The Population Genetics of Synthetic Lethals

Patrick C. Phillips and Norman A. Johnson

Department of Biology, University of Texas, Arlington, Texas 76019-0498

Manuscript received October 1, 1997
Accepted for publication May 18, 1998

ABSTRACT

Synthetic lethals are variants at different loci that have little or no effect on viability singly but cause lethality in combination. The importance of synthetic lethals and, more generally, of synthetic deleterious loci (SDL) has been controversial. Here, we derive the expected frequencies for SDL under a mutation-selection balance for the complete haploid model and selected cases of the diploid model. We have also obtained simple approximations that demonstrate good fit to exact solutions based on numerical iterations. In the haploid case, equilibrium frequencies of carrier haplotypes (individuals with only a single mutation) are comparable to analogous single-locus results, after allowing for the effects of linkage. Frequencies in the diploid case, however, are much higher and more comparable to the square root of the single-locus results. In particular, when selection operates only on the double-mutant homozygote and linkage is not too tight, the expected frequency of the carriers is approximately the quartic root of the ratio between the mutation rate and the selection coefficient of the synthetics. For a reasonably wide set of models, the frequencies of carriers can be on the order of a few percent. The equilibrium frequencies of these deleterious alleles can be relatively high because, with SDL, both dominance and epistasis act to shield carriers from exposure to selection. We also discuss the possible role of SDL in maintaining genetic variation and in hybrid breakdown.

DELETERIOUS mutations play a fundamental role in understanding the evolution of sex (Kondrashov 1988, 1993), aging (Charlesworth 1993; Partridge and Barton 1993), inbreeding depression (Charlesworth and Charlesworth 1987), and the level of phenotypic (Kondrashov and Turelli 1992) and molecular (Charlesworth *et al.* 1993; Charlesworth 1996) variation present in natural populations. Two factors with strong influence on the frequency of deleterious mutations within a population are the strength of selection against the mutation and how recessive the mutant allele is relative to the wild-type allele in heterozygotes (Crow and Kimura 1970). The more recessive the mutant, the higher its frequency under a mutation-selection balance because the mutant tends to be "hidden" from selection while heterozygous.

An alternative means by which deleterious mutants can be protected from selection is if they only display a deleterious phenotype when in combination with another deleterious allele at another locus (Fisher 1935). In its strongest form, this type of epistasis can lead to lethality of particular allelic combinations that will only be revealed via crossing and recombination of genotypes that separately harbor one of the deleterious alleles (Dobzhansky 1946). These so-called synthetic lethals were the subject of a fair amount of study a number of decades ago (reviewed in Thompson 1986b). In particular, Dobzhansky and many collaborators during the mid-1950's conducted a large scale search for synthetics in five species of Drosophila (see, for instance, Spassky *et al.* 1958; Lewontin *et al.* 1981). Since then, the study of synthetic lethals has waned substantially, in part because of concerns about the interpretation of the results (Thompson 1986b) and in part because the study of synthetics is fairly difficult. Here, we present a population genetic model of the equilibrium frequency of synthetic mutations that shows that deleterious mutations of this type may be more important than has been previously thought.

The synthetic phenotype need not be lethal to be important, and we will therefore focus on the more general class of interactions that can be called synthetic deleterious loci (SDL), or loci containing alleles that are deleterious only when found together. Gene interactions of this class actually arise fairly naturally in studies of a variety of evolutionary phenomena. For example, in gene duplications it is often assumed that the redundancy provided by the duplication event allows relaxed selection against the duplicate loci as long as they are paired with a good copy of the allele at the other locus (Fisher 1935; Ohno 1970; Ohta 1980, 1988, 1994). Selection would only operate against the gene complex when mutants are found at both loci. More generally, in any redundant system selection on the redundant pathways can be eased when mutations occur along only one branch of the pathway (Nowak *et al.* 1997).

Similar models have also been used in studies of the

Corresponding author: Patrick Phillips, Department of Biology, Box 19498, 501 S. Nedderman, University of Texas, Arlington, TX 76019-0498. E-mail: pphillips@uta.edu

evolution of sex and recombination, in which the focus has been on "reinforcing epistasis" (Kondrashov 1984, 1988). Here the fitness is dependent on the total number of mutations an individual possesses (Charlesworth 1990; Kondrashov 1995a,b). The truncation-selection version of this model is very similar to the idea of synthetic lethals, since fitness depends on the combined effects of these mutations rather than their individual contributions to fitness. For example, an individual with three mutations might have substantially lower fitness than an individual with two mutations, regardless of where those mutations occur (Milkman 1978).

The frequency of SDL alleles in natural populations depends on the mutation rate at both loci, the strength of selection against the synthetic genotype as well as the genotypes carrying the mutations, and the rate of recombination between the interacting loci. Here, we provide a theoretical framework for predicting the factors influencing the frequency of SDL in natural populations and discuss the types of experiments necessary to assess these frequencies. Although they will probably be very difficult to detect, SDL could potentially play an important role in several evolutionary processes.

MODEL

Haploid model: We will consider a random mating population of effectively infinite size that undergoes a life cycle of discrete generations, each generation consisting of separate mutation, recombination, and selection steps. A genetic system with two loci, each with two alleles, will be assumed. This yields four haplotypes: A_1B_1, the predominant (and most fit) haplotype in the population; A_1B_2 and A_2B_1, the two haplotypes that are carriers for the synthetic lethal or deleterious alleles; and A_2B_2, the synthetic deleterious haplotype. The frequencies of these haplotypes in the population will be represented respectively by x_1, x_2, x_3, and x_4. We will be primarily concerned with a mutation-selection balance between mutations toward the A_1 and B_1 alleles and selection against the A_2B_2 haplotype. Because the A_2 and B_2 alleles are expected to be rare, back-mutation will be ignored. Assuming that the mutation rate from A_1 to A_2 is given by μ and from B_1 to B_2 by ν and ignoring double mutation, then the change in allele frequency due to mutation is given by

$$x_1' = (1 - \mu - \nu)x_1, \tag{1a}$$

$$x_2' = (1 - \mu)x_2 + \nu x_1, \tag{1b}$$

$$x_3' = (1 - \nu)x_3 + \mu x_1, \tag{1c}$$

$$x_4' = x_4 + \mu x_2 + \nu x_3. \tag{1d}$$

The change in haplotype frequencies caused by recombination is then calculated by

$$x_i'' = x_i' + \delta r D, \tag{2}$$

TABLE 1

Haploid fitness model

Gamete	A_1B_1	A_1B_2	A_2B_1	A_2B_2
Frequency	x_1	x_2	x_3	x_4
Fitness	1	$1 - h_1s$	$1 - h_2s$	$1 - s$

where x_i is the frequency of the ith haplotype, r is the recombination fraction, $D = x_1'x_4' - x_2'x_3'$ is the linkage disequilibrium parameter, and δ is -1 when $i = 1$ or 4, and 1 when $i = 2$ or 3.

Selection is assumed to act primarily against the A_2B_2 synthetic haplotype, but it can also act against the A_1B_2 and A_2B_1 carrier (or repulsion) haplotypes. The fitnesses of each haplotype are given in Table 1. Using these values, the change in frequency caused by selection is given by

$$x_i''' = x_i''w_i/\bar{w}, \tag{3}$$

where w_i is the fitness of haplotype i, and $\bar{w} = 1 - s(h_1x_2 - h_2x_3 - x_4)$ is the mean fitness in the population.

Some of the equilibrium properties of diploid models similar to that given above have been described (Karlin and McGregor 1971; Christiansen and Frydenberg 1977; Allendorf 1979; Bengtsson and Christiansen 1983). We do not wish to rederive these results, but will instead focus on some novel approximations of the mutation-selection balance equilibrium achieved in the symmetric case, where $\nu = \mu$ and $h = h_1 = h_2$ (Table 1). Here, we will use the parameter, h, to describe the degree of epistatic shielding of one locus by another. In the haploid model, h functions similarly to an interlocus dominance parameter in a fashion analogous to the intralocus dominance parameter that is usually used. The important distinction here is that recombination can break up the coupling between alleles. Thus, $h = 1$ means that A_1 and B_1 are completely epistatic (in the classic sense; Wade 1992; Phillips 1998) to A_2 and B_2, $h = 1/2$ means that there is no epistasis on an additive scale, and $h = 0$ means that A_2 and B_2 are completely epistatic to A_1 and B_1.

Complete epistasis ($h = 0$): With complete epistasis and a completely symmetrical model the equilibrium conditions for Equations 1–3 can be solved analytically. First, for synthetic lethality ($s = 1$), the equilibrium frequency of the carrier haplotypes ($x = x_2 = x_3$) is given by

$$x = \{\mu(r\mu - 1) + \sqrt{\mu[r + \mu(1 - 2r - r^2) - \mu^2r(1 - 2r)]}\}/r(1 - \mu)^2. \tag{4}$$

A very good approximation to this equation can be obtained by ignoring all terms smaller than $\sqrt{\mu}$ yielding

$$x \approx \sqrt{\mu/r}, \tag{5}$$

given $r \gg \mu$. This bears a striking similarity to the standard single-locus approximation for mutation-selection balance with complete dominance, $q \approx \sqrt{\mu/s}$ (Crow and Kimura 1970, p. 259), except here the frequency of the carrier haplotype is determined by selection against the synthetic lethal and maintained at higher levels by tight linkage. An equivalent result has been found by Nowak *et al.* (1997), who report that the expected frequency of the A_1B_1 haplotype in this case would be $1 - 2\sqrt{\mu/r}$.

When the synthetic effects are merely deleterious and not lethal ($s < 1$), we can obtain a solution for the equilibrium by assuming weak selection ($\bar{w} \approx 1$). Even with this assumption, however, the solution involves an equation with over 80 terms (not shown). Nevertheless, a good approximation to the equilibrium can be found by again neglecting terms smaller than $\sqrt{\mu}$, which gives

$$x \approx \sqrt{\frac{\mu(r + s - rs)}{rs}}. \tag{6}$$

Here the frequency of the mutants is determined by the interaction of selection and recombination. The approximation can be improved, especially when r is small, by also retaining most terms on the order of μ, yielding an equilibrium value as $x^* = x(1 - x)$, where x is given by Equation 6. This weak selection approximation reduces to the strong selection approximation (5) when $s = 1$. Thus, Equation 6 represents a general approximation for complete epistasis. The likely reason why the weak selection approximation works well even with complete synthetic lethality is because with complete epistasis, selection acts only on the double mutant, whose frequency will tend to be so low that the assumption of $\bar{w} \approx 1$ is a good approximation.

The equilibrium frequency of the double mutant can also be found, but again produces a solution with more than 60 terms. An approximation is provided by

$$x_4 \approx \mu(1 - s)/s, \tag{7}$$

which is fairly close to the square of the single-locus mutation-selection balance approximation given above, especially when s is small.

Partial epistasis ($h > 0$): If selection is now allowed to operate on the carrier haplotypes, then the frequency of the synthetic haplotype should be so low that it can effectively be treated as zero. Assuming this, solving Equations 1–3 for the frequency of the carrier haplotypes (x) yields

$$x = \frac{\sqrt{4r\mu(1 - \mu)^2(1 - r\mu)(1 - hs)^2 + y^2} - y}{2r(1 - \mu)^2(1 - hs)}, \tag{8}$$

where $y = \mu(3 - 2r\mu)(1 - hs) + hs$. This can be approximated by eliminating terms smaller than μ and then using the first two terms of the Taylor series expansion of the square root function around the point $(hs)^2$ to give

TABLE 2

Diploid fitness model

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	1	1	1
B_1B_2	1	$1 - ks$	$1 - h_2s$
B_2B_2	1	$1 - h_1s$	$1 - s$

$$x \approx \frac{\mu(1 - hs)}{hs}. \tag{9}$$

This is very similar to the single locus mutation-selection balance with partial dominance, $q \approx \mu/hs$ (Crow and Kimura 1970, p. 260), especially when hs is small. Note that x does not depend on the recombination rate unless h is very close to zero, as is shown numerically below.

Diploid model: Diploid models of synthetic lethality have been extensively analyzed in the mutation-selection balance context from a variety of viewpoints (Karlin and McGregor 1971; Christiansen and Frydenberg 1977; Allendorf 1979; Bengtsson and Christiansen 1983). These studies have generally focused on the existence and stability of the equilibria found without investigating the actual frequency of the gametes attained at these equilibria. We will first again derive several approximations for equilibrium frequencies found with the assumption of symmetrical fitnesses and mutation rates, and then numerically explore how these frequencies change when the assumption of symmetry is removed.

Following Bengtsson and Christiansen (1983), the changes in gamete frequency generated by mutation and recombination are as given above in Equations 1 and 2, except that mutation occurs after selection and recombination. (The order of these steps has virtually no effect on the calculated equilibria, but does impact the ease with which the approximations can be derived.) Using the genotype fitnesses given in Table 2, the change due to selection is given by Equation 3 with the addition that w_i is the marginal fitness of a gamete, $w_i = \Sigma_j x_j w_j$, and $\bar{w} = 1 - s[2k(x_1x_4 + x_2x_3) + 2h_1x_2x_4 + 2h_2x_3x_4 + x_4^2]$.

Complete epistasis ($k = 0$, $h = 0$): The major difference between the haploid and diploid models is that the alleles comprising the synthetic genotype ($A_2A_2B_2B_2$) in the diploid case are shielded from selection by both dominance and epistasis. Because of this, the frequency of the double mutant gamete can be quite high even when the synthetic genotype is lethal. The high frequency of this gamete makes it much more difficult to derive approximations directly from the recursion equations, and we have thus far been unable to find a general solution for the diploid case. However, if the equilibrium frequency of the double mutant gamete is taken to be $x_4 \approx \sqrt{\mu/s}$ (Christiansen and Frydenberg

1977), and assuming that selection is weak enough so that $\bar{w} \approx 1$, the equilibrium for the carrier gamete is found to be

$$x = \{\sqrt{y^2 + 4r(1-2\mu)(1-\sqrt{\mu/s})[r\sqrt{\mu/s}(1-2\mu)+\mu]} - y\}/2r(1-2\mu), \quad (10)$$

where $y = 3\mu + 2r\sqrt{\mu/s}(1-2\mu)$. Expanding out the terms of (10) and collecting terms with the largest roots of μ lead to the following approximation,

$$x \approx \sqrt{\mu/r + \sqrt{\mu/s}}. \quad (11)$$

When linkage is tight ($r \to 0$), then (11) can be further approximated by $x \approx \sqrt{\mu/r}$, the haploid lethal approximation given by (5) above. Similarly, when linkage is not too tight and selection is weak, (11) can be approximated by $x \approx \sqrt[4]{\mu/s}$, which is the square root of the normal single locus approximation for the case of complete dominance. Because the ratio of the mutation rate to the selection coefficient is much less than one, the quartic root of this will be larger than the square root, and the two-locus equilibrium will be much higher than the corresponding single-locus equilibrium. Unless linkage is fairly tight ($r < O[\sqrt{\mu s}]$) the selection term should dominate this equation. The approximation given by (11) provides the easiest interpretation, but numerically, including a few more terms from (10) greatly improves the accuracy of the approximation. Algebraically this can be achieved by refining (11) by taking $x^* = x(1-x)$, as was done above.

Additive double heterozygotes ($k = \frac{1}{2}h$): Moving away from complete epistasis introduces three more parameters into the fitness model (Table 2). The degree of dominance of the marginal heterozygotes, h_1 and h_2, describes how dominant a mutant allele at one locus is when found on a genetic background that is homozygous at the other mutant allele (*i.e.*, $A_1A_2B_2B_2$ and $A_2A_2B_1B_2$). Similarly, the degree of dominance for the double heterozygote ($A_1A_2B_1B_2$) is described by the parameter, k (Table 2). We have been unable to find a general solution to the recurrence equations when these parameters are allowed to take general values. However, in the symmetrical case in which $h = h_1 = h_2$, if the double heterozygotes are constrained to have behaved additively with respect to the marginal heterozygotes ($k = \frac{1}{2}h$), we find that the equilibrium frequency of the synthetic genotype ($A_2A_2B_2B_2$) is approximately $x_4 \approx \mu/hs$, the single-locus mutation-selection balance equilibrium (see below). Assuming this frequency, then the weak selection approximation for the equilibrium frequency of the carrier gamete is

$$x = \{hs\mu[r(1-2\mu) - 2(2-\mu)] - 2r\mu(1-\mu) + \sqrt{\mu^2[y + hs(3-\mu)]^2 + y\mu(y+hs)(hs-\mu)}\}/hsy, \quad (12)$$

where $y = hs(1-\mu) + r(1-2\mu)(2-hs)$. Equation 12 can be shown to be fairly well approximated by

$$x \approx \sqrt{\mu/hs}, \quad (13)$$

which again is the square root of the standard single-locus result.

Deleterious double heterozygotes ($k > h$): If the fitness of the double heterozygotes is lower than that of the marginal heterozygotes ($k > h$), then almost all selection on the mutations will take place on the double heterozygotes. The mutants will be found primarily on the carrier gametes (A_1B_2 and A_2B_1), and therefore double heterozygotes should be more common than either marginal heterozygote. Since selection against the mutants is very effective in this scenario, the frequency of the synthetic gamete should be very low. This will be especially true if linkage is tight. If both r and x_4 are assumed to be zero and $\bar{w} \approx 1$, then the recursion equations can be solved to yield

$$x \approx \frac{\sqrt{9\mu^2 + 4ks\mu(1-\mu)} - 3\mu}{2ks(1-\mu)}. \quad (14)$$

Ignoring factors smaller than $\sqrt{\mu}$, (14) can be approximated by

$$x \approx \sqrt{\mu/ks}, \quad (15)$$

which is the same result as (13) with k substituted for h. Therefore, the equilibrium frequency of the carriers will be largely determined by which of the heterozygotes provides the smallest degree of dominance for fitness, because this class of heterozygotes will be the primary focus of selection.

Numerical calculations: The accuracy of the approximations given above was evaluated by solving Equations 1–3 via iteration from a starting point of $x_1 = 1$, with the exception that gamete frequency changes from double mutations were included in the numerical calculations. The iterations were continued until the frequency of x_2 had changed less than $10^{-3}\mu$ for at least 10,000 generations.

NUMERICAL RESULTS

Haploid model: Under a model with symmetrical fitnesses and mutation rates, the carrier haplotypes can be maintained at significant frequencies, especially when epistasis is close to complete (Figure 1A). The approximations work remarkably well (Figure 1A), and so the nature of the equilibrium can be interpreted directly from them. With complete epistasis ($h = 0$), the frequency of the carrier haplotypes is proportional to the square root of the mutation rate and roughly inversely proportional to the square root of the product of the recombination rate and the strength of selection (6). However, as selection becomes strong ($s > 0.1$), further increases in the strength of selection have fairly minor

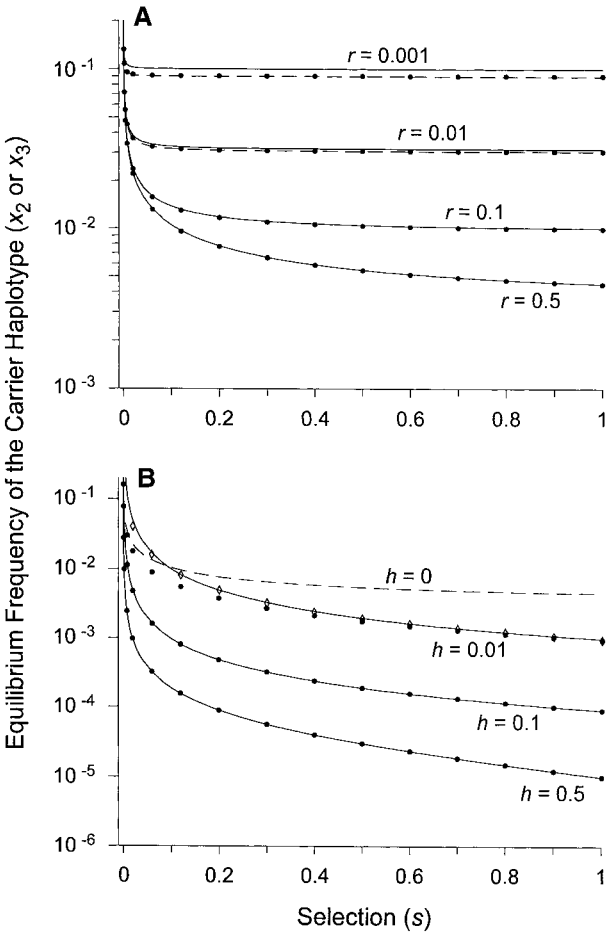

Figure 1.—The equilibrium frequency of the carrier haplotypes under different strengths of selection for haploids. Points are exact calculations based on iterating Equations 1–3 until an equilibrium was achieved. The mutation rate was set to $\mu = 10^{-5}$. (A) The effect of differing recombination rates (r) with completely synthetic epistasis ($h = 0$). The solid lines are the approximation $x = \sqrt{\mu(r + s - rs)/rs}$ and the dashed lines give the approximation $x^* = x(1 - x)$. The first approximation works extremely well unless linkage is very tight, at which time the two approximations form upper and lower bounds on the equilibrium frequency. (B) The effect of the completeness of epistasis (h). The solid lines are the approximation $x = \mu(1 - hs)/hs$. The dots give the results for free recombination ($r = 0.5$). The recombination rate has little effect on the equilibrium frequency until h becomes very small, as is shown by the diamonds for the case of $h = 0.01$ and $r = 0.001$. For weak selection and nearly complete epistasis, the approximation works better when linkage is tight (diamonds) than when there is free recombination (dots), in which case the approximation from (A) works better (dashed line).

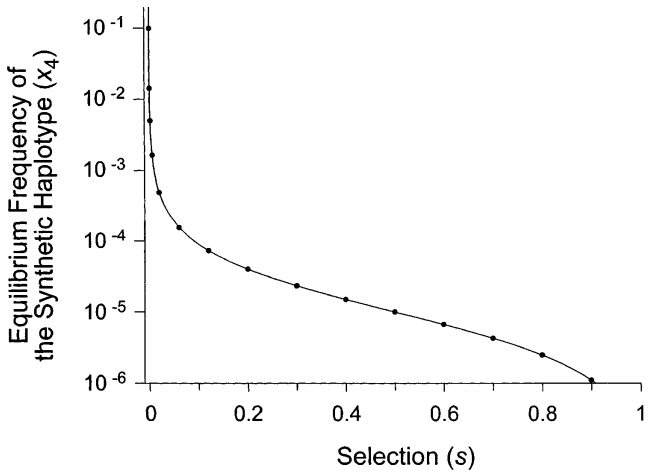

Figure 2.—Equilibrium frequency of the synthetic haplotype under different strengths of selection with complete epistasis ($h = 0$, $\mu = 10^{-5}$). Points are the exact calculations based on Equations 1–3 and the line is the approximation $x_4 = \mu(1 - s)/s$. Recombination rate has a negligible effect on the equilibrium frequency, which is undetectable on this scale.

effects on the carrier equilibrium (Figure 1). This is because, although selection continues to drive the synthetic haplotype to very low frequencies (Figure 2), the negative linkage disequilibrium generated by this selection maintains the carrier haplotypes at higher frequencies because they are not directly exposed to selection. This effect is particularly clear as linkage becomes tight, in which case relatively high frequencies of the carrier haplotypes can be maintained even in the face of strong selection (Figure 1A).

Moving away from complete epistasis ($h > 0$) leads to a precipitous decline in the frequency of the carrier haplotypes (Figure 1B). Here linkage has little effect because selection operates primarily on the carriers themselves rather than on the synthetic haplotypes. The results in this case are therefore much more similar to that predicted from an analogous single-locus mutation selection balance, in which the frequency of the carrier haplotype is inversely proportional to the degree of masking of one locus by another (9).

Diploid model: Carrier gamete frequencies are several orders of magnitude higher in the diploid case than in the haploid case (Figure 3). When mutations are completely shielded from selection by dominance and epistasis except when found as double homozygotes ($h = 0$ and $k = 0$), frequencies of the carrier gametes can easily exceed 1% (*i.e.*, proportional to the quartic root of the mutation rate; Equation 11), especially when linkage is tight (Figure 3A). This is true even when selection against the mutants is very strong. In effect, the synthetic gametes also become carriers in this case because they are shielded by dominance when heterozygous. The frequencies of the synthetic gametes are therefore also much higher when compared to the haploid situation (Figure 4A), although tight linkage can substantially reduce these frequencies (Figure 4B). Note that although the gametic frequencies are relatively high for loci under strong selection, the genotype frequencies will roughly be the square of these, so that the frequency of individuals showing the synthetic phenotype will be very low even as the frequency of the carriers is orders of magnitude higher.

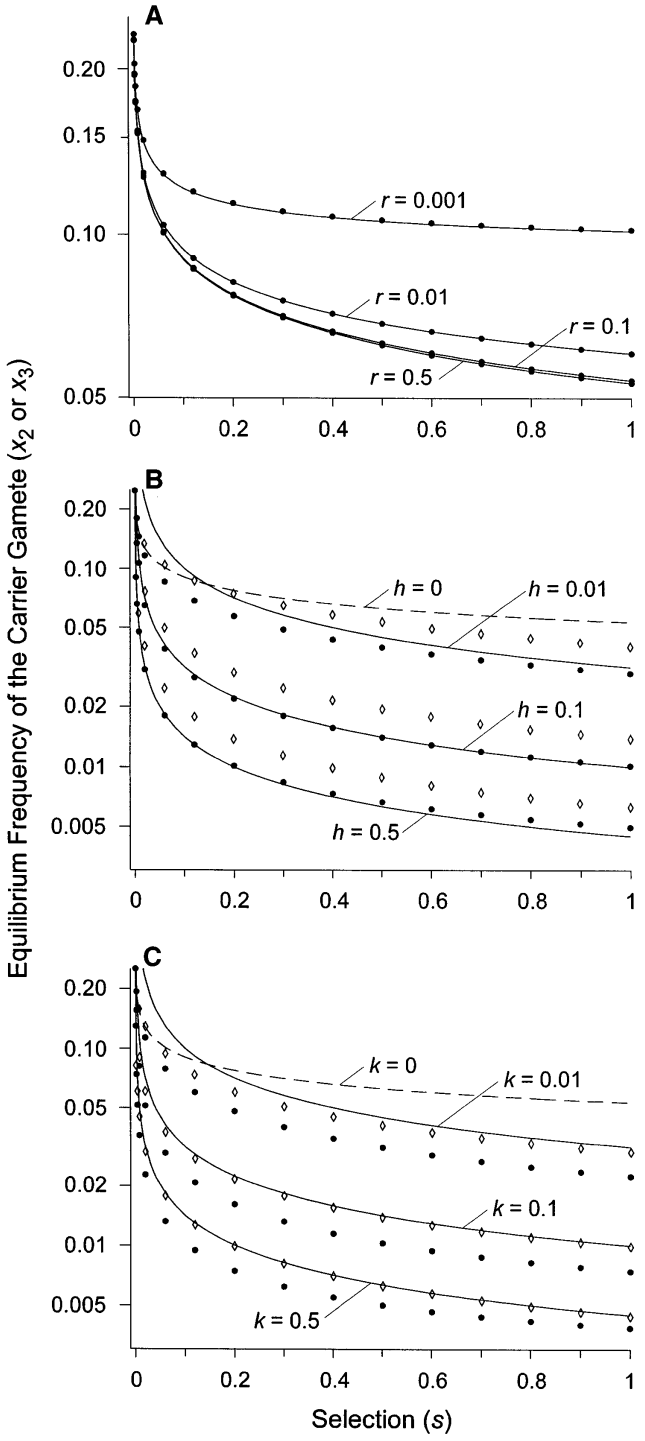

Figure 3.—The equilibrium frequency of the carrier gametes (A_1B_2 or A_2B_1) under different strengths of selection for diploids. Points are exact calculations based on iterating Equations 1–3 until an equilibrium was achieved. The mutation rate was set to $\mu = 10^{-5}$. (A) The effect of differing recombination rates (r) with completely synthetic epistasis and dominance ($h = 0$, $k = 0$). The solid lines are the approximation $x^* = x(1 - x)$, where $x = \sqrt{\mu/r} + \sqrt{\mu/s}$. (B) Partial epistasis with additive double heterozygotes ($k = h/2$). The solid lines are the approximation $x = \sqrt{\mu/hs}$. The approximation works well when there is free recombination ($r = 0.5$, solid points), but underestimates the equilibrium frequency when linkage is tighter ($r = 0.01$, diamonds), except when selection is weak. (C) Deleterious double heterozygotes ($k > h$). The solid lines are the approximation $x = \sqrt{\mu/ks}$. Here, the approximation works better when there is tight linkage ($r = 0.01$, diamonds) than when there is free recombination ($r = 0.5$, points). Varying h in this case has little influence on the equilibrium frequency ($h = 0$ here). In general, regardless of the form of epistasis operating on the synthetic genotypes, the equilibrium frequency of the carrier gametes is high, even under strong selection.

Relaxing the assumption of complete dominance and epistasis reduces the frequencies of the carrier gametes, but not dramatically so (Figure 3B). Even here, although any genotype containing both mutants is selected against, the mutants are shielded from the bulk of selection because they are primarily found in individuals with only one mutant, which in turn have normal fitness (Table 2). The same result holds if the double heterozygotes have lower fitness than the marginal heterozygotes (Figure 3C). Although the approximations are not nearly as precise as in the haploid case (especially with tight linkage), precision is hardly the issue, as both the approximations and the numerical results predict extremely high mutant frequencies (roughly the square root of the predicted haploid frequencies; Equations 13 and 15) considering the strength of selection these loci are under.

Unbalanced mutation rates: The influence of unbalanced mutation rates in complete epistasis has been solved by Christiansen and Frydenberg (1977), who found that populations will evolve to an equilibrium in which the locus with the higher mutation rate is fixed and the locus with the lower mutation rate is segregating according to a single-locus mutation-selection balance (the mutations could also go to fixation via drift in finite populations; Kimura and King 1979). The rate of approach to this equilibrium is exceedingly slow, however, as the population is first rapidly driven to a mutation-selection balance quasi-equilibrium and then very slowly approaches the mutation-mutation balance equilibrium (Christiansen and Frydenberg 1977). An example of this is given in Figure 5, which shows that the mutant alleles would be expected to be polymorphic for hundreds of thousands of generations as long as the population does not start right at a fixed equilibrium. Therefore, unbalanced mutation rates are unlikely to substantially alter the conclusions regarding the frequency of SDL given above in the symmetric case.

Unbalanced selection coefficients: The general solution to the mutation-balance equation with arbitrary selection has yet to be found and is likely to be very complex. Bengtsson and Christiansen (1983) provide some sufficient, but not necessary, conditions for a two-locus polymorphism under this model. With strongly asymmetric fitnesses, the population tends to move toward fixation of the carrier haplotype with the

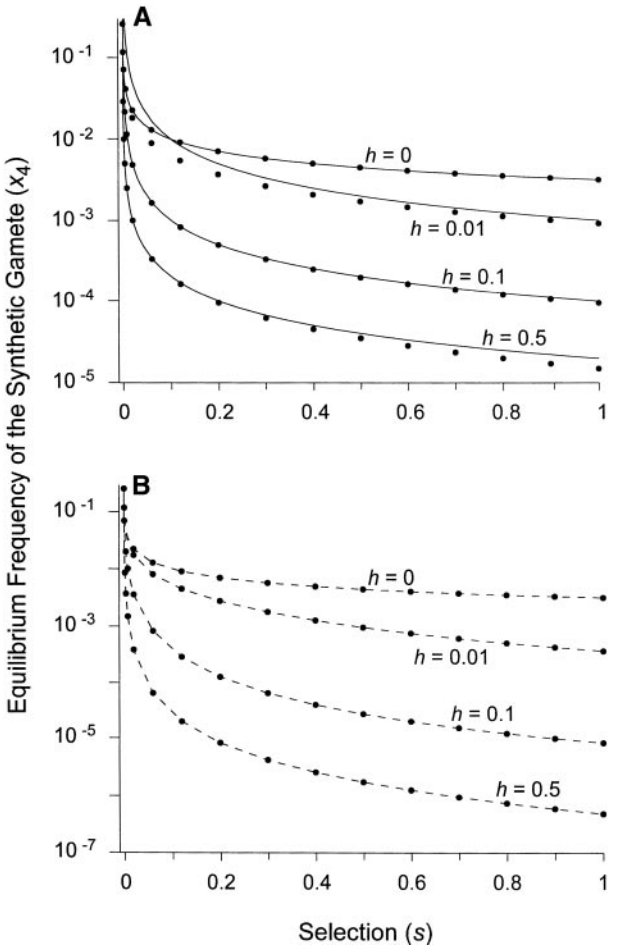

Figure 4.—Equilibrium frequency of the synthetic gamete under different strengths of selection ($\mu = 10^{-5}$). (A) Free recombination ($r = 0.5$). Points are the exact calculations and the lines are the approximations $x_4 = \sqrt{\mu/s}$ for $h = 0$ and $x_4 = \mu/hs$ for $h > 0$. (B) Tight linkage ($r = 0.01$). Points and the dashed lines connecting them are based on exact calculations. Note that linkage leads to a strong decrease in the equilibrium frequency of the synthetic gamete.

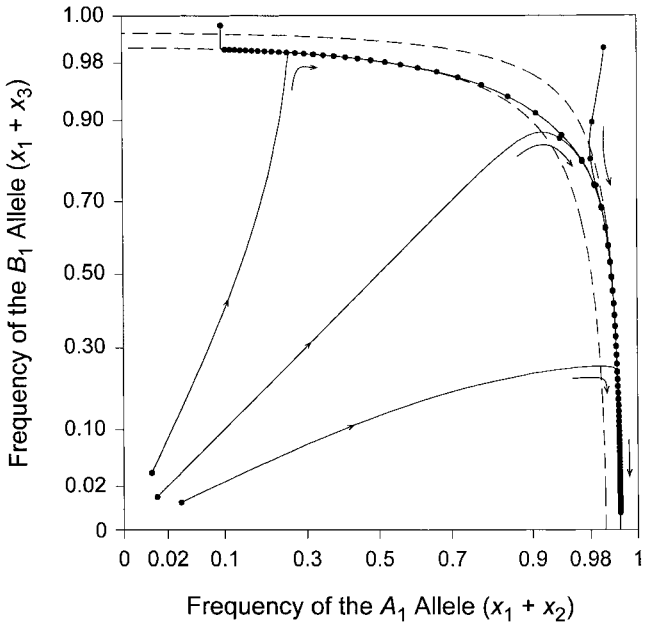

Figure 5.—Unbalanced mutation rates and the equilibrium frequency of the mutant alleles. Allele frequency trajectories for the case of complete epistasis in diploids ($h = 0$, $k = 0$) with unbalanced mutation rates ($\mu = 10^{-6}$, $\nu = 10^{-5}$), for $s = 0.1$ and $r = 0.5$. Dashed lines show the balanced-rate equilibria from Christiansen and Frydenberg (1977), for the two different mutation rates. Each solid line shows the trajectory for a population started in linkage equilibria with different initial allele frequencies. The distance between solid points represents a period of 10,000 generations. Arrows indicate the direction of evolutionary change. Each population moves from one quasi-equilibrium to the other and will eventually end up fixed for the B_2 allele and at a single locus mutation-selection balance for the A-locus. The populations are polymorphic for hundreds of thousands of generations before approaching this equilibrium, however. The frequencies are shown on the arcsine-square-root scale to accentuate the allele frequencies near one. Modeled after Christiansen and Frydenberg (1977).

higher fitness, leaving a single-locus mutation-selection balance at the other locus (Bengtsson and Christiansen 1983). The existence of a fully polymorphic equilibrium depends on a balance between all parameters, but is largely determined by the fitness of the double heterozygotes, $1 - ks$, because the majority of individuals displaying a mutant phenotype will be of this class. Loose linkage also tends to favor a polymorphic equilibrium (Bengtsson and Christiansen 1983), because in this case it is less likely that a single gamete will come to dominate the population. More work will be necessary before the nature of all equilibria is determined, but numerical calculations (not shown) suggest that the total number of possible equilibria is very large. Nevertheless, a wide variety of parameter values yield either polymorphic equilibria or the quasi-equilibria described above when the mutation rates are unbalanced.

In general, it appears that if a particular combination of parameters, h_1, h_2, and k, results in a polymorphic equilibrium, then increasing either h_1 or h_2 still yields a polymorphic equilibrium. The precise value taken by that equilibrium, however, depends to a large extent on the initial conditions. A few numerical solutions help to illustrate this point. In the most permissive situation, if $k = 0$, then a polymorphic equilibrium exists for a wide variety of h values (Figure 6A). More importantly, the equilibrium frequencies of the carrier gametes under these conditions are of the same order as those predicted for the symmetrical case described above. As selection against the double heterozygotes becomes stronger ($k > 0$), the existence of a polymorphic equilibrium depends more critically on the balance between selection on the marginal heterozygotes, but if this selection is sufficiently strong then neither of the carrier gametes will be fixed (Figure 6B). Again in this more unbalanced case, the equilibrium frequencies are comparably as high as those predicted by the symmetrical models.

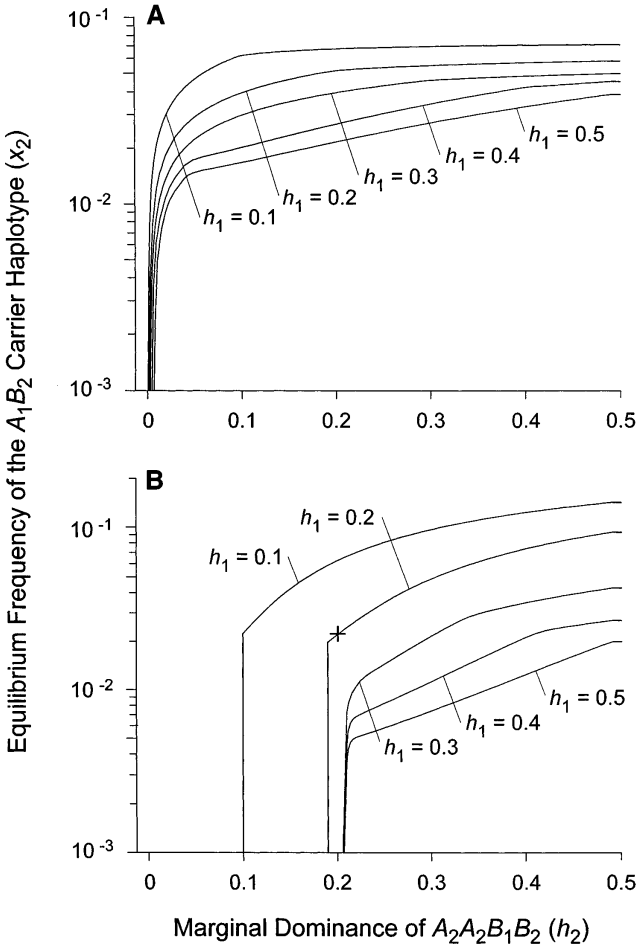

Figure 6.—Unbalanced selection coefficients and the equilibrium frequency of the carrier gametes in the diploid case with $s = 0.1$ and $r = 0.5$. Each line shows the equilibrium frequency of the A_1B_2 gamete for different values of h_1 as the value of h_2 varies from 0 to 0.5. (A) No selection against double heterozygotes ($k = 0$). (B) Some selection against double heterozygotes ($k = 0.1$). The + gives the expected frequency under the additive double heterozygote approximation, $x = \sqrt{\mu/hs}$. These equilibria are the solutions with the initial condition $x_1 = 1.0$. Other, slightly different equilibria are possible with different initial conditions (not shown). In general, as long as the polymorphic equilibrium exists, varying the balance between the marginal dominance parameters (h_1 and h_2) does not alter the qualitative result that the equilibrium frequencies are high.

DISCUSSION

Under reasonable assumptions, SDL would be expected to be found at substantial frequencies in natural populations. The predicted frequencies for haploids are similar to the analogous single-locus mutation-selection balance frequencies (Equations 6 and 9), whereas the diploid frequencies are much higher, closer to the square root of the single locus results (Equations 11, 13, and 15). The epistatic shielding from selection of one locus by another is essentially the same feature that allows deleterious recessive mutations to be maintained in populations when completely recessive. However, linked epistatic loci can be maintained at even higher relative frequencies because they are further protected by linkage, whereas recessive alleles are continually exposed to selection via segregation in the single-locus case. When dominance and epistasis are both operating, as in the diploid case, then mutant frequencies can become extremely high for loci under such strong selection, on the order of one in a thousand, or even more than one in a hundred (Figure 3).

If SDL can exist at such potentially high frequencies, why have they not received more attention? The contrast with single locus mutations is especially striking, because the theory derived here suggests that SDL frequencies can be substantially higher than the single locus mutations, yet most of population genetics theory has traditionally focused on single locus effects. There are a number of reasonable explanations why this might be the case.

First, genetics systems in which SDL occur may be extremely rare. While two or more locus interactions in general are probably ubiquitous in genetic systems (Whitlock *et al.* 1995), the particular form of interaction that leads to a synthetic phenotype may be rare or unimportant from the standpoint of fitness. Mutationally derived synthetic phenotypes have been frequently observed (*e.g.*, Davis and MacIntyre 1988; Johnson *et al.* 1988), and constitute an important means of functional analysis, particularly in yeast (Fromont-Racine *et al.* 1997). Unfortunately, the fitness effects of these mutations, particularly as single-locus heterozygotes, are usually not known. That the nature of the interaction be synthetic is important for the results presented here, as other forms of interaction would be expected to lead to much lower equilibrium frequencies. In particular, if the mutations show penetrance such that single-mutant genotypes have lower fitness, then the equilibrium frequencies would be expected to be much lower, probably close to the single-locus mutation-selection balance. If mutants show individual deleterious effects, then the fact that their interaction produces even lower fitness is somewhat inconsequential, because the mutations would be expected to be eliminated based on their own effects (Temin *et al.* 1969). Completely synthetic epistasis is likely to be as rare as complete dominance.

A more intriguing possibility is that SDL exist, but are difficult to detect. Synthetic genotypes will be very rare even if the frequencies of the carriers are very high (Figures 2 and 4). In the diploid case, even if the frequency of the synthetic gamete were as high as 10^{-3} (Figure 4), the expected frequency of the double-mutant homozygote would be closer to 10^{-6}. Thus, even if SDL exist, they would largely go unnoticed in most circumstances. Studies looking for synthetic lethals have therefore relied on generating recombination between marked chromosomes in an effort to flush out any underlying interacting loci or, more simply, have mixed

different chromosomes together. Many of these studies report recombinational synthetic lethals, but, as Thompson (1986b) has pointed out, these lethals could have arisen from mutations and not from recombination. Controls for the rate of mutation to lethality are therefore required. The only firm evidence for recombinational synthetic lethals is based on the desynthesis and resynthesis of the chromosomal segments in question—something that has been difficult to do (*e.g.*, Dobzhansky and Spassky 1960; Batten and Thoday 1969). Bichromosomal studies have been somewhat more successful (*e.g.*, Temin *et al.* 1969; Thompson 1986a), but most of these synthetic chromosomes also have individual deleterious effects, leading to the issue discussed above.

One would need to examine hundreds if not thousands of chromosomes to detect synthetics, and even more to get reasonable standard errors of the frequencies of lethals that appear upon recombination. This makes comparisons across different species or different chromosomes in the same species exceedingly difficult. In a letter to Curt Stern (25 Jan. 1955), Dobzhansky sums up the results of the large-scale search for synthetics: "The different species are interestingly different in the frequency of synthetic lethals but all of them produce such." The problem is that these interesting differences could well be due to sampling error. Highlighting this problem, Thompson (1986b) used the results of three studies with large sample sizes and adequate controls to estimate the frequency of synthetics on the second chromosome of *Drosophila melanogaster*. His estimates for the proportion of the recombinant chromosomes harboring a newly arising synthetic lethal range from 0.2 to 1.6%; and the proportion of newly arising lethals that are synthetic ranges from 26 to 78%. There is thus some evidence that loosely linked synthetics exist, although they probably do not comprise a large fraction of the standing lethals in most populations. It is possible that more tightly linked SDL, which would be extremely difficult to study, might be more common (Thompson 1986a). It is also important to note that our analysis considered only two-locus effects. None of the data collected so far can distinguish two-locus interactions from interactions involving a much larger number of loci. The more SDL involved in the synthetic pathway, the higher the equilibrium frequencies of carriers at each locus are likely to be.

From an evolutionary standpoint, SDL would likely have their largest impact on natural populations in terms of the maintenance of genetic variation. The very factor that allows the mutant alleles to reach high frequencies in carrier gametes is the same feature that limits the selective influence of these mutations within a population. The selective load on a population is low precisely because the synthetic genotypes under selection are rare even while the carrier gametes are common. In this sense, arguments regarding SDL are reminiscent of the traditional arguments regarding the maintenance of large amounts of variation under selection without an excessive burden of segregation or mutation load. As long as each selective death eliminates more than one mutation at a time (*e.g.*, the truncation selection discussed above), then the total selective load can be greatly reduced (King 1967; Milkman 1967; Sved *et al.* 1967; Crow 1992). The models of SDL presented here can therefore be viewed as specific two-locus versions of such a genetic system. Even though the frequencies of the SDL carriers are relatively high compared to deleterious mutations segregating at single loci, they are still not high enough to explain the large amounts of molecular variation often observed. Much of this variation is of course probably neutral or nearly neutral in nature. Nevertheless, SDL may have some role to play, especially if many loci are involved. SDL therefore provide a means of accounting for fairly large amounts of genetic variation in natural populations, while allowing that variation to be explicitly under selection. It is possible that this variation could be more relevant to quantitative traits than to strictly molecular variation.

Evolutionary biologists have also long recognized the potential for genetic interaction like SDL to contribute to reproductive isolation and hence speciation (Dobzhansky 1937; Mayr 1963). Suppose that a population is split in two, with little or no gene flow between the subpopulations. If carrier genotypes are segregating at SDL, then there is a possibility that the two diverging populations could be fixed for different variants at the SDL. The higher the frequencies of the carriers in the original population, the greater the likelihood of fixation of alternatives. Alternatively, duplicate genes that would behave much as SDL could undergo alternative silencing in the two populations (Werth and Windham 1991). While this process will not result in F_1 fitness reduction between the two nascent species, it could result in substantial F_2 or backcross hybrid breakdown if there were many SDL with segregating variants in the original population. A full analysis of this problem will require the addition of genetic drift to the models presented above.

The empirical evidence to date is not sufficient to allow judgment as to how common SDL are. If they are common in the sense that there is a functional basis for this type of two-locus interaction, then the theory presented here suggests that SDL should be common in the sense that they should exist at relatively high frequencies within natural populations and may therefore impact both the short- and long-term evolution of those populations.

We thank James Crow and Michael Wade for their tutelage and influence on our thinking. We also thank the American Philosophical Society for permission to quote from the correspondence of Th. Dobzhansky, and we thank their staff (especially Scott DeHavilin) for assistance. We thank A. Clark, J. Crow, M. Whitlock and two

anonymous reviewers for comments on the manuscript, and M. Lynch and J. Willis for helpful discussions. This work was supported by grants from the National Science Foundation (DBI-9722921) and the National Institutes of Health (GM54185) to P.C.P. and from the University of Texas at Arlington Research Enhancement Program to N.A.J.

LITERATURE CITED

Allendorf, F. W., 1979 Rapid loss of duplicate gene expression by natural selection. Heredity **43:** 247–258.

Batten, J. L., and J. M. Thoday, 1969 Identifying recombinational lethals in *Drosophila melanogaster*. Heredity **24:** 445–455.

Bengtsson, B. O., and F. B. Christiansen, 1983 A two-locus mutation-selection model and some of its evolutionary implications. Theor. Pop. Biol. **24:** 59–77.

Charlesworth, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. **55:** 199–221.

Charlesworth, B., 1993 Evolutionary mechanisms of senescence. Genetica **91:** 11–19.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in Drosophila melanogaster. Genet. Res. **68:** 131–149.

Charlesworth, D., and B. Charlesworth, 1987 Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. **18:** 237–268.

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

Christiansen, F. B., and O. Frydenberg, 1977 Selection-mutation balance for two nonallelic recessives producing an inferior double homozygote. Am. J. Hum. Genet. **29:** 195–207.

Crow, J. F., 1992 Twenty-five years ago in Genetics: identical triplets. Genetics **130:** 395–398.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Burgess, Minneapolis.

Davis, M. B., and R. J. MacIntyre, 1988 A genetic analysis of the α-glycerophosphate oxidase locus in *Drosophila melanogaster*. Genetics **120:** 755–766.

Dobzhansky, T., 1937 *Genetics and the Origin of Species*, Ed. 3. Columbia University Press, New York.

Dobzhansky, T., 1946 Genetics of natural populations. XIII. Recombination and variability in populations of *Drosophila pseudoobscura*. Genetics **31:** 269–290.

Dobzhansky, T., and B. Spassky, 1960 Release of genetic variability through recombination. V. Breakup of synthetic lethals by crossing over in *Drosophila pseudoobscura*. Zool. Jahr. Syst. Okol. Geog. Tiere **88:** 57–66.

Fisher, R. A., 1935 The sheltering of lethals. Am. Nat. **69:** 446–455.

Fromont-Racine, M., J. C. Rain and P. Legrain, 1997 Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat. Genet. **16:** 277–282.

Johnson, C. D., J. B. Rand, R. K. Herman, B. D. Stern and R. L. Russell, 1988 The acetylcholinesterase genes of *C. elegans*: identification of a third gene (*ace-3*) and mosaic mapping of a synthetic lethal phenotype. Neuron **1:** 165–173.

Karlin, S., and J. McGregor, 1971 On mutation selection balance for two-locus haploid and diploid populations. Theor. Pop. Biol. **2:** 60–70.

Kimura, M., and J. L. King, 1979 Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift. Proc. Natl. Acad. Sci. USA **76:** 2858–2861.

King, J. L., 1967 Continuously distributed factors affecting fitness. Genetics **55:** 483–492.

Kondrashov, A. S., 1984 Deleterious mutations as an evolutionary factor. 1. The advantage of recombination. Genet. Res. **44:** 199–217.

Kondrashov, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. Nature **336:** 435–440.

Kondrashov, A. S., 1993 Classification of hypotheses on the advantage of amphimixis. J. Hered. **84:** 372–387.

Kondrashov, A. S., 1995a Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? J. Theor. Biol. **175:** 583–594.

Kondrashov, A. S., 1995b Dynamics of unconditionally deleterious mutations: Gaussian approximation and soft selection. Genet. Res. **65:** 113–121.

Kondrashov, A. S., and M. Turelli, 1992 Deleterious mutations, apparent stabilizing selection and maintenance of quantitative variation. Genetics **132:** 603–618.

Lewontin, R. C., J. A. Moore, W. B. Provine and B. Wallace (Editors), 1981 *Dobzhansky's Genetics of Natural Populations I–XLIII*. Columbia University Press, New York.

Mayr, E., 1963 *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.

Milkman, R., 1967 Heterosis as a major cause of heterozygosity in nature. Genetics **55:** 493–495.

Milkman, R., 1978 Selection differentials and selection coefficients. Genetics **88:** 391–403.

Nowak, M. A., M. C. Boerlijst, J. Cooke and J. Maynard Smith, 1997 Evolution of genetic redundancy. Nature **388:** 167–171.

Ohno, S., 1970 *Evolution by Gene Duplication*. Springer-Verlag, Berlin.

Ohta, T., 1980 *Evolution and Variation of Multigene Families*. Springer-Verlag, Berlin.

Ohta, T., 1988 Evolution by gene duplication and compensatory advantageous mutations. Genetics **120:** 841–847.

Ohta, T., 1994 Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics **138:** 1331–1337.

Partridge, L., and N. H. Barton, 1993 Optimality, mutation and the evolution of aging. Nature **362:** 305–311.

Phillips, P. C., 1998 The language of gene interaction. Genetics **149:** 1167–1171.

Spassky, B., N. Spassky, H. Levene and T. Dobzhansky, 1958 Release of genetic variability through recombination. I. *Drosophila pseudoobscura*. Genetics **43:** 844–867.

Sved, J. A., T. E. Reed and W. F. Bodmer, 1967 The number of balanced polymorphisms that can be maintained in a natural population. Genetics **55:** 469–481.

Temin, R. G., H. U. Meyer, P. S. Dawson and J. F. Crow, 1969 The influence of epistasis on homozygous viability depression in *Drosophila melanogaster*. Genetics **61:** 497–519.

Thompson, V., 1986a Half-chromosome viability and synthetic lethality in *Drosophila melanogaster*. J. Hered. **77:** 385–388.

Thompson, V., 1986b Synthetic lethals: a critical review. Evol. Theory **8:** 1–13.

Wade, M. J., 1992 Epistasis, pp. 87–91 in *Keywords in Evolutionary Biology*, edited by E. F. Keller and E. A. Lloyd. Harvard University Press, Cambridge, MA.

Werth, C. R., and M. D. Windham, 1991 A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am. Nat. **137:** 515–526.

Whitlock, M. C., P. C. Phillips, F. B.-G. Moore and S. J. Tonsor, 1995 Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. **26:** 601–629.

Communicating editor: A. G. Clark