The Effect of Deleterious Mutations on Neutral Molecular Variation

B. Charlesworth, M. T. Morgan and D. Charlesworth

Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637-1573

Manuscript received November 23, 1992
Accepted for publication April 22, 1993

ABSTRACT

Selection against deleterious alleles maintained by mutation may cause a reduction in the amount of genetic variability at linked neutral sites. This is because a new neutral variant can only remain in a large population for a long period of time if it is maintained in gametes that are free of deleterious alleles, and hence are not destined for rapid elimination from the population by selection. Approximate formulas are derived for the reduction below classical neutral values resulting from such background selection against deleterious mutations, for the mean times to fixation and loss of new mutations, nucleotide site diversity, and number of segregating sites. These formulas apply to random-mating populations with no genetic recombination, and to populations reproducing exclusively asexually or by self-fertilization. For a given selection regime and mating system, the reduction is an exponential function of the total mutation rate to deleterious mutations for the section of the genome involved. Simulations show that the effect decreases rapidly with increasing recombination frequency or rate of outcrossing. The mean time to loss of new neutral mutations and the total number of segregating neutral sites are less sensitive to background selection than the other statistics, unless the population size is of the order of a hundred thousand or more. The stationary distribution of allele frequencies at the neutral sites is correspondingly skewed in favor of rare alleles, compared with the classical neutral result. Observed reductions in molecular variation in low recombination genomic regions of sufficiently large size, for instance in the centromere-proximal regions of Drosophila autosomes or in highly selfing plant populations, may be partly due to background selection against deleterious mutations.

THE neutral theory of molecular evolution provides an important basis for interpreting data on DNA sequence variation, since nucleotide substitutions that leave protein sequences unchanged may be under little or no selection (KIMURA 1983). Neutral theory makes predictions about patterns of nucleotide site variation within populations that can be tested against DNA sequence data (EWENS 1979; KIMURA 1983; GILLESPIE 1991; KREITMAN 1991). But the level of neutral variability can be reduced below classical neutral expectation by substitutions of alleles favored by natural selection at loci linked to the neutral sites (hitchhiking, MAYNARD SMITH and HAIGH 1974; THOMSON 1977; KAPLAN, HUDSON and LANGLEY 1989; STEPHAN, WIEHE and LENZ 1992; WIEHE and STEPHAN 1993). Reduced DNA variation in regions of restricted recombination in Drosophila has been interpreted as evidence for effects of hitchhiking (AGUADÉ, MIYASHITA and LANGLEY 1989a; STEPHAN and LANGLEY 1989; MIYASHITA 1990; BERRY, AJIOKA and KREITMAN 1991; BEGUN and AQUADRO 1991, 1992; MARTÍN-CAMPOS *et al.* 1992; STEPHAN and MITCHELL 1992; LANGLEY *et al.* 1993), suggesting that selective substitutions may not be uncommon in the Drosophila genome (BERRY, AJIOKA and KREITMAN 1991; BEGUN and AQUADRO 1992; CHARLESWORTH 1992; WIEHE and STEPHAN 1993). In this paper, we show that this conclusion may have to be partially modified, since reduced genetic diversity at neutral sites can be caused by another process, selection against linked deleterious alleles maintained by mutation ("background selection"). We also show that genetic diversity is expected to be reduced in inbreeding populations, for similar reasons.

It was previously found that the presence of loci with selection against deleterious alleles retards the fixation of alleles at linked neutral loci present at intermediate frequencies in finite populations (OHTA 1971; SVED 1972). Allele frequencies close to zero or one were not studied, however, and there has been no investigation of the general effect on neutral variability of selection acting on a genetic background of loci subject to recurrent mutation to deleterious alleles. The effect we report here was discovered in the course of work on the behavior of modifier alleles changing the selfing rate of populations with mutations to detrimental alleles occurring at many linked loci (CHARLESWORTH, MORGAN and CHARLESWORTH 1992). The mean times to fixation and loss of new neutral alleles introduced at a low frequency were found by simulation to be smaller than predicted by the standard neutral formulae (KIMURA and OHTA 1969a, b). We suggested that this effect was caused by chance associations of the neutral variants with

genotypes of different fitnesses (CHARLESWORTH, MORGAN and CHARLESWORTH 1992). An association with a rare high fitness genotype carrying few mutations will lead to rapid fixation of the neutral variant, whereas a variant in a background with low fitness will be rapidly eliminated. Such associations would be expected in highly inbreeding populations, and also in genomic regions with low frequencies of recombination. Here we show, using a simple theoretical argument and computer simulations, that this type of effect can cause an observable reduction in genetic diversity at neutral sites, assuming biologically plausible values of the per genome mutation rate to deleterious alleles and of the magnitude of selection against such mutations.

ANALYTICAL RESULTS FOR RANDOM-MATING POPULATIONS WITH NO RECOMBINATION

We shall consider two standard measures of genetic variation within a population, the nucleotide site diversity, π, and the number of segregating sites per neutral locus, s_n (NEI 1987; KREITMAN 1991). Since π is the probability that a pair of randomly chosen chromosomes differ at a given nucleotide site, it is far more influenced by high frequency neutral variants than is s_n, which includes unweighted contributions from all frequency classes. We might therefore expect the two measures of variability to be affected differently by selection against linked deleterious alleles (*cf.* TAJIMA 1989a).

When there is no recombination, the magnitude of the effect on the time to fixation and π for a large (but finite) diploid, random-mating population can be roughly predicted as follows. Studies of viability mutations in Drosophila populations suggest that selection against the heterozygous effects of deleterious mutations is such that a typical deleterious mutation can persist only for about 50 generations before elimination by selection, even in an infinitely large population (CROW and SIMMONS 1983). In a finite population, the mean time to loss of such a mutation is much shorter than this, less than 10 generations even for a population of a million individuals (KIMURA and OHTA 1969b). In the absence of recombination, a new neutral variant arising in a gamete with one or more deleterious mutations will remain associated with those mutations, and hence will be rapidly eliminated from the population. If selection against the deleterious mutations is not very weak, the least-loaded gamete class and its descendants will quickly become predominant in the population, due to their higher than average fitness. Neutral mutations can only spread to high frequencies in the entire population if they are transmitted through gametes that remain free of deleterious alleles.

As far as variants that are spreading to fixation are concerned, their drift to fixation in the least-loaded class is therefore the most important process determining the time taken, and the effective population size for this process in a large population should thus be approximated by f_0N_e, where f_0 is the equilibrium frequency of mutation-free gametes in an infinite population and N_e is the effective population size. The mean conditional time to fixation is then $4f_0N_e$, instead of the classical neutral value of $4N_e$ (KIMURA and OHTA 1969a).

The following argument shows that π should be reduced by a similar factor to that for the time to fixation. Under the infinite-sites model of neutral molecular variation (KIMURA 1969, 1971), we have

$$\pi = 2Nvh \tag{1}$$

where N is the number of breeding adults, v is the neutral mutation rate per nucleotide site, and H is the expected value of the sum of the diversity contributed by a mutation at a nucleotide site over all generations before loss or fixation of the mutation. H is given by

$$H = 2E\left\{\sum_t x_t(1 - x_t)\right\} \tag{2}$$

where x_t is the frequency at generation t of a neutral variant introduced as a single mutation at a given site. Under the classical neutral model with no selection at background loci, $H = 2N_e/N$, so that $\pi = 4N_ev$ (KIMURA 1969, 1971).

The contribution to H from a variant that persists in the population for a long enough time to have a large influence on its value will be reduced by a factor of f_0 below the classical neutral value, since the above argument shows that the effective population size for variants that are not rapidly lost from the population due to their association with deleterious alleles is f_0N_e. Hence, with background selection we have

$$\pi \approx 4f_0N_ev. \tag{3}$$

This argument breaks down if the deleterious alleles are very weakly selected against in the heterozygous state, so that $N_ehs < 1$ (KIMURA 1983, p. 44), since this means that they behave as effectively neutral, and hence will not influence the fate of a neutral mutation. Other circumstances in which this result breaks down are described in the RESULTS section below. A modification of Equation 3, which gives more accurate results for the case of small U and small N_e, is derived in the APPENDIX (Equation 6).

An alternative way of obtaining this result is through the coalescent method. The mean time to coalescence of the ancestries of two genes sampled from the population is approximately $2f_0N_e$ instead of the classical $2N_e$ (HUDSON 1990), since most of their ancestry must be contributed from a period when they were carried in mutation-free chromosomes. Equation

3 follows from the fact that π is equal to $2v$ times the mean coalescence time (HUDSON 1990). RICHARD HUDSON (personal communication) has derived a more exact formula, which takes into account the contribution to the coalescent time from the periods when members of a pair of sampled genes were carried in mutation-carrying chromosomes.

Neutral variants which are destined for loss will be eliminated from the population at a rate which is little influenced by the number of deleterious alleles with which they are associated, since random loss of a neutral variant can occur in all classes of gametes, and both stochastic and deterministic elimination take place rapidly unless population size is very large (KIMURA and OHTA 1969b; CROW and SIMMONS 1983). For $sh = 0.02$, we do not expect the mean conditional time to loss of a neutral variant to be much affected by background selection, in contrast to the mean time to fixation, until the population size exceeds several hundred thousand, when selection becomes so effective compared with drift that deleterious alleles are eliminated much faster than neutral alleles (KIMURA and OHTA 1969b). Therefore s_n will not be so strongly affected by background selection as π, since it contains a large contribution from variants that are lost rapidly from the population as well as from those that rise to high frequency. This argument also suggests that the frequency distribution of alleles at segregating sites will be skewed in favor of rare alleles, compared with the classical neutral expectation. As the population size becomes very large, this effect should disappear. Formulas for s_n and the mean time to loss for the case of small U are given in the APPENDIX.

These results are independent of the mode of selection against the deleterious alleles. Determination of f_0 in a given case requires the selection model to be specified. For simplicity, we shall assume equal effects on fitness of each locus and multiplicative fitness interactions between loci. In an infinitely large population with mutation-selection balance at many autosomal loci, the number of deleterious mutations per gamete at equilibrium follows a Poisson distribution with mean $n = U/2hs$ (KIMURA and MARUYAMA 1966; CROW 1970), where U is the mutation rate per diploid genome, s is the selection coefficient against homozygous mutations, and h is the dominance coefficient. Hence

$$f_0 = \exp\left(-\frac{U}{2hs}\right). \quad (4)$$

If $U/2hs$ is sufficiently large, $f_0 \ll 1$, and background selection against deleterious mutations is expected to have a major effect on neutral variation in regions of reduced recombination.

SIMULATION METHODS

The purpose of the simulations was to check the validity of the approximate results outlined above, and to evaluate the magnitude of the effect of background selection on the nucleotide site diversity and number of segregating sites per neutral locus, with different recombination frequencies and different rates of self-fertilization. We simulated repeated introductions of neutral variants into a population subject to recurrent mutation to deleterious alleles at many loci linked to the neutral sites, in finite diploid populations with discrete generations. The program was written in the C language for a Macintosh II computer with a Tektronix RP88 coprocessor.

In each generation, the sequence of events was mutation, reproduction and selection. The loci subject to selection were assumed to be either wild-type or mutant in state, and the neutral sites were assumed to have an initial allelic state and a mutated state. A genome contained 1024 selected loci, and 25 neutral sites at locations 500–524, *i.e.*, in the center of the set of selected loci. Haploid genotypes were stored as lists of mutant loci. The lists could undergo a recombination process, in the manner outlined by FELSENSTEIN and YOKOYAMA (1976). The loci were assumed to be on a single chromosome, or on several chromosomes, and the recombination frequency between adjacent selected loci will be denoted by r. Note that, for comparison with published data on genetic diversity in the *Drosophila melanogaster* genome (BEGUN and AQUADRO 1992), where the amounts of recombination are expressed in terms of coefficients of exchange (map distances between adjacent bands on the polytene chromosomes), division by 100 gives recombination frequencies between adjacent loci in females, assuming at least one locus per polytene band. For autosomal loci, the population recombination frequencies should be multiplied by a factor of one-half, since Drosophila males lack recombination; for sex-linked loci, a factor of two-thirds should be used (BEGUN and AQUADRO 1992).

Mutation at the loci under selection assumed a Poisson distribution of numbers of new mutations, with a mutation rate U for the whole diploid genome. In the mating process, gametes recombinant for both the selected loci and neutral sites were generated from each of two randomly chosen parental genotypes, if the population was random-mating. If self-fertilization was permitted, a fraction S of zygotes was formed by two gametes chosen from the same parent, and a fraction $1 - S$ from two random parents, where S was the assigned rate of self-fertilization of the population. The fitness of the zygote formed in this way was calculated by counting the numbers of loci for which it carried heterozygous and homozygous mutations. We assumed multiplicativity of selective effects at the selected loci, so that with selection coefficient s and dominance coefficient h the fitness of the genotype heterozygous for i and homozygous for j mutant al-

leles is $(1 - hs)^i(1 - s)^j$ (CROW 1970). For most of our simulations, we chose values of h and s that seem justified in the light of the Drosophila data on the fitness effects of detrimental mutations of minor effect, $h = 0.2$ and $s = 0.1$. This choice gives $hs = 0.02$, consistent with an estimated infinite population persistence time ($1/hs$) of 50 generations for a deleterious allele (CROW and SIMMONS 1983). The process of zygote production was repeated until a population of N new surviving zygotes was formed. Runs were started with all individuals wild-type at the selected loci.

To obtain the necessary statistics for describing the pattern of neutral variation, we have used the following procedure, which is appropriate for the infinite-sites model of molecular evolution (KIMURA 1969, 1971, 1983). By the principle of ergodicity, the mean time that an allele drifting to loss or fixation spends in a defined interval of allele frequency is proportional to the frequency with which independently introduced alleles at different sites are found in that interval, conditional on segregation at the sites in question (EWENS 1979, pp. 238–239). This enables results to be obtained in a very economical fashion, since we do not need to model the mutation process at the neutral sites, but merely to follow the fate of neutral variants introduced repeatedly into the population, until each is fixed or lost. The distribution of allele frequencies at segregating sites is then obtained from the average over many trials of the times spent in defined frequency intervals. From Equation 2, H is obtained from the mean diversity per site, $2x_t(1 - x_t)$, summed over the generations during the variants' passage to loss or fixation. From Equation 1, π is proportional to H. Similarly, s_n is proportional to T, the mean time to fixation or loss of a neutral variant (EWENS 1979, p. 238). The effect of background selection on s_n can thus be deduced from its effect on T. Note that this use of ergodicity is independent of any assumptions about selection or neutrality at the sites in question.

After an initial period of 2000 generations with mutation at the selected loci, to allow the population to reach a quasi-equilibrium state under mutation and selection, variation was introduced at all the neutral sites at a frequency of $1/2N$ at each site, 25 gametes being selected at random (with replacement) for the introduction of a variant at each of the 25 neutral sites. From that point onwards, the frequency of mutant alleles at each of the neutral sites was examined, and the genetic diversity $2x_t(1 - x_t)$ was calculated for each site at each generation, until all neutral variants were either lost or fixed. Conditional mean times to fixation and loss of the neutral variants were calculated, as well as the mean time to fixation or loss, and the distribution of times spent by segregating neutral sites in different gene frequency intervals was obtained.

With this method, the mutation rate to neutral alleles is not a parameter in the simulations, so we cannot give predicted absolute values of π and s_n. This is not necessary, however, as we are concerned solely with the values in the presence of a background of deleterious mutations, relative to those expected on the pure neutral theory. The method we employed introduces all neutral variants at a single group of linked sites. This does not affect the means from which we extract the relative π and s_n values, though it would bias estimates of correlations between variants at different neutral sites.

SIMULATION RESULTS

The effect of restricted recombination in random-mating populations: With no recombination, and with moderately strong selection acting on the selected loci, the H values and mean conditional fixation times derived from the simulations were remarkably close to the approximate theoretical values. These were calculated from the theory outlined above, replacing N_e in the standard formulae for H and fixation time (KIMURA 1969; KIMURA and OHTA 1969a) by f_0N_e. For $U = 0.01$, and $sh = 0.02$, the value of f_0 is 0.78; for $U = 0.1$ it is 0.08. The effects of background selection on genetic variability at the neutral sites from the simulations are expressed in terms of the values of π and s_n relative to their neutral expectations, calculated as described above.

Table 1 shows simulation results for complete linkage and random mating for a range of population sizes, with $U = 0.01$, $s = 0.1$ and $h = 0.2$. For each parameter set, 100,000 or more neutral variants were introduced and followed, the number being chosen such that at least 50–60 fixations of neutral variants occurred, so that a reliable estimate could be obtained of the contribution from introductions that resulted in fixation. The effects of background selection on π tended to be greater for large population size, but leveled off for $N \geq 3{,}200$, with little change above $N = 1{,}600$ (results not shown). For these larger populations, there is no significant disagreement between the theoretical and simulated results for π. A slight improvement in fit between the simulation and theoretical results for the smaller population sizes is given by Equation 6 of the APPENDIX. This takes into account the fact that even chromosomes carrying deleterious alleles make a non-zero contribution to H, which was ignored when deriving Equation 3. As might be expected, this effect is most marked for small N, and causes the reduction in π to increase with N, although the large stochastic error somewhat obscures the pattern.

The theoretical values of the mean time to loss and

TABLE 1

Effect of background selection on neutral variation in random-mating populations with no recombination

Population size, N	No. of neutral introductions	No. of fixations	Observed mean times to		Expected time to loss (neutral)	Expected fixation time		Observed values relative to neutral expectation	
			Loss	Fixation		Neutral	Corrected for selected loci	π	s_n
100	100,000	509	10.03	319	10.6	400	312	0.840	0.971
			0.100	6.24				0.020	0.008
200	100,000	244	11.04	645	12.0	800	623	0.800	0.944
			0.137	19.84				0.028	0.011
400	100,000	126	12.04	1,421	13.4	1,600	1,246	0.813	0.938
			0.204	58.68				0.044	0.015
800	200,000	123	13.27	2,448	14.8	3,200	2,492	0.768	0.921
			0.194	114.1				0.039	0.013
3,200	500,000	76	15.34	9,595	17.5	12,800	9,968	0.732	0.890
			0.232	482.7				0.046	0.014

The genetic background consisted of 1024 loci subject to deleterious mutation at rate 0.01 per diploid genome per generation. $s = 0.1$ and $h = 0.2$. Mean values of the quantities observed in the simulations are given, with standard errors below the means.

s_n relative to the classical neutral values given in the APPENDIX (Equations 8 and 10) tend to overestimate the reductions due to background selection compared with the simulation results, although the agreement improves as N increases (see Table 6). This is probably due to the fact that the formula for time to loss of a deleterious allele used to obtain Equation 7 of the APPENDIX assumes $2N_e sh \gg 1$, which is violated for the smaller population sizes (KIMURA and OHTA 1969b). As expected from the theoretical analysis, the effect of background selection on s_n and mean time to loss is much less than the effect on π and mean time to fixation, for the population sizes used in the simulations. The proportionate reduction in these statistics due to background selection increases with population size, as predicted by the analytical results.

The analytical approximation outlined above implicitly assumes that the effect of selection at background loci can be described adequately by the total mutation rate, and does not require the number of loci to be specified explicitly. Runs were done with different numbers of loci subject to mutation, to test this assumption for both the high and low mutation rates studied. These runs assumed $s = 0.1$ and $h = 0.2$. For the case of $U = 0.01$ and zero recombination, we did runs with 100 loci, instead of 1,024, *i.e.*, with about the same mutation rate per locus as in the runs with $U = 0.1$ and 1,024 loci. The relative values of π and s_n were not significantly different from the values with the larger number of loci, for population sizes of 800, 1,600 or 3,200. We also did runs with 512 instead of 1,024 loci for the case of $U = 0.1$. For zero recombination, or with some recombination ($r = 10^{-5}$ or 10^{-4}), the relative values of π and s_n were again not significantly different from the values with the larger number of loci, for populations of $N = 3,200$.

Table 2 shows the results of simulations with various mutation rates, selection coefficients, and dominance coefficients. These were intended to test the adequacy of the theoretical results for no recombination. The runs were done with a population size of 1,600. Good agreement with equation (3) was found with $hs \geq 0.02$, the biologically plausible value (CROW and SIMMONS 1983), over a wide range of predicted magnitudes of effects (see the top four lines of Table 2). When hs was very small, however, there was no longer good agreement, but the observed values of the quantities of interest were much less affected by background selection than predicted.

Accumulation of deleterious mutations by Muller's ratchet (FELSENSTEIN 1974; HAIGH 1978) in these rather small populations may contribute to this discrepancy. With $s = 0.2$, the expected number of mutations per gamete in an infinite population at equilibrium is 2.5 with $h = 0.1$, and 5 when $h = 0.05$, but the observed mean numbers after approximately 75,000 generations in runs with a population size of 1,600 were 3.8 and 7.1, respectively. With the implausibly low h value of 0.05, there was thus moderate accumulation of deleterious mutations, and very slight accumulation with $h = 0.1$. The observed relative π values in these two runs were somewhat higher than predicted from the theory ignoring such accumulation, especially with the more recessive case when mutation accumulation was more pronounced. But the ratchet is not the main reason for the disagreement with the theory. Runs with a very low selection coefficient ($s = 0.01$, $hs = 0.002$) were also done with a low mutation rate ($U = 0.01$). Since the equilibrium mean number of mutations per gamete in an infinite population is the same as with $U = 0.1$ and $hs = 0.02$, the ratchet will proceed at the same rate in the two

TABLE 2

Effects of background selection with different selection parameters

Mutation rate	Selection coefficient	Dominance coefficient	No. of runs	No. of fixations	Expected π	Observed mean π	Observed mean s_n
0.1	0.1	0.2	200,000	68	0.08	0.12	0.52
0.1	0.2	0.2	200,000	62	0.29	0.31	0.57
0.1	1.0	0.2	200,000	74	0.78	0.78	0.82
0.1	0.2	0.1	200,000	72	0.08	0.13	0.53
0.1	0.2	0.05	200,000	72	0.007	0.15	0.58
0.1	1.0	0.04	200,000	61	0.28	0.28	0.57
0.01	0.01	0.2	100,000	61	0.08	0.98	0.94

The genetic background consisted of 1024 loci subject to deleterious mutation. Mean values of the quantities observed in the simulations are given. All loci were completely linked. The population was random mating, with 1600 individuals.

cases (HAIGH 1978). With N = 1,600, the runs with weaker selection and lower mutation rate yielded results similar to those expected on the purely neutral theory i.e., there was little effect of background selection (see the last line of Table 2).

As discussed in the APPENDIX, the reason for the lack of effect of background selection in this case is the fact that, with large U/hs, a chromosome carrying a small number of deleterious mutations is at a selective advantage relative to the population as a whole. If U is small and the population size is small, a new neutral mutation arising in such a chromosome may drift to a high frequency before the chromosome accumulates sufficient deleterious mutations to be eliminated from the population. Neutral mutations arising in some of the chromosomes carrying deleterious mutations can therefore contribute significantly to genetic diversity. This effect will tend to disappear as population size increases, since the time taken for an allele to reach a high frequency increases, and there is a greater chance that the chromosome in question will be eliminated before this happens. To check this interpretation, the case of U = 0.01 and hs = 0.002 was run with a population size of 6,400 instead of 1,600. The mean values over 500,000 runs of π and s_n relative to neutral expectation were 0.18 ± 0.02 and 0.67 ± 0.01 respectively, so that both statistics were substantially reduced in value compared with neutral expectation. Genetic diversity is thus reduced by background selection even with weak selection against deleterious mutations, unless population size is very small.

Figure 1 shows the effect of recombination on π and s_n for two different mutation rates, U = 0.01 and 0.1. The runs with U = 0.1 were done using large populations, to ensure that the population would come into a quasi-equilibrium state, *i.e.*, that deleterious mutations would not accumulate at the selected loci by either Muller's ratchet or fixation (KIMURA, MARUYAMA and CROW 1963). As expected, recombination reduces the effect of selection at the background

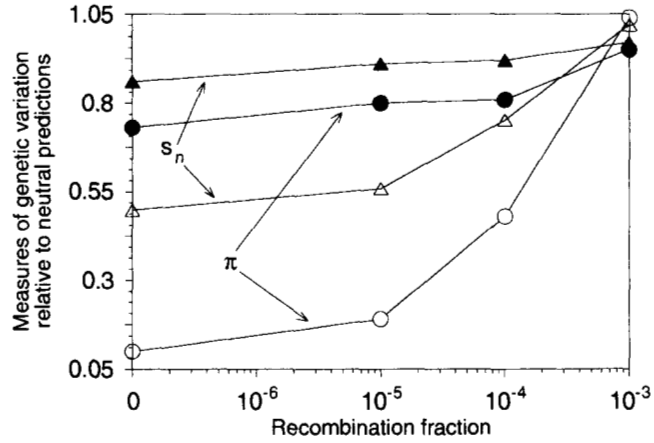

FIGURE 1.—Values of nucleotide site diversity (π) and numbers of segregating sites per locus (s_n), relative to their classical neutral expectations, for neutral variants in outcrossing populations on a genetic background with deleterious mutations, with various recombination fractions r between adjacent loci. The population size was 3,200, the selection coefficient on the mutations was 0.1, and the dominance coefficient was 0.2. Filled symbols: U = 0.01; open symbols: U = 0.1.

loci. No reduction in variability below classical neutral expectation was detectable when adjacent loci recombined with a frequency of 10^{-3}. Other population sizes also gave effects on π and fixation times that agreed with these results, though the reduction in s_n, and to a lesser extent π, was greater with larger N.

The equilibrium distribution of allele frequencies, conditioned on segregation at the neutral sites, was more J-shaped than expected from classical neutral theory when linkage was tight, with the vast majority of neutral sites having alleles at low frequency (Figure 2). The distribution when the recombination frequency is 10^{-3} is very close to the neutral distribution. With tight linkage, background selection therefore causes the distribution to resemble that for slightly deleterious rather than neutral variants (EWENS 1979, p. 239). This reflects the fact that it has a much greater effect on variants that might otherwise rise to high frequencies than on those destined to be lost rapidly (see above). As discussed in the ANALYTICAL RESULTS

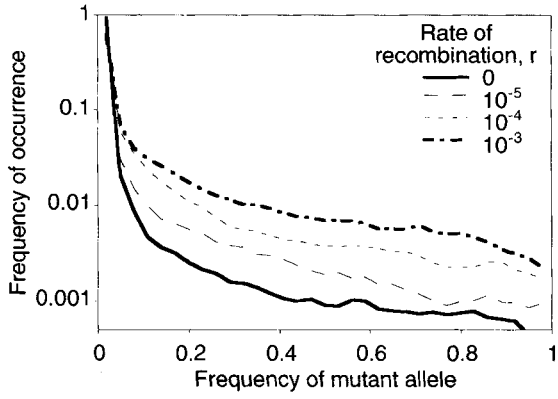

FIGURE 2.—Frequency distributions of the neutral variants in the computer simulations of outcrossing populations with deleterious mutations occurring at a rate of 0.1 per diploid genome per generation in the genetic background, for various recombination frequencies. These data were obtained from the simulation results on the numbers of times that alleles were observed in different frequency intervals (frequency intervals of 0.03 were used). The population size was 3,200, the selection coefficient was 0.1, and the dominance coefficient was 0.2.

section, this pattern is likely to become negligible when population size is very large.

The effects of inbreeding: The effects of high levels of self-fertilization are shown in Table 3 and Figure 3. The runs were done with a recombination frequency of 10^{-3} between adjacent loci, a value that yielded no discernible effect of selection at background loci in random-mating populations. The mutation rate was either 0.1 (with all loci assumed to be on a single chromosome) or 0.5 for the case when the loci were spread among five chromosomes. The standard selection parameter values of $h = 0.2$ and $s = 0.1$ were used. For calculations of the classical neutral expectations, we used effective population sizes adjusted to take account of the frequency of selfing, according to the formula $N_e = N/(1 + F)$, where F is the equilibrium inbreeding coefficient, $S/(2 - S)$ (POLLAK 1987). With complete selfing, there is effectively no recombination. In this case, the expected background selection factor for reducing N_e in the formulas for fixation time and π is $f_0 = \exp\{-U/2s\}$, the frequency of mutant-free gametes in an infinitely large selfing population at equilibrium (CHARLESWORTH, MORGAN and CHARLESWORTH 1991). With the selection and dominance coefficients used in these runs ($s = 0.1$, $h = 0.2$), $f_0 = 0.61$ for the case of $U = 0.1$ (Figure 3 and upper part of Table 3) and for $U = 0.5$ it is 0.08 (lower part of Table 3).

The expected reductions in fixation time and π were found in runs with complete selfing. For lower selfing rates, there is no explicit formula, since there is a significant amount of recombination. For $S \geq 0.75$, the genetic diversity measures were markedly lower than predicted by classical neutral theory (Figure 3), especially in runs with a mutation rate of 0.5.

DISCUSSION

A recombination frequency of only 10^{-3} in our simulations, corresponding to a map length of approximately 100 centimorgans with 1,000 loci per chromosome, is sufficient to remove any detectable effect of background selection against deleterious alleles in random-mating populations (Figure 1). This suggests that classical neutral theory should predict the pattern of silent nucleotide site variation in most regions of the genome, unless population changes or hitchhiking events have intervened to distort the picture. Deviations from neutral expectation in these regions in random-mating populations can thus be taken as evidence for such events. But when recombination or outbreeding is greatly restricted, the presence of loci subject to recurrent deleterious mutation can have significant effects on the behavior of neutral variants.

The major question is whether or not the total mutation rate for deleterious alleles in regions of restricted recombination is ever large enough that effects of the magnitude found in population studies can be explained by this mechanism. In the next sections, we discuss the magnitudes to be expected. Before doing this, it should be mentioned that runs with very different numbers of loci subject to mutation establish that the results do not depend strongly on the number of loci, but are determined chiefly by the rate of mutation in genomic segments in which recombination is infrequent (see SIMULATION RESULTS section).

Restricted recombination in Drosophila: Table 4 gives estimates of the sizes, mutation rates, frequencies of mutation-free gametes (f_0), and expected reductions in genetic diversity, for several regions of the *D. melanogaster* genome with restricted recombination (ASHBURNER 1989, Chap. 11). The estimates of the regional sizes (expressed as the proportions of the total euchromatin represented by each region) were obtained from information on the relative sizes of the different chromosome arms (ASHBURNER 1989; CHARLESWORTH, LAPID and CANADA 1992), and from data on the proportion of DNA within each chromosome arm represented by the regions in question (BOLSHAKOV, ZHARKIKH and ZHIMULEV 1985). The mutation rates were estimated by multiplying a total diploid per genome mutation rate of 1 (MUKAI *et al.* 1972; HOULE *et al.* 1992) by the sizes of each region. The map lengths were obtained from the standard genetic maps of LINDSLEY and ZIMM (1992), except for the smaller basal region of the *X*, whose length was estimated from the recombination frequency between *mal* (located at 19D/E) and the centromere (SCHALET and LEFEVRE 1976). For the *X* chromosome, which is expressed in the hemizygous states in males, $f_0 = \exp\{-3U/(2s[2h + 1])\}$, where U is the diploid mutation rate for the region of the *X* in question. For

TABLE 3

Effects of background selection on neutral variation in selfing and partially selfing populations

Population size, N	Selfing rate, S	No. of fixations	Observed mean times to		Expected (neutral) mean times to		Observed values (relative to neutral)	
			Loss	Fixation	Loss	Fixation	π	s_n
One chromosome, $U = 0.1$								
1600	1.0	71	7.26	2,011	8.07	3,200	0.72	0.88
1600	0.9	73	8.39	2,585	8.87	3,250	0.80	0.94
1600	0.75	60	9.75	3,618	9.69	4,000	0.84	1.06
1600	0.5	67	11.58	4,845	21.11	4,800	0.90	0.97
3200	1.0	55	7.27	3,256	8.76	6,400	0.53	0.79
3200	0.9	67	8.66	5,805	9.64	7,040	0.75	0.90
3200	0.5	58	13.54	10,963	13.15	9,600	1.12	1.03
3200	0	33	17.00	12,690	17.50	12,800	1.04	1.02
Five chromosomes, $U = 0.5$								
1600	1.0	65	4.53	316	8.07	3,200	0.10	0.50
1600	0.9	45	5.90	951	8.87	3,250	0.23	0.60
1600	0.75	52	8.15	2,158	9.69	4,000	0.49	0.74
1600	0.5	63	11.78	4,125	12.11	4,800	0.90	0.92

The genetic background consisted of 1000 loci subject to deleterious mutation. $s = 0.1$ and $h = 0.2$. The recombination fraction between adjacent loci was 10^{-3}. The number of simulations for each parameter set was 200,000, except for the case of no selfing, for which 500,000 replicates were run.

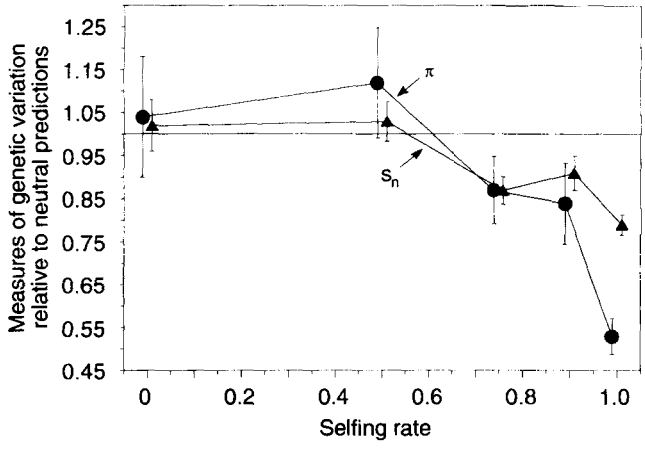

FIGURE 3.—Nucleotide site diversity values (π) and numbers of segregating sites per locus (s_n), relative to their classical neutral expectations (taking into account the effect of the relevant levels of selfing on the effective population size), for neutral variants on a genetic background with deleterious mutations in inbreeding populations of 3,200 individuals with various selfing rates, S. The error bars are the standard errors of the mean values of the statistics. The selection coefficient against the deleterious mutations was 0.1, the dominance coefficient was 0.2, and the mutation rate was 0.1.

the autosomes, Equation 4 was used. The standard values of the selection parameters, $s = 0.1$ and $h = 0.2$, were used in the calculations.

The estimated reductions in π for the different genomic regions were calculated using the following method. For the X chromosome, a population r value of 10^{-5} between adjacent loci in our simulations would correspond to a region approximately $1{,}024 \times 100 \times 10^{-5} \times 3/2 = 1.5$ centimorgans in length. Defining the tip of the X as the region distal to and including 3B (1.2 centimorgans), it is thus reasonable to assume that π for this region should correspond to that for an average r of slightly less than 10^{-5}. Since the simulations suggest that an r value of 10^{-5} is almost indistinguishable from zero, this would correspond to a reduction in π of $1 - f_0 = 0.24$. The expected reductions in π for the smaller basal regions of the X, larger autosomes, and the fourth chromosome (which has no crossing over [ASHBURNER 1989]), were obtained in a similar way. The results for the larger basal regions are less certain, as indicated in the table. For the X chromosome, an r value of 10^{-4} in our simulations would correspond to the set of genes within 15 centimorgans of the centromere (*i.e.*, to genes located proximal to 13F), for which we estimate $f_0 = 0.59$. Figure 1 shows that a change from $r = 0$ to $r = 10^{-4}$ roughly halves the expected reduction in π, suggesting that the reduction in diversity for genes located in this region might be approximately 20%. The estimate for the larger basal regions of the major autosomes was obtained by a similar procedure. These estimates should be regarded as only approximate, as the simulations assumed a central location of the neutral sites, and loci in the more distal parts of these regions might show weaker effects.

These expectations can be compared with data on molecular variation in natural populations. For the tip of the X chromosome, background selection against deleterious alleles is clearly inadequate as a full explanation of reduced variability. In some populations of *D. melanogaster* and in *Drosophila simulans*, variation in the *yellow-achaete-scute* locus at the extreme tip of the X chromosome is significantly reduced (AGUADÉ, MIYASHITA and LANGLEY 1989a; BEGUN and AQUADRO 1991; MARTÍN-CAMPOS *et al.* 1992), although there is apparently no reduction in other

TABLE 4

Expected reductions in genetic diversity in regions of the *D. melanogaster* genome

Region	Polytene bands	Proportion of total genome, mutation rate	Map length	f_0	Expected reduction in π
X chromosome					
Tip	Telomere–3B	0.026	1.2	0.76	0.24
Base	Centromere–19D/E	0.004	1.5	0.96	0.04
Base	Centromere–13F	0.05	15	0.59	≈0.2
Chromosome *2* or *3*					
Base	Chromosome *2*: 36F-43A	0.06	2	0.22	0.78
	Chromosome *3*: 76A-85A	0.06	2	0.22	0.78
Base	Chromosome *2*: 32F-49B	0.17	20	0.01	≈0.50
	Chromosome *3*: 68E-88F	0.17	20	0.01	≈0.50
Chromosome *4*		0.01	0	0.78	0.22

populations of *D. melanogaster* (BEECH and LEIGH-BROWN 1989; EANES, LABATE and AJIOKA 1989; MARTÍN-CAMPOS *et al.* 1992). From Table 4, it is apparent that our model cannot account for the drastic reduction in variability sometimes observed for *y-ac-sc*, but it is consistent with evidence for the smaller apparent reduction in variability in other populations and at other loci in this region (AGUADÉ, MIYASHITA and LANGLEY 1989b; BEGUN and AQUADRO 1991).

Similarly, the total lack of variability at *su-f* (20F) (LANGLEY *et al.* 1993) and in the *fw* locus of *Drosophila ananassae* (STEPHAN and MITCHELL 1992), both located at the base of the *X*, and for the one locus on the fourth-chromosome locus that has been sequenced, ci^D (BERRY, AJIOKA and KREITMAN 1991), cannot be accounted for by our model. The somewhat low π for *Zw* in *D. melanogaster* (located at 18E, in the larger basal region of the *X*) noted by MIYASHITA (1990) might be due to background selection. Table 4 shows that large reductions in variation are expected for genes in the smaller basal regions of the major autosomes, and a reduction of about 50% might be expected for the larger basal regions. Few data on these regions are available as yet, but a survey of nucleotide-site variability at *cta* (40F) shows reduced variability in *D. melanogaster* (MARTA WAYNE, personal communication).

Effects of asexuality or partial self-fertilization: Effects of the kind discussed above should also be seen in highly selfing or asexual species, where recombination is effectively absent. (In the latter case, the reduction factor for diversity is the equilibrium frequency of mutant-free clones, equal to exp $(-U/hs)$ for a diploid.) In both cases, the mutation rate for the entire diploid genome is appropriate for predicting the effects of background selection, so a U value of about 1 can be assumed for higher organisms (KONDRASHOV 1988; CHARLESWORTH, CHARLESWORTH and MORGAN 1990; HOULE *et al.* 1992). Our studies of selfing populations assumed a single chromosome and a mutation rate of 0.1, which is probably reasonable for one chromosome, or five chromosomes each with this mutation rate.

When the selfing rate exceeded about 75%, we found strongly reduced diversity values for neutral variants, even with a mixture of self-fertilization and random mating. It is well known that partial selfing is correlated with lower allozyme variability within populations, and the best data indicate reductions in genetic diversity for allozyme loci of about 50% in selfing plant populations compared with obligate outcrossers (HAMRICK and GODT 1990; SCHOEN and BROWN 1991), the reduction expected just from the halving of the effective population size under complete selfing (POLLAK 1987). Selfing species of slugs also show reduced allozyme diversity (MCCRACKEN and SELANDER 1980; FOLTZ *et al.* 1982). However, the breeding systems were not quantified in most of the studies, so the magnitude of the predicted reduction from this cause is not certain.

Larger differences have sometimes been found. Self-fertilizing *Mimulus micranthus* populations (estimated to be about 80% selfing) had only about one quarter of the allozyme diversity compared with the mainly outcrossing *Mimulus guttatus*, and diversity for chloroplast sequences in selfing populations was reduced to a similar extent (FENSTER and RITLAND 1992). In completely selfing species, organelle genomes will behave as though linked to the nuclear genome, so that background selection acting on the nuclear genome could cause reduced variation in organelle genomes. The effect of the rate of selfing on the level of neutral variability in organelle genomes has not yet been studied, however.

Allozyme data may not be appropriate, however, as these variants may be maintained by some form of balancing selection (GILLESPIE 1991). Indeed, if allozyme diversity is not greatly reduced in selfing, compared with outcrossing populations, this might suggest that balancing selection is occurring. It would there-

fore by very interesting to compare sequence diversity for synonymous and nonsynonymous changes, in selfers and outcrossers. There is currently little information on DNA variation, but a survey of RFLP variation in related species of tomatoes suggested a pattern similar to, but apparently more extreme than, that for allozymes (MILLER and TANKSLEY 1990). Similarly, asexual species are notoriously depauperate in genetic variability, though the interpretation of the data is complicated by recent origin and by hybridization in these cases, and there are few data on diversity within evolutionarily old clonal lineages (VRIJENHOEK, ANGUS and SCHULTZ 1977, 1978).

Reduced molecular variation in selfing or asexual populations could, of course, be due to a variety of causes. In addition to the effect that we have been investigating, which may reduce nucleotide site diversity to 50% or less of the classical neutral expectation for selfing rates close to one, the expected value for a selfing population is already 50% less than the value for random mating, due to lower effective population size (POLLAK 1987). Hitchhiking by favorable gene substitutions, which has been shown to be at least as important in selfing populations as with tight linkage (HEDRICK 1980), may also reduce variation in selfers and asexuals. Selfing populations will also lose variation at loci with heterozygote advantage (KIMURA and OHTA 1971). In addition, populations of selfing organisms can potentially be founded by single individuals, so that severe bottlenecks may also have contributed to reduced levels of variation. The data from selfing plant populations show some evidence for such bottlenecks (BARRETT and KOHN 1991), especially the finding of more variation in levels of genetic diversity between populations than in outbreeding species (SCHOEN and BROWN 1991). In many cases, asexual populations seem to have originated in the relatively recent past from a single ancestral individual (QUATTRO, AVISE and VRIJENHOEK 1991, 1992), and have thus experienced a bottleneck effect. Nevertheless, we have found that, with plausible parameter values, the reduction in genetic diversity measures caused by the process investigated here can be greater than 80% in highly selfing populations, presumably with a greater reduction for asexuals. This suggests that background selection may explain a significant portion of the observed reduction in variation in these cases.

Distinguishing between different causes of reduced molecular variation: Unfortunately, many of the possible factors mentioned above that may contribute to a reduction in molecular variation are expected to have similar observable consequences for the pattern of neutral variability. For example, TAJIMA's test will only partially discriminate between them. The test evaluates the significance of the difference, D, between the value of $4N_e v$ obtained on the neutral theory from the mean number of pairwise differences between sequences in the sample, and the value estimated from the number of segregating sites in the sample (TAJIMA 1989a). After complete hitchhiking or loss of heterozygosity caused by fixation of a single chromosome, the neutral variation in a population is due entirely to mutations that have arisen since that event. The restoration of neutral variation by mutation and drift after its elimination by population bottlenecks and hitchhiking events is a long-term process, and is accompanied by a transient excess of rare alleles compared with the statistical equilibrium under drift and mutation (MARUYAMA and FUERST 1984, 1985; WATTERSON 1984; TAJIMA 1989b). A recent population expansion without a bottleneck will have a similar effect. These events will produce negative values of TAJIMA's D, as with background selection if population size is moderate (a few tens of thousands or less; see Table 6 of the APPENDIX). This makes background selection hard to discriminate from these other processes. If effective population size is known to be very large, however, as seems to be the case in the Drosophila species that have been used in studies of DNA sequence variation (KREITMAN 1991), D is expected to be very small, and so a significant value of D would be inconsistent with background selection.

Hitchhiking events or population bottlenecks that merely reduce levels of genetic variation could produce the opposite effect: with such partial effects, loss of heterozygosity is accompanied by severe loss of rare alleles, so that the value of s_n relative to π is lower than expected on equilibrium neutral theory (MARUYAMA and FUERST 1985; TAJIMA 1989b). Such a partial effect will usually be expected for population bottlenecks, since surviving populations are unlikely to have remained small for a sufficient time that all variability is lost. Reductions of allele numbers in populations after bottlenecks have been observed in several species of plants (BARRETT and KOHN 1991) and animals (JANSON 1987; LEBERG 1992). A positive value of D would thus indicate partial hitchhiking or bottleneck effects. The possible effects are summarized in Table 5.

The utility of tests such as TAJIMA's D is also limited by the fact that small numbers of sampled genomes (as are usually employed in studies of DNA variation) provide very poor information on the number of segregating sites, and on the shape of the allele frequency spectrum, since low frequency variants are either excluded from the sample or have their frequencies inflated to the reciprocal of the sample size. The results described in this paper relate only to population parameters. It is clearly important to study the properties of the corresponding sample statistics, in order to determine whether or not the expected effects of background selection could be detected in

TABLE 5

Expected effects of different factors on the two measures of neutral variation (diversity and number of segregating sites), and on TAJIMA's *D* statistic

Process	Regions affected	Effects on diversity measures
Bottlenecks		
Partial (*i.e.*, genetic variation reduced but not completely eliminated)	All loci, including organelles, but not in all populations of a species	Loss of diversity, severe loss of rare alleles, π reduced less than s_n, $D > 0$
Complete (especially in self-fertilizing populations)	As for partial bottlenecks	π and s_n initially reduced to 0, π recovers more slowly than s_n, $D < 0$
Hitchhiking		
Partial (*i.e.*, genetic variation reduced but not completely eliminated before a recombination event)	Only certain loci, especially in regions of restricted recombination (all loci in selfers, including organelles), some or all populations	As for partial bottlenecks, $D > 0$
Complete (*i.e.*, single haplotype fixed)	As for partial hitchhiking	As for complete bottlenecks, $D < 0$
Background selection		
Outcrossing populations	Regions of restricted recombination	π more reduced than s_n, $D < 0$, (unless population size is very large, in which case $D \approx 0$)
Selfing populations	All genomic regions, including organelles	As for outcrossers

samples of realistic size. Our initial results from such a study indicate that background selection is unlikely to produce a significantly negative D in the case of samples of genes from the Drosophila centromeric or telomeric regions, even if population sizes are much lower than seems to be the case in the Drosophila species studied to date (see above). Findings of significantly negative D values for such samples, as in the study of *y-ac-sc* in *D. melanogaster* by MARTÍN-CAMPOS *et al* (1992), would therefore strongly suggest that other factors have been involved.

The best hope for discriminating between different possibilities lies in the fact that background selection can only affect the level of neutral variation under sharply defined conditions: greatly restricted outbreeding, or very restricted recombination over a sufficiently large genomic region that the net mutation rate to deleterious alleles is high enough to cause a low equilibrium frequency of mutation-free gametes (see Figures 1 and 3). Given that these conditions are met, neutral sites in the genomic regions that meet the requirements should be affected uniformly, in all populations of the species and in related species with the same genome organization. In contrast, population bottlenecks should have genome-wide effects, but might affect different populations or taxa differently. Hitchhiking events should be confined to regions of restricted recombination or to highly inbreeding populations, but are likely to be sporadic in their occurrence, so that different populations or species will probably vary in the extent to which they show reduced variation. The differences between populations of *D. melanogaster* in the level of variation in the *y-ac-sc* region (MARTÍN-CAMPOS *et al.* 1992) further suggest that local hitchhiking or population size changes are responsible for reductions in variability in this case.

Extensions of the selection model: The results described above have assumed fixed values of s and h for all loci, and multiplicative fitness effects across loci, for analytical simplicity and convenience in running the simulations. In reality, s and h must vary across loci (CROW and SIMMONS 1983), and it is important to ask whether this has any serious consequences for our conclusions. In addition, interactions between loci may cause deviations from multiplicative fitness effects. As we have seen, the effect of background selection is largely determined through the frequency of mutation-free chromosomes, f_0, in a nonrecombining region of the genome. The effect of deviations from our simplifying assumptions about the mode of selection can therefore be understood by considering their effects on f_0.

The effect of variation in the selection parameters between loci can be studied as follows. In a random-mating population, with multiplicative fitnesses and with variation in sh (assumed to be uncorrelated with the mutation rate per locus), the number of mutations per chromosome will follow a Poisson distribution whose mean is given by replacing $U/2sh$ in Equation 4 by the product of $U/2$ and the mean of $1/sh$. The latter is equivalent to the mean persistence for a detrimental gene isolated from a natural population, estimated by CROW and SIMMONS (1983) to be approximately 50. Hence, the "standard" value of sh used in our studies is appropriate even if there is variation among loci in s and h, as far as the estimate of the maximal effect of background selection for a given mutation rate in a random-mating population is concerned.

For the case of a wholly selfing population, the

mean of $1/s$ replaces $1/sh$ in the formula for f_0 (see RESULTS). The data of HOULE *et al.* (1992) suggest that the mean value of s for the effects of new detrimental mutations on net fitness in a Drosophila population cage environment is considerably less than 0.1 (see their Table 1). The mean of $1/s$ is approximately equal to $(1 + C^2)/\bar{s}$, where $\bar{s}$ and C are the mean and coefficient of variation of s. The value of $s = 0.1$ that we have used in our studies of selfing populations may thus overestimate f_0, and hence underestimate the effect of background selection.

These considerations suggest that, if anything, the existence of variation among loci in the effects of mutant alleles on fitness may increase the effect of background selection on neutral variation. The only situation in which this would not be the case is if the distribution of effects were strongly skewed or bimodal, such that $2N_esh$ is of the order of 1 or less for a large fraction of deleterious mutations. As discussed in the ANALYTICAL RESULTS section, such weakly selected mutations would not interfere with the fate of linked neutral alleles. The above estimates of f_0 would then overestimate the impact of background selection. But with the effective population sizes of a million or more required to account for observed levels of silent-site DNA diversity in natural populations of Drosophila (KREITMAN 1991), extremely small selection coefficients would be needed for this effect to be important. Mutations subject to such weak selection would not be detectable in the laboratory measures of the effects of mutations on fitness cited above, and so this possibility can almost certainly be disregarded, except for species with small effective population sizes.

KONDRASHOV (1988) has argued that synergistic effects of deleterious mutations on fitness (*i.e.*, effects that increase with the number of mutations carried by an individual) may be widespread and have important consequences for the evolution of breeding systems. The effects of synergistic selection on f_0 can be examined as follows. CHARLESWORTH (1990), following KIMURA and MARUYAMA (1966), studied mutation-selection balance in a diploid sexual population with no recombination and with synergistic selection. His results show that the variance at equilibrium under synergistic selection is smaller than the mean, indicating a departure from the Poisson distribution found with multiplicative fitnesses. The decrease in variance suggests that f_0 will be smaller than the Poisson value for the same equilibrium mean, so that the values we have used are conservative. This is confirmed by direct calculation of f_0 from the recurrence relations. For the selection parameters used in the top part of Table 2 of CHARLESWORTH (1990), with $U = 0.1$ and 0.01 we find $f_0 = 0.002$ and 0.25 respectively, compared with corresponding Poisson values of 0.004 and 0.28. These reductions in f_0 are unlikely to be measurable experimentally.

Conclusions: Overall, while it is clear that the more extreme examples of reduced nucleotide site diversity in sections of the Drosophila genome cannot be accounted for by background selection against deleterious mutations, so that some hitchhiking events must be invoked, examples of less extreme, but unusually low, diversity values are consistent with this model. It may therefore be at least a contributory factor to the correlation between molecular variability and rate of recombination described by BEGUN and AQUADRO (1992). In addition, it is probable that this process contributes to the reduced levels of molecular variation seen in asexual and highly selfing taxa.

We are grateful to JAMES CROW, who independently proposed a similar mechanism, for helpful discussions. We also thank RICHARD HUDSON and two anonymous reviewers for suggestions for improving the manuscript.

LITERATURE CITED

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989a Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989b Restriction-map variation at the *zeste-tko* region in natural populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6:** 123–130.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

BARRETT, S. C. H., and J. R. KOHN, 1991 Genetic and evolutionary consequences of small population size in plants: implications for conservation, pp. 3–30 in *Genetics and Conservation of Rare Plants*, edited by D. A. FALK and K. E. HOLSINGER. Oxford University Press, New York.

BEECH, R. N., and A. J. LEIGH-BROWN, 1989 Insertion-deletion variation at the *yellow-achaete-scute* region in two natural populations of *Drosophila melanogaster*. Genet. Res. **53:** 7–15.

BEGUN, D. J., and C. F. AQUADRO, 1991 Molecular population genetics of the distal portion of the *X* chromosome in Drosophila: evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics **129:** 1147–1158.

BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

BOLSHAKOV, V. N., A. A. ZHARKIKH and I. F. ZHIMULEV, 1985 Intercalary heterochromatin in Drosophila II. Heterochromatic features in relation to local DNA content along the polytene chromosomes of *Drosophila melanogaster*. Chromosoma **92:** 200–208.

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

CHARLESWORTH, B., 1992 New genes sweep clean. Nature **356:** 475–476.

CHARLESWORTH, B., D. CHARLESWORTH and M. T. MORGAN, 1990 Genetic loads and estimates of mutation rates in very inbred plant populations. Nature **347:** 380–382.

CHARLESWORTH, B., A. LAPID and D. CANADA, 1992 The distribution of transposable elements within and between chromo-

somes in a population of *Drosophila melanogaster*. II. Inferences on the nature of selection against elements. Genet. Res. **60:** 115–130.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1991 Multilocus models of inbreeding depression with synergistic selection and partial self-fertilisation. Genet. Res. **57:** 177–194.

CHARLESWORTH, D., M. T. MORGAN and B. CHARLESWORTH, 1992 The effect of linkage and population size on inbreeding depression due to mutational load. Genet. Res. **59:** 49–61.

CROW, J. F., 1970 Genetic loads and the cost of natural selection, pp. 128–177 in *Mathematical Topics in Population Genetics*, edited by K.-I. 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. CARSON and J. N. THOMPSON. Academic Press, London.

EANES, W. F., J. LABATE and J. W. AJIOKA, 1989 Restriction map variation associated with the *yellow-achaete-scute* region in four populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6:** 492–502.

EWENS, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, Berlin.

FELSENSTEIN, J., 1974 The evolutionary advantage of recombination. Genetics **78:** 737–756.

FELSENSTEIN, J., and S. YOKOYAMA, 1976 The evolutionary advantage of recombination. II. Individual selection for recombination. Genetics **83:** 845–859.

FENSTER, C. B., and K. RITLAND, 1992 Chloroplast DNA and isozyme diversity in two Mimulus species (Scrophulariaceae) with contrasting mating systems. Am. J. Bot. **79:** 1440–1447.

FOLTZ, D. W., H. OCHMAN, J. S. JONES, S. M. EVANGELISTI and R. K. SELANDER, 1982 Genetic population structure and breeding systems in arionid slugs (Mollusca: Pulmonata). Biol. J. Linn. Soc. **17:** 225–241.

GILLESPIE, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

HAIGH, J., 1978 The accumulation of deleterious genes in a population. Theor. Popul. Biol. **14:** 251–267.

HAMRICK, J. L., and M. J. GODT, 1990 Allozyme diversity in plant species, pp. 43–63 in *Plant Population Genetics, Breeding, and Genetic Resources*, edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WEIR. Sinauer, Sunderland, Mass.

HEDRICK, P. W., 1980 Hitch-hiking: a comparison of linkage and partial selfing. Genetics **94:** 791–808.

HOULE, D., D. K. HOFFMASTER, S. ASSIMACOPOULOS and B. CHARLESWORTH, 1992 The genomic rate of mutation for fitness in *Drosophila*. Nature **359:** 58–60.

HUDSON, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–42.

JANSON, K., 1987 Genetic drift in small and recently founded populations of the marine snail *Littorina saxatilis*. Heredity **58:** 31–37.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

KIMURA, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutation. Genetics **61:** 893–903.

KIMURA, M., 1971 Theoretical foundations of population genetics at the molecular level. Theoret. Popul. Biol. **2:** 174–208.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

KIMURA, M., and T. MARUYAMA, 1966 The mutational load with epistatic gene interactions in fitness. Genetics **54:** 1337–1351.

KIMURA, M., T. MARUYAMA and J. F. CROW, 1963 The mutational load in small populations. Genetics **48:** 1303–1312.

KIMURA, M., and T. OHTA, 1969a The average number of generations until fixation of a mutant gene in a finite population. Genetics **61:** 763–771.

KIMURA, M., and T. OHTA, 1969b The average number of generations until extinction of an individual mutant gene in a finite population. Genetics **63:** 701–709.

KIMURA, M., and T. OHTA, 1971 *Theoretical Topics in Population Genetics*. Princeton University Press, Princeton, N.J.

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

KREITMAN, M., 1991 Detecting selection at the level of DNA, pp. 204–221 in *Evolution at the Molecular Level*, edited by R. K. SELANDER, A. G. CLARK and T. S. WHITTAM. Sinauer, Sunderland, Mass.

LANGLEY, C. H., J. MACDONALD, N. MIYASHITA and M. AGUADÉ, 1993 Lack of correlation between interspecific divergence and intraspecific polymorphism at the *suppressor of forked* region of *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **90:** 1800–1803.

LEBERG, P. L., 1992 Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. Evolution **46:** 477–494.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego, Calif.

MARTÍN-CAMPOS, J. M., J. P. COMERON, N. MIYASHITA and M. AGUADÉ, 1992 Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics **130:** 805–816.

MARUYAMA, T., and P. A. FUERST, 1984 Population bottlenecks and nonequilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. Genetics **108:** 745–763.

MARUYAMA, T., and P. A. FUERST, 1985 Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population derived from a large steady-state population by means of a bottleneck. Genetics **111:** 675–689.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. Genet. Res. **231:** 1114–1116.

MCCRACKEN, G. F., and R. K. SELANDER, 1980 Self-fertilization and monogenic strains in natural populations of terrestrial slugs. Proc. Natl. Acad. Sci. USA **77:** 684–688.

MILLER, J. C., and S. D. TANKSLEY, 1990 RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theoret. Appl. Genet. **80:** 437–448.

MIYASHITA, N., 1990 Molecular and phenotypic variation of the *Zw* locus region in *Drosophila melanogaster*. Genetics **125:** 407–419.

MUKAI, T., S. I. CHIGUSA, L. E. METTLER and J. F. CROW, 1972 Mutation rate and dominance of genes affecting viability. Genetics **72:** 335–355.

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

OHTA, T., 1971 Associative overdominance caused by linked detrimental mutations. Genet. Res. **18:** 277–286.

POLLAK, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117:** 353–360.

QUATTRO, J. M., J. C. AVISE and R. C. VRIJENHOEK, 1991 Molecular evidence for multiple origins and hybridogenetic fish clones (Poeciliidae: Poeciliopsis). Genetics **127:** 391–398.

QUATTRO, J. M., J. C. AVISE and R. C. VRIJENHOEK, 1992 Mode of origin and sources of genotypic diversity in triploid gynogenetic fish clones (Poeciliopsis: Poeciliidae). Genetics **130:** 621–628.

SCHALET, A., and G. LEFEVRE, 1976 The proximal region of the *X* chromosome, pp. 847–902 in *The Genetics and Biology of Drosophila*, Vol. 1b., edited by M. ASHBURNER and E. NOVITSKI. Academic Press, London.

SCHOEN, D. J., and A. H. D. BROWN, 1991 Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proc. Natl. Acad. Sci. USA **88:** 4494–4497.

STEPHAN, W., and C. H. LANGLEY, 1989 Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121:** 89–99.

STEPHAN, W., and S. J. MITCHELL, 1992 Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. Genetics **132:** 1039–1045.

STEPHAN, W., T. H. E. WIEHE and M. W. LENZ, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theoret. Popul. Biol. **41:** 237–254.

SVED, J. A., 1972 Heterosis at the level of the chromosome and at the level of the gene. Theoret. Popul. Biol. **3:** 491–506.

TAJIMA, F., 1989a Statistical method for testing the neutral mutation hypothesis. Genetics **123:** 585–595.

TAJIMA, F., 1989b The effect of change in population size on DNA polymorphism. Genetics **123:** 597–601.

THOMSON, G., 1977 The effect of a selected locus on a linked neutral locus. Genetics **85:** 752–788.

VRIJENHOEK, R. C., R. A. ANGUS and R. J. SCHULTZ, 1977 Variation and heterozygosity in sexually versus clonally reproducing populations of Poeciliopsis. Evolution **31:** 767–781.

VRIJENHOEK, R. C., R. A. ANGUS and R. J. SCHULTZ, 1978 Variation and clonal structure in a unisexual fish. Am. Nat. **112:** 41–55.

WATTERSON, G. A., 1984 Allele frequencies after a bottleneck. Theoret. Popul. Biol. **26:** 387–407.

WIEHE, T. H. E., and W. STEPHAN, 1993 Analysis of a genetic hitch-hiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. (in press).

Communicating editor: R. R. HUDSON

APPENDIX

For small mutation rates and with no recombination, an approximate treatment of the effects of small population size on π and s_n can be developed as follows, for the case of a random-mating population. If $U/2hs < 1$, the mean fitness of chromosomes that carry one or more deleterious mutations will be lower than that of the whole population. If U is small, such chromosomes will usually be eliminated from the population before they experience further mutations at the selected loci. A neutral mutation arising in a chromosome carrying i deleterious mutations can thus be treated as though it remains permanently associated with these mutations. The total diversity, H_i, contributed by such a new neutral mutation during its persistence in the population can therefore be approximated by the equivalent of Equation 2 of the text, modified for the case of a new deleterious mutation whose heterozygous carriers have fitness $w_i = (1 - hs)^i/\exp(-U/2)$ relative to the population mean fitness. Writing $s_i = 1 - w_i$ for the effective selection coefficient against the neutral mutation, H_i relative to the classical neutral value is given approximately by Equation 3.18 of KIMURA (1983, p. 45) as:

$$H_i = \frac{(e^{4N_e s_i} - 4N_e s_i - 1)}{4N_e s_i (e^{4N_e s_i} - 1)}. \quad (5)$$

If $4N_e s_i \gg 1$, as is likely to be true in most cases of interest, H_i can be approximated by $1/2N_e s_i$.

It should be noted that KIMURA's equation assumes that the fitness of the heterozygote is exactly intermediate between the fitnesses of the two homozygotes at the locus, whereas we have assumed partial recessivity of the deleterious mutations. This should not cause serious error, since the deleterious mutations will not rise to a high frequency in a population of reasonable size, and so their homozygous effects are largely irrelevant.

Let f_i be the equilibrium frequency of chromosome carrying i mutations, which is given by a Poisson distribution with mean $U/2hs$ on the assumptions used here. Summing over all classes, the equilibrium nucleotide site diversity is thus given by

$$\pi \approx 4N_e v \left(f_0 + \sum_{i=1}^{\infty} f_i H_i\right). \quad (6)$$

With very large $N_e hs$, the bracketed terms other than f_0 can be neglected, and this expression then reduces to Equation 3 of the text.

The mean time to fixation of a new mutation, t_f, is equal to f_0 times the classical neutral value (see ANALYTICAL RESULTS section). This result is independent of N_e, provided that $2N_e hs > 1$, since chromosomes carrying deleterious mutations are doomed to loss.

The mean time to loss of a neutral mutation, t_l, can also be calculated approximately. For a neutral variant introduced into a chromosome carrying i deleterious mutations ($i > 0$), the mean time to loss, t_i, is equivalent to that for the mean time to loss of a deleterious mutation with selective disadvantage s_i in the heterozygote. Using the results of KIMURA and OHTA (1969b), the value of this relative to the neutral value is given by

$$t_i \approx \frac{0.423 - \ln(2s_i)}{\ln(2N)}. \quad (7)$$

Summing over all classes, the relative mean time to loss is given by

$$t_l \approx \left(f_0 t_0 + \sum_{i=1}^{\infty} f_i t_i\right) \quad (8)$$

where the relative mean time to loss of a neutral mutation in a chromosome free of deleterious mutations is $t_0 \approx \ln(2f_0 N)/\ln(2N)$.

As noted in the SIMULATION METHODS section, the mean time to loss or fixation, T, of a new mutation is proportional to s_n. For a neutral mutation in a chromosome carrying i deleterious mutations ($i > 0$), the

TABLE 6

Comparisons of simulation results and theoretical results for low U

N	100	200	400	800	1600	3200	10^5	10^6	∞
π									
Theoretical	0.846	0.813	0.796	0.787	0.783	0.781	0.779	0.779	0.778
Simulated	0.840	0.800	0.813	0.768		0.732			
t_f									
Theoretical	0.778	0.778	0.778	0.778	0.778	0.778	0.778	0.778	0.778
Simulated[a]	0.800	0.806	0.888	0.765		0.750			
t_l									
Theoretical	0.901	0.887	0.976	0.867	0.859	0.853	0.832	0.824	0.778
Simulated	0.946	0.920	0.899	0.897		0.877			
s_n									
Theoretical	0.878	0.870	0.856	0.856	0.850	0.845	0.828	0.821	0.778
Simulated	0.971	0.944	0.938	0.921		0.890			

All results are expressed relative to the classical neutral values. The simulation results are taken from Table 1. Random mating and complete linkage were assumed. $s = 0.1$, $h = 0.2$ and $U = 0.01$.

[a] These estimates have high standard errors, because only small numbers of fixation events occurred.

type of argument used above gives this time relative to the neutral value as

$$T_i \approx \frac{0.423 - \ln(2s_i)}{(\ln[2N] + 1)} \tag{9}$$

and so s_n relative to the classical neutral value is given by

$$s_n \approx \left(f_0 T_0 + \sum_{i=1}^{\infty} f_i T_i\right) \tag{10}$$

where $T_0 \approx (1 + \ln[2f_0N])/(1 + \ln[2N])$.

The predictions of these formulas for different population sizes can be compared with the simulation results of Table 1, for the case $U = 0.01$, $sh = 0.02$, which satisfies the assumptions made above. The results are shown in Table 6, where all the population statistics are expressed relative to the classical neutral values.

If $U/2hs < 1$ but $1/U$ is small compared with the mean time to loss of a chromosome carrying a deleterious mutation, these results will tend to underestimate the effect of background selection in reducing neutral variability when N is small, since chromosomes will accumulate additional mutations, and hence be eliminated more rapidly than allowed for here. If $U/2hs > 1$, chromosomes carrying one or a few deleterious mutations will have higher mean fitnesses than the overall population, and so it is no longer appropriate to use the formulas for deleterious mutations when analyzing their fate. Under these conditions, background selection will be less effective for a given value of N than predicted by the considerations used here, since neutral mutations will be eliminated less rapidly than expected. If U is large, this effect will be less marked than if it is small, since chromosomes with small numbers of deleterious mutations will rapidly become less fit than the overall population, and are then eliminated rapidly (see discussion in the SIMULATION RESULTS section).