations can lead to much larger differences in mean log fitness than crossing lines with the same number of deleterious mutations. Furthermore, these differences should be large enough to be experimentally detectable. Crossing lines with relatively low numbers of deleterious mutations with lines with relatively large numbers of deleterious mutations would, therefore, appear to be the best way to use differences in mean log fitness to test for epistasis between deleterious mutations.

**Mean log fitness, equilibrium populations and the recombination load:** In any population at equilibrium, there will be variation in the number of deleterious mutations

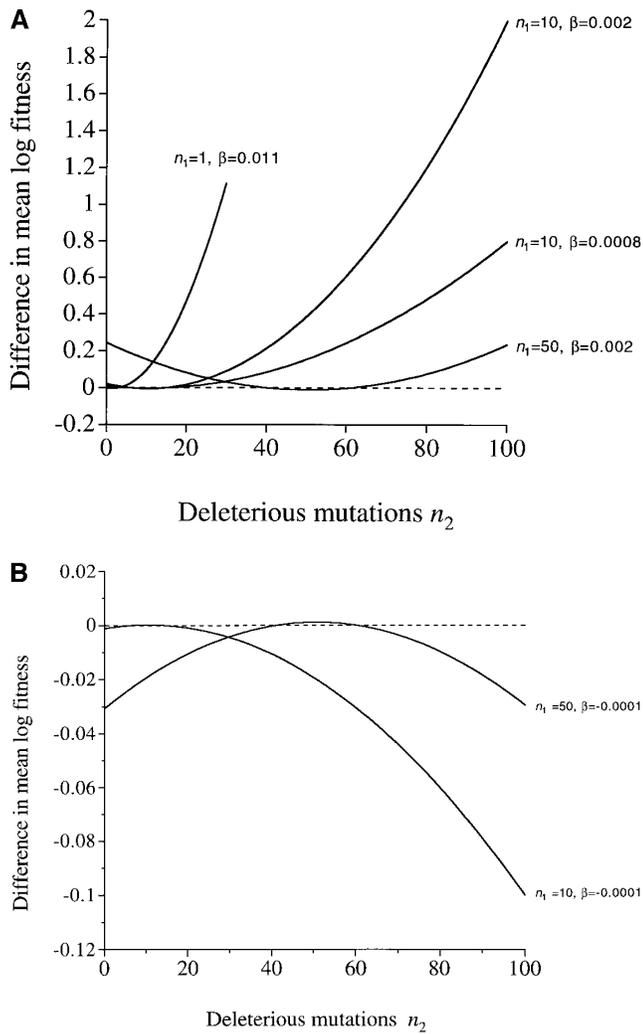


Figure 3.—The predicted difference in mean log fitness between offspring and their parents as the number of deleterious mutations in one parent ( $n_2$ ) varies. The other parent is assumed to have a fixed number of deleterious mutations ( $n_1$ ). (A) Synergistic epistasis. (B) Antagonistic epistasis.

per individual. Consequently, random mating with either synergistic or antagonistic epistasis may lead to a decrease in mean log fitness with some matings, and an increase with others. The aim of this section is to consider: (1) the change in mean log fitness that is predicted from random mating in a sexual population with synergistic epistasis, and (2) how often the difference in numbers of deleterious mutations, between two individuals drawn at random from a population, will be large enough to predict that their offspring have a higher mean log fitness. In addition to testing for synergistic epistasis these results have implications for the possible role of deleterious mutations in causing a recombination load (an immediate reduction in fitness due to recombination; Charlesworth and Barton 1996).

Consider a haploid population in which the number of deleterious mutations per individual is normally distributed with a given mean ( $\bar{n}$ ) and variance ( $V_n$ ) (Charles-

TABLE 1

The average difference between mean offspring log fitness and mean parent log fitness (recombination load) in sexual populations at equilibrium

Mutation rate ( $U$ )	Deleterious mutations per individual		Average difference in log fitness ( $\overline{\Delta \log(\bar{w})}$ )
	Mean ( $\bar{n}$ )	Variance ( $V_n$ )	
2.0	50.8	49.2	-0.000320
1.5	43.6	42.4	-0.000240
1.0	35.0	34.2	-0.000160
0.5	24.1	23.7	-0.000080
0.1	9.8	9.8	-0.000014

All cases, moderate synergistic epistasis was assumed ( $\alpha = 0.002$ ,  $\beta = 0.0008$ ). The values for  $\bar{n}$  and  $V_n$  are from Charlesworth (1990), and differ due to the linkage disequilibrium that is caused by epistasis.

worth 1990; Barton 1995). Substituting into Equation 3, the mean difference between mean offspring log fitness and mean parent log fitness [ $\overline{\Delta \log(\bar{w})}$ ] will be

$$\overline{\Delta \log(\bar{w})} = \frac{\beta}{4}(V_n - \bar{n}). \quad (4)$$

This value increases with the extent of epistasis ( $\beta$ ), and as the difference between the mean and the variance in number of deleterious mutations per individual increases (see also Charlesworth and Barton 1996). Note also that increasing the extent of epistasis leads to greater linkage disequilibrium, and so increases the difference between the mean and the variance in number of deleterious mutations per individual (Charlesworth 1990; Barton 1995; Charlesworth and Barton 1996).

Charlesworth (1990) calculated the mean and the variance in number of deleterious mutations per individual for five different genomic deleterious mutation rates. In all cases, moderate synergistic epistasis was assumed ( $\alpha = 0.002$ ,  $\beta = 0.0008$ ). The subsequent predicted mean difference between mean offspring log fitness and mean parent log fitness are given in Table 1. This table illustrates two points about the predicted difference: in all cases it is extremely small ( $<0.0004$ ), making it practically impossible to detect experimentally (see also Charlesworth and Barton 1996); and it rises with the mutation rate, because of an increasing difference between the mean and variance in the number of deleterious mutations per individual. It is worth noting here that the above predictions assume that fitness itself is measured. While it is the actual fitness that determines the variance in the number of deleterious mutations, it will often only be experimentally possible to measure a component of fitness.

We also carried out simulations in order to determine how often the difference in numbers of deleterious mu-

**TABLE 2**  
**The difference between mean offspring log fitness and mean parent log fitness**  
**(recombination load) in simulated sexual populations**

Mutation rate ( $U$ )	Observed deleterious mutations per individual <sup>b</sup>		Mean difference in log fitness ( $\Delta\log(\bar{w})$ ) <sup>a</sup>	Proportion of matings in which $\Delta\log(\bar{w}) > 0$
	Mean ( $\bar{n}$ )	Variance ( $V_n$ )		
2.0	50.74	45.82	-0.000788 (0.001081)	0.312
1.5	43.74	41.34	-0.000955 (0.000959)	0.320
1.0	34.83	33.98	-0.000126 (0.000980)	0.314
0.5	24.14	22.36	-0.000400 (0.000575)	0.292
0.1	9.88	9.54	-0.000145 (0.000244)	0.284

Deleterious mutations were assumed to be normally distributed, with the mean and variance predicted for that mutation rate (Table 1; Charlesworth 1990). In all cases, moderate synergistic epistasis was assumed ( $\alpha = 0.002$ ,  $\beta = 0.0008$ ).

<sup>a</sup> 95% confidence intervals are shown in parentheses.

<sup>b</sup> Observed in 1000 simulated individuals.

tations, between two individuals drawn at random from a population, is large enough to predict that their offspring have a higher mean log fitness. We assumed a normal distribution, and randomly assigned mutations to 1000 individuals in 500 mating pairs. The mean log offspring fitness and mean log parent fitness were calculated for each mating pair, and the process repeated for each of the values of  $\bar{n}$  and  $V_n$  corresponding to the mutation rates examined by Charlesworth (1990). The results are given in Table 2. As predicted, very small differences between mean offspring log fitness and mean parent log fitness were observed. These differences were not significantly different from zero, despite sample sizes of 500 and no measurement error, demonstrating how difficult they would be to detect experimentally. However, in all cases, approximately 30% of the matings led to the mean offspring log fitness being higher than the mean parent log fitness ( $\Delta\log(\bar{w}) > 0$ ). In these matings the number of deleterious mutations in the two parents were different enough that their variance was more than in their offspring ( $V_p > V_o$ ). The reason why there were only very small average differences overall, is that the difference in mean log fitness between parents and their offspring is larger when the offspring log fitness is greater (the parents have different numbers of deleterious mutations; Figure 3), than when the offspring log fitness is lower (the parents have "similar" numbers of deleterious mutations; Figure 2).

These results illustrate a possibly important point to bear in mind when crossing lines with "similar" levels of mutations and examining the change in mean log fitness. The fact that approximately 70% of matings led to the mean offspring log fitness being lower than the mean parent log fitness ( $\Delta\log(\bar{w}) < 0$ ) suggests that it might be useful to cross numerous different lines and then analyze the results with a sign test. However, because  $|\Delta\log(\bar{w})|$  is, on average, smaller when negative, any variance in the

data (e.g., environmental or measurement) is likely to make more negative values of  $\Delta\log(\bar{w})$  become positive than vice versa. This would make acceptance of the null hypothesis more likely (type II error). In addition, this problem would be increased if there was skew in the error. Nonetheless, these results also emphasise the importance of crossing numerous distinct lines (i.e., replication at the line level); 30% of matings between individuals that came from the same population, and so might have been expected to have "similar" numbers of deleterious mutations, resulted in the variance in numbers of deleterious mutations per individual being greater in the parents than in their offspring ( $V_p > V_o$ ). In these matings, the mean offspring log fitness is predicted to be greater than the mean parent log fitness.

Equation 4 predicts the recombination load due to epistasis between deleterious mutations in a haploid sexual population at equilibrium. This equation would also hold for a diploid species in which all deleterious mutations were heterozygous. Our results (Tables 1 and 2) therefore agree with Charlesworth and Barton (1996), that this is too small to be able to explain the recombination load that has been observed in experiments with *D. melanogaster*.

#### SKEW IN OFFSPRING LOG FITNESS

de Visser *et al.* (1997a) pointed out that, in a haploid species, examining whether the distribution of log fitness amongst the offspring of sexual crosses are symmetrical or skewed can provide information about the way in which deleterious mutations interact. As noted above, these offspring will have a symmetrical distribution of mutations per individual, with the mean equal to that of their parents. If deleterious mutations interact multiplicatively, then the distribution of offspring log fitness will also be symmetrical. However, if deleterious muta-

tions show epistasis, then this distribution will be skewed. Synergistic epistasis would lead to a negative skew, and antagonistic epistasis to a positive skew. As with mean log fitness, this is due to the effects of the relatively low (synergistic epistasis) or high (antagonistic epistasis) fitness of individuals with relatively large numbers of deleterious mutations.

If we assume that deleterious mutations are normally distributed in the progeny (Charlesworth 1990; Barton 1995; Charlesworth and Barton 1996), then by taking moments over Equation 1, the skew in offspring log fitness ( $g_i$ ) simplifies to

$$g_i = \frac{-\left(\frac{M_p^3\beta^3}{8} + \frac{3M_p^2\beta(\alpha + M_p\beta)^2}{4}\right)}{\left(\frac{M_p^2\beta^2}{8} + \frac{M_p(\alpha + M_p\beta)^2}{2}\right)^{3/2}}. \quad (5)$$

The numerator represents the third central moment of log fitness, while the denominator is the cube of its standard deviation. Example relationships for different values of  $\alpha$  and  $\beta$  are given in Figure 4, and illustrate three important points. First, as expected, synergistic epistasis leads to a negative skew, and antagonistic epistasis to a positive skew. Second, the predicted values of skew are very small for realistic parameter estimates, and would be very hard to detect experimentally. For large samples ( $n > 150$ ) the standard error of the skew statistic is  $\approx \sqrt{6/n}$  (Sokal and Rohlf 1981), and so the minimum sample size required to detect a certain amount of skew, assuming no error variance in the fitness estimates, will be  $\approx 24/g_i^2$ , which will be very large for the values of skew predicted here. For example, the minimum sample sizes required to detect the skew predicted by the estimates from *D. melanogaster* ( $\alpha = 0.002$ ,  $\beta = 0.0008$ ) are several hundred for most mean numbers of deleterious mutations in the parents. Replicating such experiments could be extremely hard.

Finally, the shape of the relationship between predicted skew and mean number of deleterious mutations depends upon the values of  $\alpha$  and  $\beta$ . Equation 5 predicts a domed or inverse domed shape, reaching a maximum/minimum when  $\partial g_i / \partial M_p = 0$ , which occurs when  $M_p = \alpha/\beta$  (synergistic epistasis;  $\beta > 0$ ) or when  $M_p = -\alpha/\beta$  (antagonistic epistasis;  $\beta < 0$ ). With synergistic epistasis, the maximum predicted skew is  $-(\alpha^3/8 + 3\alpha^4/\beta)/(\alpha^2/8 + 2\alpha^3/\beta)^{3/2}$ , which increases with  $\beta$  and decreases with  $\alpha$ . The form of the relationship between skew and number of deleterious mutations, over the appropriate range of mean number of deleterious mutations in the parents, therefore depends upon the relative magnitude of  $\alpha$  and  $\beta$ . For example, if  $\alpha \gg \beta$  then the maximum skew occurs with very high mean numbers of deleterious mutations, and so the magnitude of skew will generally increase with rising numbers of deleterious mutations. In contrast, if  $\alpha \approx \beta$ , then the maximum skew occurs with only one deleterious mutation in each parent, and so the magnitude of skew will decrease with

rising numbers of deleterious mutations. Between these two extremes there will be a large range of values of  $\alpha$  and  $\beta$  where the skew will peak at intermediate numbers of deleterious mutations. The most important consequence of this variable relationship is that it would be almost impossible to carry out control crosses where one would expect less skew due to deleterious mutations. Such a control would be crucial to allow for epistasis between favorable alleles, a point that we shall return to in the discussion.

Another possible problem with this method is that, when measuring fitness, there may be skew in the error. This could arise environmentally or through the method by which fitness is estimated. The crucial point here is that such a source of variation generates a third moment which would increase with the variance. In order for this to have no effect on the predicted skew it would require the extremely restrictive assumption that this third moment scale with the variance  $^{3/2}$ .

## DISCUSSION

de Visser *et al.* (1996, 1997a) suggested three methods to test for epistasis between deleterious mutations in haploid species: (1) crossing two individuals with similar numbers of deleterious mutations and comparing the mean log fitness of the two parent lines with their offspring; (2) crossing two individuals with very different numbers of deleterious mutations and comparing the mean log fitness of the two parent lines with their offspring, and (3) examining whether the distribution of log fitness amongst sexually produced offspring is symmetrical or skewed. Our results suggests that method 2 (crossing two individuals with very different numbers of deleterious mutations) is by far the best method, and the only one likely to give clear results.

There are several problems with the other two methods. Method 1 (crossing two individuals with similar numbers of deleterious mutations) is unsuitable because: (i) the number of the deleterious mutations in the two parents need only differ by a small amount to give a difference in mean log fitness between parents and their offspring equal to zero or in the opposite direction to that predicted with exactly equal numbers of deleterious mutations, (ii) even when the two parents have exactly the same number of deleterious mutations, it predicts very small differences in mean log fitness, and (iii) because such small differences are predicted, it is hard to carry out a control where smaller differences are predicted due to epistasis between deleterious mutations. Method 3 (testing for skew in log fitness) is unlikely to give clear results because: (i) the predicted relationship between skew and average number of deleterious mutations can be domed, and so it is impossible to carry out control crosses where one would expect less skew due to epistasis between deleterious mutations, (ii) very small values of skew are predicted that would be very hard to detect statistically, (iii) the enormous

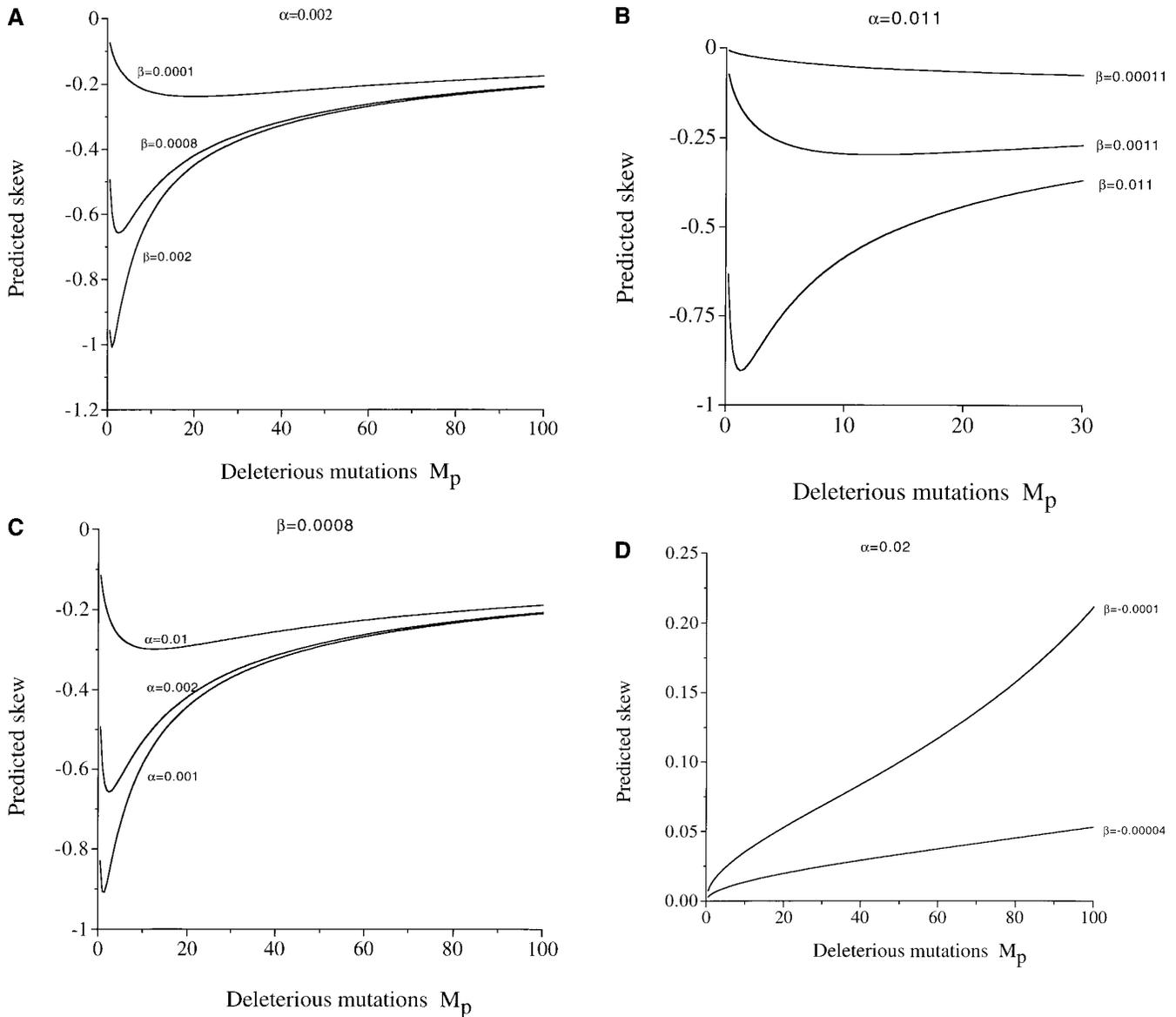


Figure 4.—The relationship between predicted skew and mean number of deleterious mutations in the parents  $M_p$ . (A) Synergistic epistasis ( $\alpha = 0.002$ ,  $\beta = \text{variable}$ ). (B) Synergistic epistasis ( $\alpha = 0.014$ ,  $\beta = \text{variable}$ ). (C) Synergistic epistasis ( $\alpha = \text{variable}$ ,  $\beta = 0.0008$ ). (D) Antagonistic epistasis ( $\alpha = 0.02$ ,  $\beta = \text{variable}$ ).

sample sizes required to detect skew in a cross between two lines means that it would be hard to replicate with crosses between different parents, and (iv) it does not allow for skew in the error.

The problems for methods 1 and 3 would be increased by experimental measurement (replication) error, which our theoretical predictions do not take into account. The importance of this would depend enormously upon the type of organism used in any experiments. If a species is being used where genotypes can be cloned and replicated to a high degree, then the problem can be effectively ignored. However, if this is not possible, then it could make detecting small effects much harder. This would increase the problems for methods 1 and 3, where very small effects are predicted. It would be much less of a problem for method 2 (cross-

ing individuals with different numbers of deleterious mutations), where large differences in mean log fitness are predicted.

The methods that examine mean log fitness (methods 1 and 2) require assumptions to be made about the number of deleterious mutations in different individuals (de Visser *et al.* 1996, 1997a). The number of deleterious mutations in an individual will never be exactly known, and so must be inferred by their relative fitness. This is a problem because the fitness consequences and form of epistasis of deleterious mutations differ (Keightley 1994, 1996; Whitlock *et al.* 1995; Elena and Lenski 1997; Otto and Feldman 1997). For example, a reduction in fitness of 4% could be caused by one mutation reducing fitness 4% or by two mutations which each reduce fitness by 2%. This is particularly a problem for the

method that involves crossing individuals with “approximately” equal numbers of mutations (method 1). However, this problem will be reduced in crosses between individuals with very different fitnesses (method 2), because the fitness consequences of deleterious mutations are thought to be generally small (Mukai *et al.* 1972; Ohnishi 1977a,b; Crow and Simmons 1983; Keightley 1994, 1996; Keightley and Ohnishi 1998; Kibota and Lynch 1996; but see Keightley and Caballero 1997). Consequently, large fitness differences between individuals are likely to represent large numbers of deleterious mutations. Replicating crosses with different parent lines would reduce the possibilities that mutations of small effects were usually the cause of fitness decreases, but that one had unfortunately chosen to cross two individuals where the fitness differences were due to mutations with large effects. Such replication would also reduce the possibility that variation in the extent of epistasis occurs, and that one had crossed individuals whose mutations showed particularly large or small extents of epistasis.

We have used a quadratic function to represent various forms of epistasis. One of the advantages of this function is that by varying the parameters it is possible to consider a wide variety of forms of epistasis: the ratio  $\beta/\alpha$  measures the degree of epistasis (Charlesworth *et al.* 1990). Another commonly used function is truncation selection, which is an extreme case of synergistic epistasis (Kondrashov 1982). With truncation selection, individuals who have less than a certain number of deleterious mutations ( $T$ ) have a fitness of 1.0, and individuals with  $T$  or more mutations have a fitness of 0. We have not presented results with a truncation function because its extreme properties lead to it not being useful for the purposes of this paper: (1) all individuals have a fitness of 1.0 or 0;  $\log(0)$  cannot be calculated, and so it is not possible to work on a log fitness scale; (2) parents must have less than  $T$  deleterious mutations because they are alive and reproducing, and so the fitness of offspring is always equal to or less than their parents, and (3) skew is maximized when offspring fitness is either very low (lots of 0's and few 1.0's) or very high (lots of 1.0's and few 0's).

**Experimental design, controls and replication:** Our theoretical results suggest that only method 2 (crossing two individuals with different numbers of deleterious mutations) can be used to test for epistasis between deleterious mutations. A basic methodology for applying this to a haploid species would be as follows. Individuals with relatively low numbers of deleterious mutations would come from (base) populations maintained in conditions that had minimized the accumulation of deleterious mutations: large population size, plenty of opportunity for competition and selection, and, if possible, sexual reproduction. Individuals with relatively large numbers of deleterious mutations would come from lines which had maximized the accumulation of deleterious mutations: several generations with single individ-

ual population bottlenecks, as benign conditions as possible to minimize selection, and asexual reproduction. Mutations should be accumulated in several independent replicate lines. Mutation accumulation could be speeded up with artificial mutagens. The experimental crosses would come from crossing the mutation accumulation lines with individuals from the base populations. The control crosses would be to cross individuals from the base populations.

Control crosses are crucial because similar differences in mean log fitness, or skew, could be predicted by epistasis between beneficial alleles. Indeed, the inability to carry out a control is the biggest problem for method 3 (testing for skew in log fitness). Our proposed control for method 2 is to cross individuals with similar and low levels of deleterious mutations (a case of method 1). With regard to epistasis between beneficial alleles, the differences in mean log fitness between parents and their offspring in the control crosses should equal that in the experimental crosses. However, epistasis between deleterious mutations is predicted to lead to much greater differences in mean log fitness between parents and their offspring in the experimental crosses (Figure 3) than in the control crosses (Figure 2). Consequently, epistasis between deleterious mutations would lead to the mean difference between mean offspring log fitness and mean parent log fitness differing between the experimental and the control crosses.

Replication needs to be carried out at the parent level. This is crucial to avoid pseudoreplication (Hurlbert 1984) and subsequent false results that could occur for a number of reasons, such as extreme variation in mutation/epistasis effects (see above), or some particular gene combination that increased or decreased fitness. Several lines which have acquired deleterious mutations independently, through natural or artificial mutation, should be used. The data from each of these lines should then either provide a single data point in the analysis, or be analyzed by an appropriate nested approach (*e.g.*, Crawley 1993, p. 147).

Another issue that needs consideration is the nature of deleterious mutations present (segregating) in natural populations versus the properties of newly arising deleterious mutations. The problem here is that selection will lead to these two distributions being different. While it is the latter of these two distributions that is crucial to estimate, it is the former that may be more accessible to empirical study. This problem can be reduced if selection is minimized when creating and maintaining lines with high deleterious mutation loads. Related to this, it is also worth noting that if one's aim is to test the importance of deleterious mutations in maintaining sexual reproduction, it is crucial to work on sexual species: selection may have shaped the distribution of effects in newly arising deleterious mutations differently in sexual and asexual species.

**The results from *Chlamydomonas moewussii*:** Our results suggest that de Visser *et al.*'s (1996, 1997a) data do not

provide clear evidence for or against epistasis between deleterious mutations in *C. moewussi*. The main problem is that they used the two methods which we have found are unlikely to give clear results. In their first paper (de Visser *et al.* 1996), they crossed individuals with similar numbers of deleterious mutations and compared the mean log fitness of the two parent lines with their offspring (method 1). The main problem with this method is that the number of the deleterious mutations in the two parents need only differ by a small amount to give a difference in mean log fitness between parents and their offspring equal to zero or in the opposite direction to that predicted with exactly equal numbers of deleterious mutations. de Visser *et al.* (1996) only carried out one cross for each of their different UV irradiation treatments (*i.e.*, no replication at the parent level), and we have no idea how 'similar' the numbers of deleterious mutations in the two strains were. This is crucial because when crossing two lines with 'similar' numbers of mutations, any result can be explained by any form of epistasis. In addition, the predicted differences in mean log fitness with method 2 may be undetectably small (Figure 2).

In their second paper (de Visser *et al.* 1997a) they examined whether the distribution of log fitness amongst sexually produced offspring was symmetrical or skewed (method 3). The most important problem with this method is that it is very hard to construct control crosses where less skew would be expected due to deleterious mutations. For example, they found greater skew in the crosses between individuals that were likely to contain more deleterious mutations, and suggested that this was indicative of synergistic epistasis. However, our results suggest that, although the extent of skew initially increases with the number of mutations for very small numbers of deleterious mutations, it then decreases (Figure 4, A and B). The number of deleterious mutations in a population at equilibrium depends upon the mutation rate and the extent of any epistasis between deleterious mutations (Crow 1970; Kondrashov 1982; Charlesworth 1990). This makes it almost impossible to make a clear *a priori* prediction about how the skew should vary between different crosses. Indeed, the predicted equilibrium number of mutations in populations (Charlesworth 1990; Table 1) is generally enough to suggest that reasonable levels of synergistic epistasis would lead to crosses between individuals with greater numbers of deleterious mutations having less skew (Figure 4, A and B), the opposite direction to that found by de Visser *et al.* The other general problems for this method also stand.

There may also have been some problems with the measures of log fitness used in the two papers. The parameters of the logistic growth model, maximum growth rate ( $r$ ) and carrying capacity in batch culture ( $K$ ) were used as the measures of fitness. While these may be related to fitness, the exact relationship is crucial because the tests applied are very dependent upon the

scale of measurement. For example, theory has been developed in terms of discrete generations, where fitness ( $w$ ) is the number of offspring in one generation, and the population grows at a rate  $wt = e^{\log(w)t}$ . However, with overlapping generations, the maximum growth rate is  $e^r$ . This suggests that  $r$  plays the role of  $\log(w)$ , and not  $\log(r)$ . In contrast,  $K$  may be proportional to fitness when there is weak density dependent selection (Charlesworth 1980). In addition, it was necessary to remove several data points from the analyses because growth was continuing or the logistic model did not fit. This may have biased the results: skew is particularly sensitive to outliers.

**Applying the methodology to diploids:** de Visser *et al.* (1996, 1997a) initially suggested that these methods should be applied to haploids, where the problems of homozygosity and dominance are avoided. Although easiest to apply in haploids, we believe that this methodology can also be applied to diploids. One way in which this would be possible would be if all deleterious mutations were heterozygous. Individuals with all deleterious mutations in the heterozygote state could be achieved by crossing two individuals with high mutation loads that carried different deleterious mutations. This would be likely if the individuals came from completely different populations, and if they had acquired additional deleterious mutations independently (through natural or artificial mutation). These individuals could then be used as the parents with high mutational loads in method 2.

Another way of applying method 2 is possible with facultatively sexual diploid species. Consider two individuals, A and B, which would have some heterozygote and some homozygote deleterious mutations. These individuals should then be maintained asexually (clonal lineages), and additional deleterious mutations acquired independently (through natural or artificial mutation). These new mutations should be in the heterozygote state, and we shall refer to the new individuals with additional mutations as MA and MB. Two types of crosses should then be carried out: (1) the original individuals with each other ( $A \times B$ ), and (2) each mutated individual with the opposite non-mutated ( $MA \times B$ ;  $MB \times A$ ). The difference between mean offspring and mean parent log fitness in the first ( $A \times B$ ) cross would be due to epistasis between deleterious mutations, dominance effects, and epistasis between beneficial alleles (*i.e.*, all forms of non-additive genetic interactions). The difference between mean offspring and mean parent log fitness in the second group of crosses ( $MA \times B$  and  $MB \times A$ ) would be the sum of the difference in the first cross ( $A \times B$ ) and any epistasis due to the new deleterious mutations. So any difference between the first and second crosses would indicate epistasis. Both this and the previous method should be replicated with different parents, and independent acquisition of additional mutations.

de Visser and Hoekstra (1998) have recently applied method 3 to diploid species. They tested for skew in data from the literature on fitness-related traits in sev-

eral diploid species. They found that fitness-related traits in plants showed almost exclusively negative skewness, while those in fungal species did not. In addition, they argued that dominance and error variance were unlikely to be responsible for the skew observed with plant species. However, as they (de Visser and Hoekstra 1998) pointed out, their data cannot distinguish between synergistic epistasis between deleterious mutations and antagonistic epistasis between beneficial alleles. This result emphasises how control crosses cannot be carried out with method 3 that would allow epistasis between deleterious mutations to be unambiguously demonstrated. In addition, the observed skews were very large (up to  $-36.7$ ), and cannot be predicted by synergistic epistasis between deleterious mutations.

To conclude, despite their considerable importance, empirical data testing for epistasis between deleterious mutations are severely lacking. We have shown that one of the three possible methods proposed by de Visser *et al.* (1996, 1997a) should provide a relatively simple way to collect such data in both haploid and diploid species. Experiments using these methods, therefore, have the opportunity to provide valuable data for several areas in population genetics and evolutionary biology. Other possible methods include testing for a non-linear decline of log fitness in mutation accumulation experiments (Mukai 1969), and examining the fitness consequences of a known number of marker mutation or transposable element insertions (de Visser *et al.* 1997b; Elena and Lenski 1997). A general point that must be considered for all of these methods is that fitness must be measured under conditions as realistic as possible (Dudash 1990; Kondrashov and Houle 1994; West *et al.* 1996).

We thank Brian Charlesworth, Andrew Clark, Laurence Hurst, Peter Keightley, Alexey Kondrashov, Curt Lively, Margaret Mackinnon, Katrina Lythgoe, Sally Otto, Andrew Read and Arjan de Visser for useful discussion and comments on the manuscript. This work was supported by the Biotechnology and Biological Sciences Research Council.

#### LITERATURE CITED

- Barton, N. H., 1995 A general model for the evolution of recombination. *Genet. Res.* **65**: 123–144.
- Burt, A., 1995 The evolution of fitness. *Evolution* **49**: 1–8.
- Charlesworth, B., 1980 *Evolution in Age-Structured Populations*. Cambridge University Press, Cambridge, London.
- Charlesworth, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**: 199–221.
- Charlesworth, B. 1998 The effect of synergistic epistasis on the inbreeding load. *Genet. Res.* (in press).
- Charlesworth, B., and N. H. Barton, 1996 Recombination load associated with selection for increased recombination. *Genet. Res.* **77**: 27–41.
- Charlesworth, B., D. Charlesworth and T. M. Morgan, 1990 Genetic loads and estimates of mutation rates in highly inbred plant populations. *Nature* **347**: 380–382.
- Crawley, M., 1993 *GLIM for Ecologists*. Blackwell Scientific, Oxford.
- Crow, J. F., 1970 Genetic loads and the cost of natural selection, pp. 128–177 in *Mathematical Topics in Population Genetics*, edited by K. Kojima. Springer-Verlag, Berlin.
- Crow, J. F., and M. J. Simmons, 1983 The mutation load in *Drosophila*, pp. 1–35 in *The Genetics and Biology of Drosophila*, edited by M. Ashburner, H. L. Carlson and J. N. Thompson. Academic Press, London.
- de Visser, J. A. G. M., and R. F. Hoekstra, 1998 Synergistic epistasis between loci affecting fitness: evidence in plants and fungi. *Genet. Res.* (in press).
- de Visser, J. A. G. M., R. F. Hoekstra and H. van den Ende, 1996 The effect of sex and deleterious mutations on fitness in *Chlamydomonas*. *Proc. R. Soc. Lond. Ser. B* **263**: 193–200.
- de Visser, J. A. G. M., R. F. Hoekstra and H. van den Ende, 1997a An experimental test for synergistic epistasis and its application in *Chlamydomonas*. *Genetics* **145**: 815–819.
- de Visser, J. A. G. M., R. F. Hoekstra and H. van den Ende, 1997b Test of interaction between genetic markers that affect fitness in *Aspergillus Niger*. *Evolution* **51**: 1499–1505.
- Dudash, M. R., 1990 Relative fitness of selfed and outcrossed progeny in a self-compatible, protandrous species, *Sabatia Angularis* L. (Gentianaceae): a comparison in three environments. *Evolution* **44**: 1129–1139.
- Elena, S. F., and R. E. Lenski, 1997 Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**: 395–398.
- Hurlbert, S. H., 1984 Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* **54**: 187–211.
- Keightley, P. D., 1994 The distribution of mutation effects on viability in *Drosophila*. *Genetics* **138**: 1315–1322.
- Keightley, P. D., 1996 Nature of deleterious mutation load in *Drosophila*. *Genetics* **144**: 1993–1999.
- Keightley, P. D., and A. Caballero, 1997 Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **94**: 3823–3827.
- Keightley, P. D., and O. Ohnishi, 1998 EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. *Genetics* **148**: 753–766.
- Kibota, T. T., and M. Lynch, 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**: 694–696.
- Kondrashov, A. S., 1982 Selection against harmful mutations in large sexual and asexual populations. *Genet. Res.* **40**: 325–332.
- Kondrashov, A. S., 1984 Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genet. Res.* **44**: 199–218.
- Kondrashov, A. S., 1993 Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**: 372–387.
- Kondrashov, A. S., and D. Houle, 1994 Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proc. R. Soc. Lond. Ser. B* **258**: 221–227.
- Mukai, T., 1969 The genetic structure of natural populations of *D. melanogaster*. VIII. Synergistic interactions of spontaneous mutant polygenes controlling viability. *Genetics* **61**: 749–761.
- Mukai, T., S. I. Chigusa, L. E. Mettler and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**: 335–355.
- Ohnishi, O., 1977a Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effect of polygenic mutations. *Genetics* **87**: 529–545.
- Ohnishi, O., 1977b Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. III. Heterozygous effect of polygenic mutations. *Genetics* **87**: 547–556.
- Otto, S. P., and M. W. Feldman, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* **51**: 134–147.
- Sokal, R. R., and F. J. Rohlf, 1981 *Biometry*. W. H. Freeman, New York.
- West, S. A., K. E. Flanagan and H. C. J. Godfray, 1996 The relationship between parasitoid size and fitness in the field, a study of *Achrysocharoides zwoelferi* (Hymenoptera, Eulophidae). *J. Anim. Ecol.* **65**: 631–639.
- Whitlock, M., P. Phillips, F. B. G. Moore and S. J. Tonsor, 1995 Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**: 601–629.