

The Pattern of Neutral Molecular Variation Under the Background Selection Model

D. Charlesworth, B. Charlesworth and M. T. Morgan¹

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

Manuscript received June 6, 1995

Accepted for publication September 1, 1995

ABSTRACT

Stochastic simulations of the infinite sites model were used to study the behavior of genetic diversity at a neutral locus in a genomic region without recombination, but subject to selection against deleterious alleles maintained by recurrent mutation (background selection). In large populations, the effect of background selection on the number of segregating sites approaches the effect on nucleotide site diversity, *i.e.*, the reduction in genetic variability caused by background selection resembles that caused by a simple reduction in effective population size. We examined, by coalescence-based methods, the power of several tests for the departure from neutral expectation of the frequency spectra of alleles in samples from randomly mating populations (Tajima's, Fu and Li's, and Watterson's tests). All of the tests have low power unless the selection against mutant alleles is extremely weak. In *Drosophila*, significant Tajima's tests are usually not obtained with empirical data sets from loci in genomic regions with restricted recombination frequencies and that exhibit low genetic diversity. This is consistent with the operation of background selection as opposed to selective sweeps. It remains to be decided whether background selection is sufficient to explain the observed extent of reduction in diversity in regions of restricted recombination.

SELECTION against deleterious alleles maintained by recurrent mutation ("background selection") causes a reduction in the amount of genetic variability at linked neutral sites (Charlesworth *et al.* 1993; Charlesworth 1994; Hudson 1994; Hudson and Kaplan 1994). This is because a new neutral variant can only remain in the population for a long period of time if it is maintained in gametes that are free of deleterious alleles, so that the effective population size in the presence of deleterious mutations is less than that based on the numbers of individuals in the population [see also Fisher (1930, Chapter 6); Peck (1994); Barton (1995)]. We previously showed that there can be large effects of background selection on genetic diversity in random-mating populations with no genetic recombination and in populations reproducing exclusively or largely by self-fertilization (Charlesworth *et al.* 1993). These effects decrease rapidly with increasing recombination frequency or rate of outcrossing (Charlesworth *et al.* 1993; Hudson 1994; Hudson and Kaplan 1994). We concluded that observed reductions in molecular variation in low-recombination genomic regions, such as the centromere-proximal regions of *Drosophila* autosomes (Langley 1990; Begun and Aquadro 1992; Aquadro *et al.* 1994; Kreitman and Wayne 1994), or in highly selfing plant populations

(Hamrick and Godt 1990; Schoen and Brown 1991; Wolff 1991; Fenster and Ritland 1992; Hanfstein *et al.* 1995) may be partly due to background selection against deleterious mutations.

Our previous simulation work on background selection employed a method that introduced neutral variants at a number of sites and followed their frequencies until either loss or fixation. This method enabled us to study the values of diversity measures relative to their neutral expectations for a range of parameter values, for instance with various recombination frequencies between the selected loci (Charlesworth *et al.* 1993). With the population sizes studied previously (up to 6400 diploid individuals) we found that of the two measures of genetic diversity commonly used in studies of molecular variation (Nei 1987; Kreitman 1991), the mean number of differences per nucleotide site between a pair of alleles at the neutral locus (π) was more strongly reduced than the number of segregating sites (s_n). This reflects the greater numbers of low-frequency variants compared with neutral expectation (Charlesworth *et al.* 1993, Figure 2). Approximate analytical results suggested, however, that no difference in the effect of background selection on π and s_n would be expected in very large populations (Charlesworth *et al.* 1993).

The purpose of the present work is to investigate more fully the differences between the behavior of the two diversity measures. This is clearly important, as different factors that affect genetic diversity may have different effects on the distribution of the frequencies of variants at a neutral locus (Charlesworth *et al.* 1993;

Corresponding author: D. Charlesworth, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637-1573. E-mail: debchar@pondside.uchicago.edu

¹ Present address: Department of Biology, McGill University, 1205 Ave. Dr. Penfield, Montreal, Quebec, Canada.

BRAVERMAN *et al.* 1995; SIMONSEN *et al.* 1995). Such differences may provide a means of discriminating between alternative hypotheses to explain reduced genetic variability, such as background selection and hitchhiking. Our previous simulations were confined to small population sizes. Furthermore, they did not provide information on the properties of samples from populations and so did not permit us to assess the power of tests of the divergence of empirical data on allele frequency spectra from neutral expectation. We here present the results of simulations in which we examine both the diversity properties of very large randomly mating populations and the results of tests of divergence from neutrality using samples of realistic size.

SIMULATION METHODS

Two different programs were used to simulate the infinite sites model of molecular variation (KIMURA 1971) with the addition of selected loci in the genetic background. We assumed complete linkage between all loci, both those subject to selection against deleterious mutations (termed the selected loci in what follows) and also the neutral locus whose variation is of interest. Multilocus simulations thus need only keep track of the frequencies of gametes carrying different numbers of deleterious mutant alleles, and the genotypes at individual selected loci need not be followed. In both types of simulation, we assumed a randomly mating population of N diploid individuals. The loci subject to selection were assumed to be either wild type or mutant in state, with identical selection coefficients at each locus. In each generation, the sequence of events was mutation, reproduction, and selection. The numbers of new mutations per diploid individual that arise each generation were assumed to be Poisson distributed with mean U .

With large population size, homozygotes for deleterious mutations can be ignored, so that selection on diploid genotypes can be modeled by performing selection on haploid genotypes with fitnesses 1 for nonmutant and $1 - sh$ for mutant alleles, where s and h are, respectively, the selection and dominance coefficients for mutant alleles (CROW 1970). For the multilocus case, the fitnesses of haploid genotypes were usually calculated multiplicatively from these fitnesses, *i.e.*, $w_j = (1 - sh)^j$ is the fitness of the genotype with j heterozygous mutations. We also carried out some studies with synergistic epistasis among mutant effects on fitness (KIMURA and MARYAMA 1966; CROW 1970; KONDRASHOV 1988; CHARLESWORTH 1990). In this case the quadratic-exponential fitness model was used, such that $w_j = \exp -j(\alpha + \frac{1}{2}\beta j)$ (CHARLESWORTH 1990). For both selection models, the equilibrium frequency distribution of gametes in each mutant class was assumed to be the stable state of the population during the course of a run. The vector of numbers of gametes in each class was found by multiplying the frequencies (f_i) by twice the number of diploid individuals in the population. This set of $2Nf_i$ values was stored as an array of real numbers. With multiplicative fitness interactions among the selected loci, the relevant distribution of numbers of mutant alleles per gamete is Poisson with mean $U/(2sh)$ (KIMURA and MARUYAMA 1966; CROW 1970). Most runs employed values of h and s of 0.1 and 0.2, *i.e.*, $sh = 0.02$, corresponding to a biologically reasonable infinite-population mean persistence time ($1/sh$) of 50 generations for a deleterious allele (CROW and SIMMONS 1983). With synergistic epistasis, the equilibrium distribution of numbers of mutations per gamete was found by iteration of the recursion

equations of KIMURA and MARUYAMA (1966) for the case of a sexual population with no recombination. Following CHARLESWORTH (1990) the selection coefficient against a heterozygous mutation in an equilibrium population with synergistic epistasis is given by the ratio of U to the equilibrium mean number of mutations per individual. This can be equated to sh in the multiplicative fitness model, enabling direct comparisons to be made between the two models.

Pseudo-sampling model: The first type of simulation was an infinite sites model with repeated introductions of variants at initially nonsegregating neutral loci, similar to the model studied previously (CHARLESWORTH *et al.* 1993). A pseudo-sampling scheme, similar to that of LI (1980) and KIMURA and TAKAHATA (1983) for single locus and two locus simulations, was used to increase the speed of the simulations and to enable us to examine the relative behavior of the two commonly used measures of genetic diversity, π and s_n , in very large populations. A vector of frequencies of gametes carrying different numbers of deleterious mutations in each generation was generated from that in its parental generation by mutation, followed by selection and drift. Under suitable conditions, the first two processes, mutation and selection, can be treated deterministically, and only drift need be dealt with stochastically.

To simulate a pair of segregating alleles at a neutral locus, the program follows changes in the frequencies of gametes with the variant allele within each class characterized by its number of mutations at the selected loci. When a mutation at the neutral locus occurs, a single individual with the new variant is assumed to be present in one class, chosen randomly from the equilibrium distribution of the numbers of mutant alleles per gamete. Because there is no recombination and back-mutation is neglected, there is no movement from classes of genotypes with more selected mutations to classes with fewer mutations, but new individuals enter each class by mutation from classes with fewer mutations (so new neutral mutations occurring in genotypes without deleterious alleles will not necessarily remain in mutant-free genotypes), while selection against genotypes with mutations reduces the numbers of individuals in the class. The program contains expressions for the frequencies, p_i , of gametes with i deleterious mutations and that carry the neutral variant. The transition probability from a gamete with j deleterious mutations to one with i mutations is the product of the marginal fitness w_j of a gamete with j mutations ($w_j = \sum_i p_i w_{ji}$) and the probability m_k that such a genotype undergoes $k = i - j$ mutations. The new frequency of gametes with i mutations that carry the variant is thus

$$p'_i = \frac{\sum_j p_j w_j m_k}{\bar{w}} \quad (1)$$

where \bar{w} is the population mean fitness (KIMURA and MARUYAMA 1966).

Drift of the neutral alleles is then simulated by sampling gametes within each class of genotypes with respect to the selected loci. When the number of individuals with the rarer allelic type at the neutral locus was < 15 in a given class, a random Poisson sampling event was performed to generate a new value for the number of gametes with the rarer allele. When both allelic types were present in 15 or more individuals of a class, the sampling method of ITO was performed as described by LI (1980). The magnitude of the change in allele frequency was calculated as $\sqrt{p_i(1 - p_i)/2Nf_i}$. The sign of the change was chosen by drawing a uniform pseudo-random number between 0 and 1; positive or negative changes were made, depending on whether the number was greater or less than 0.5. As in simulations in which all genetic processes are

treated stochastically (CHARLESWORTH *et al.* 1993), the frequencies of loss and fixation, the times to loss and fixation, and the sum of the genetic diversity during the sojourn of a neutral variant in the population were followed for many replicate introductions of neutral mutations into these populations. The values of π and s_n relative to their values in the purely neutral cases were determined from these statistics as described by CHARLESWORTH *et al.* (1993).

For population sizes of 3200 and 6400 diploid individuals (the largest populations that we were previously able to run), but not with smaller population sizes, the results were very similar to those of simulations in which all stages of the life cycle were modeled stochastically (see Table 1 below). As our interest here is in large population sizes, however, it appears that the approximate simulation method is very satisfactory for our purposes.

Coalescence model: The second program generated results that allowed us to examine the standard genetic diversity measures (NEI 1987; KREITMAN 1991) and values of both TAJIMA's and FU and LI's D statistics (TAJIMA 1989; FU and LI 1993) in samples of alleles at a single neutral locus subject to mutation according to the infinite sites model (KIMURA 1971). The diversity measures for such a sample are the mean number of nucleotide differences between pairwise combinations of sampled genes and the number of segregating sites in the sample (denoted here by k and S_n , respectively, to distinguish them from the population values). In addition, the genes in the sample can be classed into different allelic categories according to the history of the neutral mutations that they carry, such that two genes can belong to the same allelic class only if they have identical genealogies and carry the same neutral mutations. The sample can then be characterized by its allelic configuration (n_1, n_2, \dots, n_m) , where m is the total number of different allelic classes in the sample, and n_i is the number of times that the i th class is represented (EWENS 1972). A useful statistic for measuring the departure of the allele frequency spectrum from neutral expectation is provided by the sample homozygosity $F = \sum_i n_i^2 / (\sum_i n_i)^2$ (WATTERSON 1977, 1978). The greater abundance in the population of rare alleles under background selection compared with neutrality (CHARLESWORTH *et al.* 1993) means that the expectations of the sample values of the two kinds of D statistics should be smaller than in the neutral case, whereas F is expected to be larger.

To study these sample statistics, we generated genealogical trees of the neutral alleles whose sequence is examined, using the standard coalescence approach (HUDSON 1990), modified to take into account selection at the linked loci in the genetic background. The amounts of evolutionary change at the neutral locus are determined in the standard manner, using the times between nodes on the tree and a random process of assignment of neutral mutations to the branches (HUDSON 1990).

To make the method clear, we first describe the process for a neutral locus in a selectively neutral genetic background and then extend it to the case of linkage to selected loci in the genetic background. This process involves three stages, namely: (1) setting up a tree culminating in a sample of alleles at the neutral locus (with a preset sample size), including determining the generation at which each coalescent event occurred; (2) determining the number of mutations at the neutral locus between each node of the tree and its ancestral node; and (3) determining the genealogy of the sampled alleles at the neutral locus and the numbers of site differences between them. Stages 2 and 3 are the same whether there is or is not background selection, but the first stage differs.

Neutral case: A set of samples of neutral alleles was generated as described by HUDSON (1990). For each such sample,

we determined S_n , the number of segregating sites in the sample, recorded how many of each allelic type are present, and used the matrix of numbers of site differences between allelic types to calculate the average number of pairwise differences (k) between the $n(n-1)/2$ different genes in the sample, including the zero differences between different alleles in the sample that are of the same allelic type. With real molecular data, the estimate of π is found from $k = \pi \times$ the number of bases in the sequence studied. We used k and S_n to calculate TAJIMA's standardized D statistic (TAJIMA 1989), D_T . S_n values were scaled by the quantity $a_1 = \sum_{i=1}^{n-1} (1/i)$. This facilitates comparisons with the k values, because both quantities have the same expectations under neutrality, $M = 4Nv$ (WATTERSON 1975; TAJIMA 1989). Here, we will use $\theta = S_n/a_1$ to denote the estimator of M obtained by equating the observed value of S_n to its expectation.

We also computed the sum of the numbers of mutations in the external branches of the tree, as well as the total number of mutations in the tree. These two quantities were used to calculate FU and LI's standardized D statistic, D_F , using their method without an outgroup, as the phylogeny and branch lengths of the trees are, of course, completely known in our simulations (FU and LI 1993). As discussed below, this implies that the method will be less powerful when applied to real data (for which branch lengths must be estimated from the data) than suggested by our results. Samples containing no variation at the neutral locus were not included in these calculations, as the values of the D statistics are zero for such samples. The program was run for 2000 replicate samples, and the values of the quantities just described were stored so that their distributions could be calculated.

Background selection case: The procedure for the case with natural selection against deleterious mutations in the genetic background is based on the same method, with the additional feature that the routine that generates the tree also keeps track of the numbers of mutations in the gametes. To do this, the program sets up a matrix of the probabilities that a gamete with i mutations in a given generation derives from an ancestor in the preceding generation that has j mutations, similar to the approach used by HEY (1991) for generalizations of the classical coalescent process (see also HUDSON and KAPLAN 1994). We denote these probabilities by Q_{ij} . To calculate these values, taking into account selection on deleterious mutations, we use the selection models described above. As before, we assume that the population is at equilibrium under mutation and selection, with fixed frequencies of genotypes with different numbers of deleterious mutations.

Given a vector of frequencies, f_j , of gametes having different numbers of mutations, the transition matrix to the next generation is readily calculated, using similar logic to that leading to Equation 1. If m_k is the probability that a gamete experiences $k = i - j$ mutations, the probability that a gamete with j mutations will produce a descendant gamete with i mutations ($i \geq j$) is

$$p_{ij} = \frac{f_j w_i m_k}{\bar{w}} \quad (2)$$

where $i \geq j$. By BAYES' theorem, the probability that a gamete with i mutations derives from a gamete with j mutations is

$$Q_{ij} = \frac{p_{ij}}{\sum_j p_{ij}} \quad (3)$$

Once the program has been initialized by calculating the Q_{ij} values, the tree for a sample can be set up in a similar manner to that described above for the neutral case. In the first generation, n gametes with specified numbers of deleteri-

TABLE 1
Results of pseudo-sampling simulations of large random-mating populations with no recombination for various mutation rates to deleterious alleles

Population size, N^a	Observed mean times to		Expected time to loss (neutral)	Expected fixation time		Observed values relative to neutral expectation (\pm SE)	
	Loss	Fixation		Neutral	Corrected for selected loci	π	s_n
$U = 0.1$ ($f_0 = 0.082$)							
800	8.34	428.1	14.67	3,066	393	0.1584 \pm 0.0034	0.5370 \pm 0.0041
*800	9.75	1,167	14.67	3,066	393	0.30 \pm 0.022	0.60 \pm 0.008
1,600	8.42	753	16.06	6,131	645	0.1191 \pm 0.0028	0.4960 \pm 0.0029
1,600	9.18	781	16.06	6,131	645	0.11 \pm 0.007	0.52 \pm 0.004
3,200	8.58	1,222	17.50	12,618	1,177	0.1042 \pm 0.0027	0.4670 \pm 0.0023
*3,200	9.21	1,288	17.50	12,618	1,177	0.09 \pm 0.006	0.48 \pm 0.002
5,000	8.66	1,631	18.39	19,716	1,760	0.0820 \pm 0.0030	0.4448 \pm 0.0026
6,400	8.74	2,176	18.91	25,491	2,234	0.0806 \pm 0.0021	0.4389 \pm 0.0017
*6,400	9.38	2,182	18.91	25,491	2,234	0.09 \pm 0.005	0.46 \pm 0.003
10,000	8.08	3,468	19.81	40,000	3,280	0.0939 \pm 0.0026	0.3910 \pm 0.0017
100,000	8.45	32,427	24.41	400,000	32,966	0.0762 \pm 0.0021	0.3337 \pm 0.0011
200,000	8.01	73,404	25.80	800,000	65,791	0.0824 \pm 0.0024	0.3012 \pm 0.0011
500,000	7.13	157,691	27.63	2,000,000	164,301	0.0778 \pm 0.0024	0.2507 \pm 0.0010
1,000,000	6.33	326,705	29.02	4,000,000	328,461	0.0774 \pm 0.0022	0.2131 \pm 0.0009
5,000,000	4.33	1,653,190	32.24	20,000,000	1,641,837	0.0801 \pm 0.0024	0.1331 \pm 0.0009
$U = 0.05$ ($f_0 = 0.287$)							
1,600	10.10	2,002	16.06	6,155	1,881	0.3084 \pm 0.0087	0.6143 \pm 0.0066
3,200	10.56	3,764	17.51	12,683	3,751	0.3091 \pm 0.0090	0.5928 \pm 0.0061
*3,200	11.32	3,764	17.51	12,683	3,751	0.33 \pm 0.018	0.64 \pm 0.006
6,400	11.01	7,122	18.90	25,366	7,385	0.2799 \pm 0.0077	0.5687 \pm 0.0050
*6,400	11.52	7,122	18.90	25,366	7,385	0.31 \pm 0.038	0.60 \pm 0.012
10,000	11.29	10,762	19.80	40,000	11,566	0.3117 \pm 0.0099	0.5630 \pm 0.0058
100,000	11.49	115,578	24.41	400,000	114,720	0.2739 \pm 0.0079	0.4679 \pm 0.0039
200,000	11.15	227,118	25.80	800,000	229,310	0.2834 \pm 0.0080	0.4353 \pm 0.0038
500,000	10.641	521,656	27.63	2,000,000	573,099	0.2680 \pm 0.0075	0.3855 \pm 0.0034
$U = 0.025$ ($f_0 = 0.535$)							
1,600	12.14	3,648	16.09	6,236	3,420	0.5681 \pm 0.0166	0.7557 \pm 0.0119
3,200	12.97	7,463	17.52	12,750	6,907	0.5903 \pm 0.0174	0.7495 \pm 0.0113
6,400	13.47	12,351	18.91	25,501	13,732	0.5295 \pm 0.0146	0.7140 \pm 0.0093
10,000	13.96	22,975	19.80	39,845	21,410	0.5234 \pm 0.0154	0.7063 \pm 0.0092
25,000	14.75	53,565	21.65	100,000	53,583	0.5339 \pm 0.0178	0.6868 \pm 0.0097
50,000	15.30	114,398	23.02	200,000	107,084	0.4901 \pm 0.0211	0.6622 \pm 0.0107
100,000	15.39	217,948	24.41	400,000	214,085	0.5026 \pm 0.0162	0.6356 \pm 0.0082
200,000	17.03	454,255	25.80	800,000	428,270	0.6735 \pm 0.0649	0.7003 \pm 0.0030
500,000	15.69	1,250,360	27.63	2,000,000	1,070,553	0.5798 \pm 0.0245	0.5895 \pm 0.0105
1,000,000	14.23	2,119,820	29.02	4,000,000	2,141,106	0.5374 \pm 0.0220	0.5020 \pm 0.0050

Runs that were done without approximation, by full Monte Carlo simulation of the genotypes of the entire population, are indicated by asterisks.

^aDue to approximations involved in the simulations, these employed population sizes slightly less than the stated N values, and the expected fixation times were based on the actual sizes used in the simulations.

ous mutations are formed by sampling from the f_i distribution. The first step in each generation (going back in time) is to obtain the number of mutations in the ancestral gamete of each gamete. These numbers are obtained by picking random Poisson deviates, using the Q_{ij} probabilities as the expected values. Coalescence of alleles at the neutral locus can occur only between two gametes with the same number of deleterious mutations, so that each class of gamete with respect to the number of deleterious mutations is dealt with separately. For one such class, with k deleterious mutations, the program calculates the number L_k of such nodes (gametes) present in the ancestry of the sample at the time point in question. The number of coalescence events that will occur in that mutational class is then found using a random binomial devi-

ate with parameters $1/(2Nf_k)$ and $L_k(L_k - 1)/2$. When the population size N is large, we can, to a good approximation, neglect the possibility that the number of events within a given class is greater than one, although simultaneous coalescent events involving different gamete classes are permitted. The nodes that will undergo coalescence are picked randomly from those present in the given generation and mutational class, and the coalescence event(s) performed as described above.

Two thousand samples generated for each set of parameters were subjected to the same calculations as described above for the neutral model. To examine whether the effect of background selection on genetic diversity can be detected in samples taken from populations, the simulation outputs were sub-

jected to power analysis. The distributions of statistics in the samples without background selection were used to find the critical values such that deviations exceeding these values would occur at less than given defined levels of significance, for instance a frequency of 0.05 (one-sided). From the distributions of the statistics in samples with background selection operating, we could then determine the proportions of the samples of given size that had deviations in the direction of interest that were greater than the corresponding critical values for neutral samples of the same size.

In the case of WATTERSON's homozygosity statistic, F , the distribution of sample F values for a given number of allelic classes, m , can be computed directly (WATTERSON 1977, 1978), and a FORTRAN computer program was written to do this. This procedure for testing the departure of allele class configurations from neutrality was used rather than SLATKIN's proposal that comparisons of samples of given size and number of alleles with the fraction expected under neutrality can be used for such a test for departure of a sample from neutrality (SLATKIN 1994). SLATKIN's method is not designed to test any particular type of departure from neutrality. In the present case, however, where one is interested in testing for an excess of rare alleles caused by hitchhiking or background selection, it seems preferable to make the more conventional comparison with the fraction of samples having F values in excess of certain critical values.

RESULTS

Infinite sites population model: The approximate analytical results generated previously suggested that, at least with low mutation rates, no difference in the effect of background selection on the two different measures of genetic diversity (π and s_n) would be expected in very large populations, although s_n is affected less than π when population size is small (CHARLESWORTH *et al.* 1993). Using the pseudo-sampling methods outlined above, we have now run simulations with much larger population sizes than before.

Table 1 shows some results assuming multiplicative fitnesses, for $sh = 0.02$ and with mutation rates of 0.1, 0.05, 0.025. The table shows the observed times to loss and fixation and the expected values based on standard neutral theory (KIMURA 1983). The expected fixation times are also shown after correction for the expected effect of background selection, obtained by multiplying the effective population size by the frequency of mutation-free gametes, $f_0 = \exp - U/(2sh)$ (CHARLESWORTH *et al.* 1993). The effects of background selection on the two measures of genetic diversity are also shown, in terms of their values relative to the neutral expectations for the same population sizes and mutation rates.

For population sizes above 3200, the table shows that, for all mutation rates, the mean time to fixation and π agree closely with those predicted theoretically, as was also found when populations were simulated in full in our previous study. But the mean time to loss and s_n are expected to converge very slowly on their asymptotic values, and even population sizes as high as 10^6 are not expected to have reached their limiting values for these statistics (CHARLESWORTH *et al.* 1993). This slow conver-

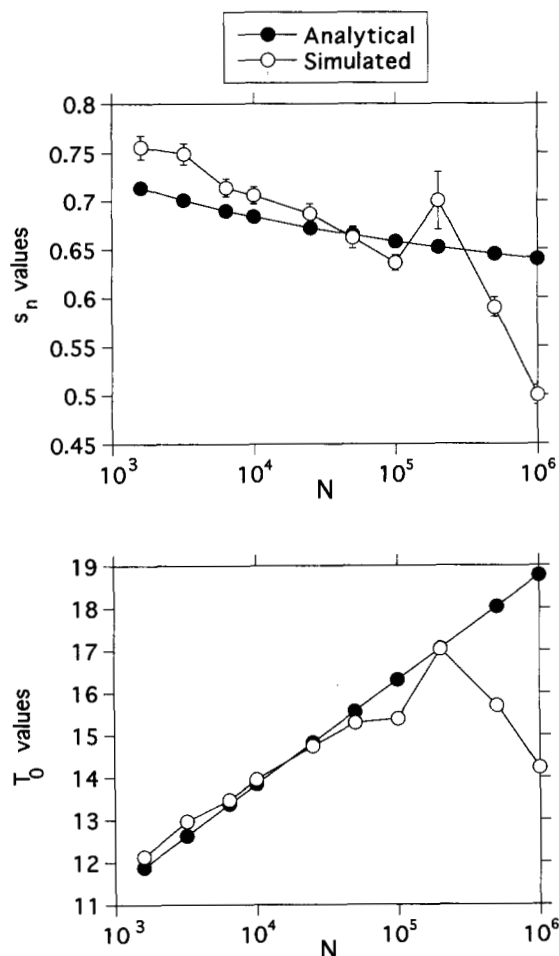


FIGURE 1.—Mean values of s_n and time to loss (T_0) from sets of simulations (\circ), compared with the analytical predicted values (\bullet) for a U value of 0.025 ($sh = 0.02$). SEs are shown for the s_n values but not for the T_0 values as they are too small to be visible.

gence can be seen in Table 1. For the higher mutation rates, it is clear that even populations of over a million individuals have substantial departures of allele frequency distributions from those obtained under a purely neutral process. Figure 1 shows the relation between s_n and N predicted by equation 10 of the Appendix of CHARLESWORTH *et al.* (1993) for the case of $U = 0.025$, together with the simulation results. The two agree quite well for $N > 3200$, but the simulated values converge on f_0 considerably faster than predicted. A similar pattern is found for the mean time to loss.

With lower sh values, the s_n values were larger for a given value of f_0 than shown in Table 1. For instance, with sh equal to 0.002, instead of 0.02, the values were 0.491 with $U = 0.1$, and 0.592 with $U = 0.05$. This is expected, since s_n is an increasing function of the mean time to loss of gametes with deleterious mutations (CHARLESWORTH *et al.* 1993), which will be closer to the neutral value if selection is very weak, whereas π is determined largely by f_0 . In the absence of recombination, background selection thus produces a greater de-

TABLE 2
Mean values and standard errors of the sample statistics

Sample size	Background selection $U = 0.01$, expected $f_0 = 0.779$				Background selection $U = 0.1$, expected $f_0 = 0.082$			
	k	θ	D_T	D_F	k	θ	D_T	D_F
$M = 1$								
20	0.719	0.753	-0.084	-0.147	0.086	0.093	-0.091	-0.185
	0.014	0.012	0.023	0.024	0.004	0.004	0.051	0.053
50	0.564	0.731	-0.084	-0.185	0.086	0.094	-0.140	-0.202
	0.013	0.011	0.027	0.030	0.004	0.003	0.038	0.042
100	0.740	0.788	-0.180	-0.396	0.091	0.107	-0.204	-0.483
	0.015	0.010	0.023	0.027	0.004	0.003	0.036	0.044
$M = 10$								
20	7.85	7.84	-0.0900	-0.314	0.818	0.864	-0.170	-0.284
	0.094	0.067	0.021	0.022	0.015	0.012	0.021	0.023
50	7.47	7.56	-0.132	-0.193	0.691	0.791	-0.293	-0.461
	0.083	0.052	0.019	0.022	0.013	0.010	0.021	0.025
100	7.17	7.68	-0.290	-1.118	0.880	0.943	-0.163	-0.266
	0.084	0.050	0.021	0.027	0.014	0.010	0.022	0.024

The table shows results from 2000 replicate runs with two different mutation rates for selected loci in the genetic background. Calculations of the D statistics include only samples that contained some genetic variation.

departure of the allele frequency spectrum from the neutral pattern if selection against deleterious alleles is weak than if it is strong, provided that the mutant alleles are maintained close to their deterministic equilibrium frequencies.

Coalescence model: The above results apply to the parameters of the population as a whole. It is possible that the effects of background selection could be too small to be easily detectable in samples from the population. Simulations based on the coalescent process were therefore done to ask two questions. First, can the effects of background selection on diversity measures be detected in samples of reasonable size (as opposed to the entire population)? Second, are significant values of test statistics for departures of allele frequency spectra from neutral expectation likely to be detected?

In the simulations, a mutation rate to deleterious mutations (U) of 0.1 or 0.01 per diploid genome per generation, multiplicative fitnesses with $sh = 0.02$, and a population size of $N = 25,000$ were assumed for most of the runs. From the results described above, this population size implies the existence of larger departures of frequency spectra from neutral expectation than is likely for the much larger effective sizes characteristic of most natural populations used for studies of DNA variability. Two different M values were used, $M = 1$ and $M = 10$, representing the range from low to high values for empirical studies. The quantities of interest were generated as described in the previous section.

The distributions of the statistics of interest for samples of various numbers of genomes were calculated from runs both with and without background selection. In the absence of selection, the approximate simulation method used here produces results that are similar to

those obtained by TAJIMA (1989) and by FU and LI (1993) for the same parameter values. The mean values of k and S_n are close to M , as expected (WATTERSON 1975; TAJIMA 1989). The mean values of the D statistics in small samples are consistently negative, as found by TAJIMA (1989) and FU and LI (1993).

To make comparisons of the D statistics with samples simulated with background selection, neutral simulations with values of M approximately equal to the values expected in the presence of selection at linked loci were run (*i.e.*, M was discounted by a factor of f_0 , representing the expected reduction in nucleotide diversity due to background selection) (CHARLESWORTH *et al.* 1993). As the true parameter values are not known for real samples, it is natural to compare empirical data with theoretical expectations that match the observed amounts of diversity. We should therefore make similar comparisons when we examine the power of statistical tests. Thus, for the higher mutation rate of $U = 0.1$, we compared the results with neutral runs with M reduced to $\sim 8\%$ of the values with selection, *i.e.*, for $M = 10$ we compared with neutral runs done with $M = 0.8$ (for $n = 100$) or $M = 1$ (for smaller samples with $n = 20$ or 50). The M values were adjusted by using a lower mutation rate for the neutral runs.

Tables 2–5 show the results of simulations with background selection. It is evident from the mean values of the sample statistics displayed in Table 2 that background selection had the expected effect on the genetic diversity measures, in that the means of both k and θ were reduced to the degree expected compared with neutrality with the same M values (CHARLESWORTH *et al.* 1993). The distributional properties of k and θ presented in Table 3 show that there is considerable power

TABLE 3

Power tests for samples of various sizes for values of k and θ , with two M values, and two different mutation rates for selected loci in the genetic background

Sample size	P value	$M = 1$		$M = 10$	
		k	θ	k	θ
$U = 0.01$					
20	0.01	—	—	0.021	0.009
	0.025	—	—	0.060	0.054
	0.05	—	—	0.089	0.119
	0.1	0.097	0.097	0.169	0.235
50	0.01	—	—	0.022	0.046
	0.025	0.030	0.030	0.077	0.123
	0.05	0.067	0.030	0.145	0.216
	0.1	0.127	0.149	0.232	0.331
100	0.01	—	—	0.083	0.063
	0.025	0.046	0.021	0.143	0.096
	0.05	0.124	0.021	0.181	0.202
	0.1	0.217	0.126	0.299	0.289
$U = 0.1$					
20	0.01	—	—	0.959	1.000
	0.025	—	—	0.993	1.000
	0.05	—	—	0.997	1.000
	0.1	0.579	0.579	0.974	1.000
50	0.01	—	—	0.974	1.000
	0.025	0.658	0.658	0.991	1.000
	0.05	0.743	0.658	0.996	1.000
	0.1	0.822	0.933	1.000	1.000
100	0.01	—	—	0.963	1.000
	0.025	0.705	0.593	0.977	1.000
	0.05	0.791	0.593	0.994	1.000
	0.1	0.835	0.886	0.998	1.000

The table shows percentages of runs that yielded lower values than the critical points for the same M values and no background selection. Only samples that contained some genetic variation were included.

to detect the effect of background selection on genetic variability, even with samples as small as 20, at least with high M values. When Mf_0 is low because of strong background selection, there is a high probability that no diversity will be present in a sample. There is no great difference between the performance of k and θ as diversity measures, despite the smaller variance of θ under neutrality (WATTERSON 1975; TAJIMA 1989). When a given critical point for the neutral case is a zero (no genetic variability), it is impossible for background selection to produce a lower value, so the frequencies of these cases cannot be given and are indicated in the table by dashes.

These conclusions do not imply that the effect of background selection on the allele frequency spectrum will be easily detectable. Tables 3 and 4 show this clearly. The D statistics were affected in the qualitatively expected way, *i.e.*, their means are all negative and larger in magnitude than in the comparable neutral cases. However, when we compare the distributions of the D statistics with the results of neutral simulations with

TABLE 4

Power tests for samples of various sizes for TAJIMA's and FU and LI's statistics (D_T and D_F , respectively)

Sample size	P value	$U = 0.01$		$U = 0.1$	
		D_T	D_F	D_T	D_F
20	0.01	0.018	0.015	0.009	0.005
	0.025	0.037	0.048	0.019	0.032
	0.05	0.082	0.070	0.051	0.074
	0.1	0.138	0.117	0.130	0.101
50	0.01	0.015	0.009	0.012	0.008
	0.025	0.028	0.018	0.035	0.052
	0.05	0.058	0.059	0.078	0.091
	0.1	0.113	0.128	0.160	0.157
100	0.01	0.044	0.092	0.043	0.068
	0.025	0.082	0.112	0.081	0.139
	0.05	0.116	0.289	0.135	0.217
	0.1	0.186	0.400	0.230	0.327

The table shows percentages of runs with $M = 10$ that yielded samples with variation that gave values than more extreme than the critical points for the case of no background selection. Only samples that contained some genetic variation were included. For the lower U values, comparisons were made with neutral runs with $M = 8$. For the higher U , comparisons were made with neutral runs with $M = 1$, which is close to the value of 0.8 expected under background selection with this mutation rate.

matching levels of diversity, their power to detect departures from neutrality in samples appears very limited (Table 4), even when the allele frequency spectrum in the population is strongly affected (Table 3). Even in sets of samples that exhibited extreme effects on the distributions of the diversity measures (*e.g.*, the case of strong background selection with $U = 0.1$ and high M , see Table 3) and with samples as large as 100, the proportion of samples that yielded test values more extreme than the critical points in the matched comparisons was less than three times as high as for the samples sets from neutral runs. The standard errors of the statistics for many of these cases were often lower than for the corresponding neutral cases, indicating that the shape of the distributions is changed by background selection, such that tails of the distributions are less extreme than with neutrality. Such an effect clearly means that comparisons with the critical points from the distributions obtained by neutral simulations may not be of much value in testing for the presence of background selection.

Table 4 shows only the samples generated with high M values, because if significant effects on the frequency spectra are undetectable with such samples, it is impossible for samples with lower diversity to yield detectable effects, even if they are very large. This was borne out by runs with $M = 1$ with the higher mutation rate used in Tables 2 and 5, and with $n = 100$. For background selection with $U = 0.1$, significant values of both TAJIMA's D and FU and LI's D statistics were found in fre-

TABLE 5

Results of WATTERSON's tests with background selection assuming a mutation rate, U , of 0.1

No. of alleles	Fraction of samples with F values greater than the critical point		
	$M = 10$ $n = 50$	$M = 10$ $n = 100$	$M = 100$ $n = 100$
2	205 0.3951	83 0.4458	3 —
3	387 0.1344	175 0.1543	7 —
4	459 0.1242	341 0.1584	19 0.3158
5	414 0.1377	362 0.1492	19 0.2125
6	254 0.1220	391 0.1381	137 0.1971
7	142 0.1761	299 0.1505	216 0.1528
Overall distribution	1861 0.1628	1651 0.1641	462 0.1818

Results are from 2000 samples of size $n = 50$ or 100 with various M values. The results are shown for the $P = 0.05$ level only, for each sample constitution with respect to the number of alleles given at the top of each column of frequencies. Cases when the number of samples with a given number of alleles was too low to perform the test are indicated by—. The results when all allele numbers were combined are also shown.

quencies that were less than or equal to the expected frequencies under neutrality for all four critical points examined. This is because a large number of samples lack variability, so that the tests cannot be applied. For $U = 0.01$, the frequencies were very slightly above the expected frequencies under neutrality (a maximum of 23% greater).

Runs with $M > 10$ gave higher frequencies of samples outside the critical values. For instance, with $M = 100$ and a mutation rate of 0.1 in the genetic background, comparisons with critical values derived from neutral runs with $M = 8$ and samples of 100 alleles yielded 38% of samples beyond the 1% point for F_U and L_I 's statistic, and 66% beyond the 5% point. The frequencies were lower for the T_{AJIMA} test (10 and 23% for the 1 and 5% points, respectively). Very large M values in the absence of background selection were, however, needed to obtain this statistical power. With $M = 20$, the frequencies were similar to those with $M = 10$.

As described in METHODS, we also studied the performance of WATTERSON's homozygosity test (WATTERSON 1977, 1978). As the distribution of values of WATTERSON's test depends on the number of alleles in the sample, we did the tests separately for the sets of samples with different numbers of alleles but also tabulated the results over all numbers of alleles for which reason-

ably large numbers of samples were obtained. Table 5 shows the results from samples with $U = 0.1$, for comparison with the results given in Table 4 for the other two tests. With $M = 10$ and $n = 50$ or 100, the fractions of samples with higher values than the critical points are increased by a factor of two to three, compared with the neutral expectations. For instance, with $n = 100$, 15% of samples with seven alleles, and 18% of all samples, exceeded the 5% critical values. Table 5 shows that samples with small numbers of alleles are not very useful, as might be expected because the different critical points tend to be very similar to one another with small allele numbers. In general, even for samples with five or more allelic types, this test had low power, comparable with that of T_{AJIMA} 's test, to detect the effect of background selection on the frequency spectrum but, unlike T_{AJIMA} 's and F_U and L_I 's tests, there was no evident improvement in power in simulations with very high M values.

Effect of changes in the selection regime: It is important to consider the effect of the selection regime on the properties of samples from populations. The sh values used above were chosen because they correspond to the estimate of the harmonic mean effect of a heterozygous detrimental mutation on the fitness of *Drosophila* in nature (CROW and SIMMONS 1983). Potential effects in other species are, of course, also of interest, but at present few data are available either for estimating the relevant parameters or for testing predicted diversity values. In any species, it is likely that there is a wide distribution of selection coefficients around the mean, possibly with a long tail of rather weakly selected mutations (KEIGHTLEY 1994). Since the simulations described above indicated that weaker selection tends to cause a larger deviation of the allele frequency spectrum from neutral expectation, there is thus a possibility that the results for large sh may underestimate the power of tests for deviations from neutrality.

We studied the effect of the magnitude of sh by similar analyses to those just described, assuming $M = 10$. We chose a low value of f_0 (0.08) to produce strong background selection and studied samples of size 100, so that the samples generated were ones that would be most likely to yield statistically significant statistic values. Sets of runs with various selection coefficients but the same f_0 were done by changing both sh and the mutation rate. The results are displayed in Figure 2. As expected, lower sh values generated higher frequencies of significant test statistics, though the frequencies of values significant at the 5% level remained modest even when the sh value was assumed to be 10 times lower than the value of 0.02 used above (Figure 2). At most, about one-third of the tests were found to be significant at the 5% level, even with sh as small as 0.0002, if the population size was 500,000 or more. Lower M values and smaller sample sizes will be expected to give fewer significant cases.

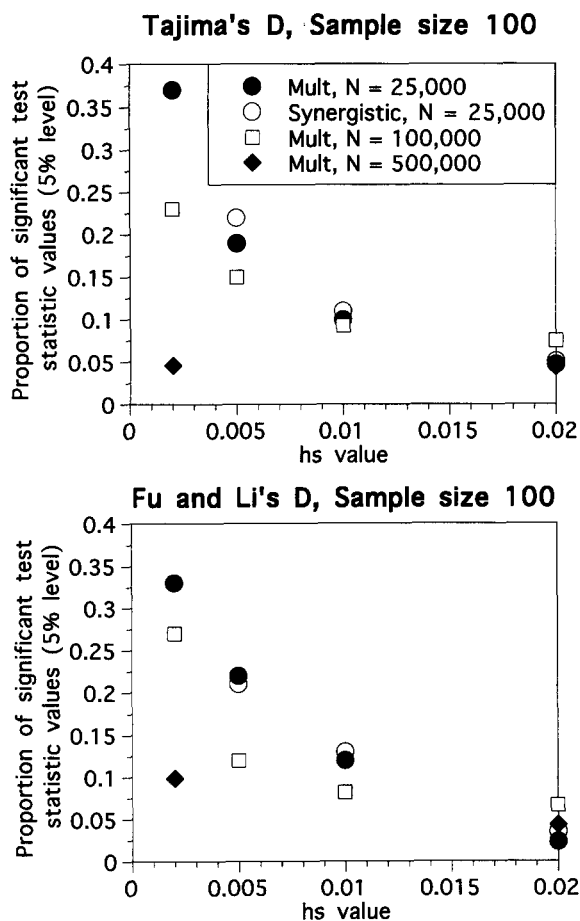


FIGURE 2.—Fractions of TAJIMA's D and FU and LI's D values that were significant at the 5% level, out of sets of 2000 samples of size 100 simulated with background selection. The mutation rate U to deleterious alleles and the product sh were chosen such that the value of f_0 was 0.08 in all cases. The D values were compared with their distributions in sets of neutral samples simulated with the same selection model and population size, but with $M = 0.8$ so that diversity was similar to that in the samples with background selection. Populations with $N = 25,000$, 100,000 and 500,000 were simulated under the assumption of multiplicative selection (denoted by Mult in the figure), and $N = 25,000$ was assumed for the runs with synergistic selection. The α and β values for synergistic selection were as follows: for $hs \approx 0.02$, α was 0.08, β was 0.032, and U was 0.1; for $hs \approx 0.01$, α was 0.004, β was 0.0016, and U was 0.05; for $hs \approx 0.005$, α was 0.002, $\beta \approx 0.0008$, and U was 0.025.

As the results from the infinite sites model (see above) showed that the effect of background selection on the allele frequency spectrum tends to decrease with increasing population size, we also studied the effect of changing the population size. With a population of $N = 100,000$ or 500,000, instead of 25,000 as above, the frequencies of significant results were lower for each sh value studied (Figure 2). This conclusion was true even when both N and sh were changed so as to preserve the same value of Nsh . With either $Nsh = 50$ ($N = 25,000$ and $sh = 0.002$ vs. $N = 100,000$ and $sh = 0.0005$) or $Nsh = 500$ ($N = 25,000$ and $sh = 0.02$ vs. $N = 100,000$

and $sh = 0.005$), the frequency of significant tests was lower with the lower selection, showing that the effect of sh was more important than that of population size.

There has been much discussion of the evolutionary role of synergistic fitness interactions among deleterious mutations (KIMURA and MARYAMA 1966; CROW 1970; KONDRASHOV 1988; CHARLESWORTH 1990). For this reason, we have also studied samples generated assuming synergistic interactions between the loci under selection (see METHODS). For the three strengths of selection studied with synergism (equivalent to sh values of approximately 0.02, 0.01 and 0.0052), the frequencies of significant results were very similar to those for the comparable cases with multiplicative selection (Figure 2). This suggests that we can have confidence in the conclusions obtained by assuming multiplicative fitnesses.

DISCUSSION

Our results for the case of no recombination, using the ITO approximate method of stochastic simulation, confirm our earlier conclusion that background selection causes an excess of rare variants over neutral expectation in populations of small to moderate size, and show that the difference between the behavior of π and s_n diminishes as population size becomes very large (Table 1 and Figure 1). Figure 1 shows that that the ratio of s_n to the classical neutral values approaches its asymptotic value of $f_0 = 0.535$ faster than expected from the analytical result. As the population size increases, so that selection becomes more effective, the mean time to loss (which is the major influence on s_n) is increasingly overestimated by the analytic approximation. This is not unexpected, since this formula ignores the possibility that a gamete carrying a single deleterious mutation acquires a further mutation (and so is selected against more strongly) before elimination from the population (CHARLESWORTH *et al.* 1993).

Convergence of the ratio of s_n to its neutral expectation is, however, very slow for $U = 0.1$ or 0.05 (Table 1). $U = 0.1$ is probably at the upper limit possible for the centromeric regions of *D. melanogaster* autosomes (Table 4 of CHARLESWORTH *et al.* 1993) and is unlikely to be exceeded for the centromeric regions of other species. Such regions represent the largest sections of the genome that have near-zero rates of genetic recombination and, hence, offer the best opportunity for studying the effects of restricted recombination on genetic variability. In *D. melanogaster*, the level of genetic diversity in the *Adh* region of the genome suggests that the effective population size of the species is of the order of one million (KREITMAN 1983). Our results therefore suggest that there will be an appreciable excess of rare variants at segregating neutral sites in the centromeric regions of the major autosomes of this species, if background selection is operating at the ex-

pected intensity (although this may not be detectable in samples by tests such as TAJIMA's and FU and LI's tests; see below).

With samples of realistic sizes, the simulations of the coalescent process with selection show that the effect on the level of genetic diversity should be detectable (Table 3). This is in accord with the only currently available data from *Drosophila*, which indicate reduced variability at the DNA level at loci in regions of reduced recombination (see below). The tests of departure from neutrality studied here offer (in principle) the possibility of determining whether observed instances of reduced genetic diversity could be explained by background selection. But significant departures of allele frequency spectra from expectations under neutrality (as assessed by the values of TAJIMA's or FU and LI's statistics, or by Watterson's test) are unlikely to be found under background selection with the "standard" sh value of 0.02 that corresponds to the estimated harmonic mean selection coefficient against a heterozygous detrimental mutation in *Drosophila* (MUKAI 1964; MUKAI *et al.* 1972; CROW and SIMMONS 1983). This remains true even when the true population size is small, so that the π and θ values (relative to neutral expectation) for the population are expected to differ widely. This conclusion is based on our simulations employing populations of size 25,000. In real populations of *Drosophila*, which appear to have very large effective population sizes (KREITMAN 1983), it is even less likely that one will detect an effect. Even a fourfold increase in population size yields samples with much lower frequencies of statistically significant D values, for a given value of sh (see above). The low power of TAJIMA's D statistic for detecting the effect of background selection has also been noted by HUDSON and KAPLAN (1994).

A limitation of our results is that they are based on the unrealistic assumption of identical selection coefficients for all selected mutations. This restrictive assumption enables us to use the technically very convenient coalescence method. It remains to be seen how our conclusions might be changed if samples were generated with mutations at the selected loci drawn from some biologically plausible distribution of heterozygous selection coefficients. Since, however, even selection one hundred times weaker than the current best estimate of the mean value yields only modest frequencies of significant test statistics if effective population size is large (Figure 2), our conclusion appears robust; reduction of genetic diversity due to background selection seems unlikely to be frequently associated with significant values of any of the test statistics examined here, in contrast to what is found with hitchhiking (see below).

A difficulty that is inherent in the nature of the empirical data is that strong effects of background selection, which are most likely to cause large departures of allele frequency spectra from neutral expectation, tend to result in a high frequency of samples with little or no

variability, so that the ability to test for such departures is very limited (see Table 3). Significant differences from the neutral expectation were seen with appreciable frequency only when we modeled very high diversity ($M = 100$) and strong background selection ($U = 0.1$). In this set, FU and LI's statistic yielded more than twice the frequency of significant results than TAJIMA's test. A similar superiority of FU and LI's test is also seen in many of the results given in detail in Table 4, though the difference is usually not as large. These findings suggest that TAJIMA's test is not as good as FU and LI's test at detecting the effects of background selection. It is important to note, however, that this comparison is based only on our simulations, in which the phylogenetic trees are completely known. With real data sets, it is necessary to perform this test to construct trees from the data, or from other sources of information, so the error inherent in this process might outweigh this benefit. There is also no reason from our results alone to believe that FU and LI's test must be superior as a test for other departures from neutrality, such as hitchhiking, for example. It is therefore probably best to use both tests on real data sets. Our results also suggest that WATTERSON's test is not appreciably better at detecting the effects of background selection than TAJIMA's test (Tables 4 and 5). As WATTERSON's test is considerably more troublesome to perform, it does not seem to be a satisfactory means of testing for the causes of lowered genetic diversity.

These results imply that data on very long sequences would be necessary to detect the effects of background selection on allele frequency spectra. A diversity value per nucleotide site of 0.005, a typical value for a gene in a freely recombining section of the *D. melanogaster* genome (LANGLEY 1990), is equivalent to a sequence of length $M/0.005$ or 20,000 bases for $M = 100$. Surveys of such large tracts are unlikely to be possible in studies that use sequencing as the only methodology. Methods such as single-strand conformation polymorphism, which can survey large tracts of DNA (*e.g.*, AGUADÉ *et al.* 1994), may enable such data to be obtained.

However, to distinguish between background selection and an alternative such as hitchhiking, it is only necessary that one of the processes considered yields significant test results while the other does not. The fact that significant values of the test statistics for departure from neutrality are not expected in most realistic sets of data thus does not make discrimination between alternative models impossible. Simulation studies and power tests, such as we have performed for the background selection model, have also been carried for the hitchhiking model (BRAVERMAN *et al.* 1995; SIMONSEN *et al.* 1995). Extreme hitchhiking events, such that a single haplotype is fixed, should reduce π to a greater extent than θ (see discussions in CHARLESWORTH *et al.* 1993 and in BRAVERMAN *et al.* 1995). This is because, as new diversity begins to accumulate after the event that

TABLE 6
Results of TAJIMA's tests for 50 loci studied in *Drosophila melanogaster*
population surveys of DNA level diversity

Recombination	No. of values	π	θ	No. of values	TAJIMA's d (scaled)
All loci and populations treated individually					
<i>L</i>	35	1.30 \pm 0.255	1.37 \pm 0.306	32	-0.238 \pm 0.128
<i>H</i>	47	5.39 \pm 0.546	5.11 \pm 0.618	40	-0.116 \pm 0.053
Significance		$P < 0.001$	$P < 0.001$		$P = 0.95$
One population per locus (U.S. populations where possible)					
<i>L</i>	14	1.69 \pm 0.447	2.04 \pm 0.650	13	-0.167 \pm 0.188
<i>H</i>	28	6.03 \pm 0.809	6.55 \pm 0.752	28	-0.080 \pm 0.048
Significance		$P < 0.001$	$P < 0.001$		$P = 0.37$
Using π and θ values averaged for each locus (Zimbabwe population considered separately)					
<i>L</i>	18	1.68 \pm 0.379	1.77 \pm 0.433	18	-0.052 \pm 0.184
<i>H</i>	32	5.95 \pm 0.706	6.35 \pm 0.719	32	-0.063 \pm 0.069
Significance		$P < 0.001$	$P < 0.001$		$P = 0.97$

The table shows estimates of variability in terms of the two standard measures (π and θ), for regions of low recombination (*L*) compared with other chromosomal regions (*H*). All values are multiplied by 10^3 . Values of TAJIMA's statistic are shown in terms of unstandardized d values ($\pi - \theta$), scaled as described in the text. SEs of the statistics are also given, and the levels of significance of the differences between results from the *L* and *H* recombination regions were tested by MANN-WHITNEY U tests. Note that numbers are lower for the TAJIMA statistics than for the diversity measures, because for some loci no diversity was detected, so that the statistic could not be calculated. The initial data source was the compilation of KREITMAN and WAYNE (1994). For sequencing studies published since that compilation, this was supplemented by data for silent and noncoding (intron and flanking) sites where available (LEICHT *et al.* 1994; SIMMONS *et al.* 1994) or with unpublished data for silent sites in coding regions compiled by E. N. MORIYAMA and J. R. POWELL. We also included the SSCP study of AGUADÉ *et al.* (1994) and additional unpublished data supplied by C. F. AQUADRO, J. HEY, H. HILTON, E. C. KINDAHL, J. PRITCHARD and S. W. SCHAEFFER.

destroys diversity, the number of alleles should recover faster than π , and this should produce negative values of TAJIMA's D , or of F_U and L_i 's D . In contrast to our results for background selection, significant values of TAJIMA's and F_U and L_i 's test statistics in samples of realistic size occurred in a high proportion (>60% of one-tailed tests) of simulations in which hitchhiking caused greatly reduced diversity values (BRAVERMAN *et al.* 1995; SIMONSEN *et al.* 1995). WATTERSON's test performed about as well as the other tests.

BRAVERMAN *et al.* (1995) compared their simulation results with five studies of molecular variation in *D. melanogaster*, where the loci surveyed were in regions of very low recombination and had substantially reduced levels of genetic diversity compared with typical values, and where there were enough segregating sites in the samples that tests of neutrality compared with the hitchhiking alternative had reasonable power. Only one of the cited studies in which TAJIMA's D was calculated yielded a significant value, while diversity was reduced to ~10% of the standard value for *D. melanogaster*. Hitchhiking of the intensity required to explain this reduction in diversity would be very likely to produce significant D values. These authors therefore concluded that severe hitchhiking alone cannot explain the low

genetic diversity observed. In contrast, our power analyses indicate that significant D values are unlikely to be detected if background selection had caused the low genetic diversity at these loci. In the published literature to date (see Table 6), three out of 51 values of TAJIMA's standardized D value significant at the 5% level have been found (two for the *y-ac-sc* region, and one for *Adh-dup*).

In addition to examining the results from DNA variability surveys for the presence of statistically significant TAJIMA's test results, it is also interesting to ask whether large negative values tend to be found for loci in regions of restricted recombination, as should be the case if hitchhiking were the major cause of low genetic variability in these regions. We therefore conducted a review of the published literature on DNA sequence variation in *D. melanogaster*, the only relevant data sets available to date, to ask whether samples of alleles from loci in regions of reduced recombination have excess rare alleles more than do samples from loci in freely recombining chromosomal regions. Unfortunately, values of TAJIMA's standardized D statistic are only rarely reported; the presentation of published data is highly heterogeneous, often not permitting its calculation. Furthermore, different populations are often pooled,

which may inflate estimates of variability, and produce discrepancies between π and θ (see STROBECK 1987, pp. 151–152). Taking the reported data at face value, it is, however, possible to compare π and θ values, *i.e.*, to calculate the unstandardized TAJIMA's d , whose expectation is zero under neutrality. We did this for all studies that we could find, whether based on restriction enzyme or SSCP surveys of loci and their surrounding sequences, or on sequencing studies of individual loci. Where both restriction and sequencing studies were available for the same locus in a given population, we used the restriction enzyme data, as these generally had larger sample sizes and usually include noncoding sequences.

Out of 82 pairs of values, including 50 loci, we classified 35 (18 loci) as coming from regions with low recombination (loci in *X* chromosomal bands 1–3A4, bands 21, 38–43E, and 59E–60 of chromosome 2, and bands 61–62A, 78–84 and 100 of chromosome 3), and 47 as representing regions with more recombination (see LANGLEY *et al.* 1988). Because selection events such as hitchhiking might occur within one population but not another, it is most appropriate to calculate TAJIMA's d values for data from single populations. For measures of diversity, however, it may be preferable to calculate estimates from data combined from as many populations as possible. We therefore treated the data in several different ways (see Table 6).

Genetic variability was significantly reduced in the low recombination regions as defined above, for both π and θ values, averaging 0.24–0.32 of the values estimated for the other loci, depending on how the data were treated (Table 6). Because differences between π and θ , from which TAJIMA's d values are calculated, will tend to be smaller when the diversity measures are themselves small, *i.e.*, in the regions with low recombination, we scaled each value by dividing by the mean θ value for the loci in its set (data from either low recombination regions or from loci in the other regions) and compared the scaled d values. Unlike the diversity measures, there was no significant difference in d between the two recombinational environments, the values being slightly negative for both sets of loci with a slight tendency for the samples from loci in low recombination regions to have larger negative values (Table 6). The data thus suggest that there may be a slight overall excess of rare alleles, compared with the neutral expectation, due either to selection on the sites studied or on linked sites. The departure is, however, significant only in one of the six tests (for the *H* loci in the upper set of results in Table 6), and this result is rendered doubtful by the evident lack of independence between the different data sets for the same loci.

The data thus provide no clear evidence favoring hitchhiking but appear consistent with the reduction of variability in low recombination regions having been caused by background selection. The d values from

these studies were similar to values resulting from our simulations, when scaled in the same manner. For the samples of size 100, for instance, with the parameter values of Table 2 and either high or low mutation rates to deleterious alleles, the d value scaled in the same way was -0.066 . These results cannot, however, be considered as strong evidence in favor of background selection as the sole cause of reduced genetic variation, relative to neutral expectation, as other possibilities that preserve the neutral pattern of alleles frequencies may also exist. GILLESPIE (1994) has, for instance, shown that a model of fluctuating selection, involving the ultimate fixation or loss of weakly selected alleles, can also lead to reduced genetic diversity without generating a large difference between the effect on π and θ , so that significantly negative D values would presumably not be found in samples. Thus, the observation of reduced variability without significant departures of allele frequencies from neutral expectation does not necessarily mean that background selection is the force involved.

In our previous study of the effects of background selection (CHARLESWORTH *et al.* 1993), we concluded that this process alone was not a plausible explanation for the great reduction in genetic diversity in some regions of the *D. melanogaster* genome, such as the tip of the *X* chromosome and the fourth chromosome. The reason was that it is unlikely that rates of deleterious mutation in such small regions of the genome would be high enough to make background selection a powerful force. This conclusion should now, however, be qualified. Good approximate methods for large populations have recently been developed that enable modeling of the effects of background selection in the presence of genetic recombination (R. R. HUDSON and N. L. KAPLAN 1995; M. NORDBORG, B. CHARLESWORTH and D. CHARLESWORTH, unpublished results). These methods can be used to predict levels of genetic diversity in different regions of the *Drosophila* genome, for which some of the relevant parameters can be estimated, and data on reduced variability in the centromeric regions of chromosomes can be fitted reasonably well (R. R. HUDSON and N. L. KAPLAN, unpublished results; B. CHARLESWORTH and R. R. HUDSON, unpublished results). The data on reduced variability at loci on the tip of the *X* and the fourth chromosome can also be explained if the effects of transposable element insertions are taken into account (B. CHARLESWORTH and R. R. HUDSON, unpublished results). As noted by HUDSON (1994), transposable elements may play an important role in background selection, since the available evidence suggests that such elements are usually mildly deleterious in their effects on fitness and are distributed over many sites in the genome (CHARLESWORTH *et al.* 1994). However, the strength of selection against transposable elements that are found segregating in natural populations is probably very weak (*sh* is estimated to be $\sim 2 \times 10^{-4}$) (CHARLESWORTH *et*

al. 1992). A recombination rate of the order of the strength of selection between a selected locus and a neutral site largely removes the effect of background selection (HUDSON and KAPLAN 1994; M. NORDBORG, B. CHARLESWORTH and D. CHARLESWORTH, unpublished results), so that transposable elements can have a major effect only in regions where recombination is highly suppressed (B. CHARLESWORTH and R. R. HUDSON, unpublished results). Since transposable elements are so weakly selected, their effects on background selection are more likely to be associated with negative *D* statistics than are those of mutations with larger effects on fitness, so that negative *D* statistics might sometimes be found for genes located in regions of highly reduced exchange (see Figure 2).

All the results presented here deal with random mating populations. It will also be valuable to study background selection in selfing populations, as high levels of self-fertilization are expected to greatly reduce genetic diversity (CHARLESWORTH *et al.* 1993). Since empirical studies to quantify genetic diversity in plants with high selfing rates are now being done [as well as comparisons with related outcrossing species, see FENSTER and RITLAND (1992) and HANFSTINGL *et al.* (1995)], it is important to know what diversity properties should be examined and which tests should be performed on the data to gain the most insight from the results.

We thank the following for permission to cite their unpublished work: C. F. AQUADRO, N. H. BARTON, D. J. BEGUN, J. M. BRAVERMAN, G. A. CHURCHILL, E. C. KINDAHL, J. HEY, H. HILTON, R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, E. N. MORIYAMA, J. R. POWELL, J. PRITCHARD, S. W. SCHAEFFER, K. L. SIMONSEN and W. STEPHAN. We also thank several reviewers, C. F. AQUADRO, J. M. BRAVERMAN and MAGNUS NORDBORG for comments on the manuscript, and MAGNUS NORDBORG for help with programming. This work was supported by National Science Foundation grant DEB9217683, by the Darwin Trust of Edinburgh, and the Underwood fund of the Biotechnology and Biological Sciences Research Council, U.K.

LITERATURE CITED

- AGUADÉ, M., W. MEYERS, A. D. LONG and C. H. LANGLEY, 1994 Reduced DNA sequence polymorphism in the *su(s)* and *su(wa)* regions of *Drosophila melanogaster* as revealed by SSCP and stratified DNA sequencing. *Proc. Natl. Acad. Sci. USA* **91**: 4658–4662.
- AQUADRO, C. F., D. J. BEGUN and E. C. KINDAHL, 1994 Selection, recombination, and DNA polymorphism in *Drosophila*, pp. 46–56 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING. Chapman and Hall, London.
- BARTON, N. H., 1995 Linkage and the limits to natural selection. *Genetics* **140**: 821–841.
- 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.
- BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- CHARLESWORTH, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**: 199–221.
- CHARLESWORTH, B., 1994 The effect of background selection against deleterious alleles on weakly selected, linked variants. *Genet. Res.* **63**: 213–228.
- CHARLESWORTH, B., A. LAPID and D. CANADA, 1992 The distribution of transposable elements within and between chromosomes 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, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- CHARLESWORTH, B., P. SNIEGOWSKI and W. STEPHAN, 1994 The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215–220.
- 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*, Vol. 3c, edited by H. L. CARSON, M. ASHBURNER and J. N. THOMSON. Academic Press, London.
- EWENS, W. J., 1972 The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**: 87–112.
- FENSTER, C. B., and K. RITLAND, 1992 Chloroplast DNA and isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *Amer. J. Bot.* **79**: 1440–1447.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GILLESPIE, J. H., 1994 Alternatives to the neutral theory, pp. 1–17 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING. Chapman and Hall, London.
- 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, MA.
- HANFSTINGL, U., A. BERRY, E. A. KELLOGG, J. T. COSTA, W. RÜDIGER *et al.*, 1994 Haplotype divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection *Genetics* **138**: 811–828.
- HEY, J., 1991 A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. *Theoret. Pop. Biol.* **39**: 30–48.
- HILTON, H., R. M. KLIMAN and J. HEY, 1994 Using hitchhiking genes to study adaption and divergence during speciation within the *Drosophila melanogaster* species complex. *Evolution* **48**: 1900–1913.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process. *Oxford Surveys of Evolutionary Biology* **7**: 1–45.
- HUDSON, R. R., 1994 How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? *Proc. Natl. Acad. Sci. USA* **91**: 6815–6818.
- HUDSON, R. R., and N. L. KAPLAN, 1995 Gene trees with background selection, pp. 140–153 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING. Chapman and Hall, London.
- KEIGHTLEY, P. D., 1994 The distribution of mutant effects on viability in *Drosophila melanogaster*. *Genetics* **138**: 1315–1322.
- KIMURA, M., 1971 Theoretical foundations of population genetics at the molecular level. *Theoret. Pop. 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., and N. TAKAHATA, 1983 Selective constraint in protein polymorphism: study of the effectively neutral model by using an improved pseudosampling method. *Genetics* **80**: 1048–1052.
- KONDRASHOV, A., 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**: 435–440.
- KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**: 412–417.
- KREITMAN, M., 1991 Detecting selection at the level of DNA, pp. 202–221 in *Evolution at the Molecular Level*, edited by R. K. SELANDER, A. G. CLARK and T. S. WHITTAM. Sinauer, Sunderland, MA.
- KREITMAN, M., and WAYNE, M. L. 1994 Organization of genetic variation at the molecular level: lessons from *Drosophila*, pp. 157–183 in *Molecular Ecology and Evolution: Approaches and Applications*, edited by B. SCHIERWATER, B. STREIT, G. P. WAGNER and R. DE-SALLE. Birkhäuser Verlag, Basel.

- LANGLEY, C. H., 1990 The molecular genetics of *Drosophila*, pp. 75–91 in *Population Biology of Genes and Molecules*, edited by N. TAKAHATA and J. F. CROW. Baifukan, Tokyo.
- LANGLEY, C. H., E. A. MONTGOMERY, R. H. HUDSON, N. L. KAPLAN and B. CHARLESWORTH, 1988 On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* **52**: 223–235.
- LEIGHT, B. G., S. V. MUSE, M. HANCZYC and A. G. CLARK, 1995 Constraints on sequence evolution in the gene encoding the myosin alkali light chain in *Drosophila*. *Genetics* **139**: 299–308.
- LI, W.-H., 1980 Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics* **95**: 237–258.
- MUKAI, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**: 1–19.
- 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.
- PECK, J.R., 1994 A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* **137**: 597–606.
- 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.
- SIMMONS, G. M., W. KWOK, P. MATULONIS and T. VENKATESH, 1994 Polymorphism and divergence at the *prune* locus in *Drosophila melanogaster*. *Mol. Biol. Evol.* **11**: 666–671.
- SIMONSEN, K. L., G. A. CHURCHILL and C. F. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**: 413–439.
- SLATKIN, M., 1994 An exact test for neutrality based on the Ewens sampling distribution. *Genet. Res.* **64**: 71–74.
- STROBECK, C., 1987 Average number of nucleotide differences in a sample from a single population: a test for population subdivision. *Genetics* **117**: 149–153.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis. *Genetics* **123**: 585–595.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* **7**: 256–276.
- WATTERSON, G. A., 1977 Heterosis or neutrality? *Genetics* **35**: 789–814.
- WATTERSON, G. A., 1978 The homozygosity test of neutrality. *Genetics* **88**: 405–417.
- WOLFF, K., 1991 Analysis of allozyme variability in three *Plantago* species and a comparison to morphological variability. *Theoret. Appl. Genet.* **81**: 119–126.

Communicating editor: N. TAKAHATA