

The Effects of Hill-Robertson Interference Between Weakly Selected Mutations on Patterns of Molecular Evolution and Variation

Gilean A. T. McVean and Brian Charlesworth

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

Manuscript received September 28, 1999

Accepted for publication February 17, 2000

ABSTRACT

Associations between selected alleles and the genetic backgrounds on which they are found can reduce the efficacy of selection. We consider the extent to which such interference, known as the Hill-Robertson effect, acting between weakly selected alleles, can restrict molecular adaptation and affect patterns of polymorphism and divergence. In particular, we focus on synonymous-site mutations, considering the fate of novel variants in a two-locus model and the equilibrium effects of interference with multiple loci and reversible mutation. We find that weak selection Hill-Robertson (wsHR) interference can considerably reduce adaptation, *e.g.*, codon bias, and, to a lesser extent, levels of polymorphism, particularly in regions of low recombination. Interference causes the frequency distribution of segregating sites to resemble that expected from more weakly selected mutations and also generates specific patterns of linkage disequilibrium. While the selection coefficients involved are small, the fitness consequences of wsHR interference across the genome can be considerable. We suggest that wsHR interference is an important force in the evolution of nonrecombining genomes and may explain the unexpected constancy of codon bias across species of very different census population sizes, as well as several unusual features of codon usage in *Drosophila*.

THE efficacy of selection acting simultaneously at linked sites can be considerably reduced in regions of low recombination (Fisher 1930; Muller 1964; Hill and Robertson 1966). Linkage disequilibrium between alleles at selected loci, generated by the stochastic nature of mutation and sampling in finite populations, “interferes” with the action of selection at any one locus (Felsenstein 1974). When the rate of recombination between selected sites is high, associations between alleles are rapidly broken down, but when recombination is rare or absent, interference may have a significant impact on patterns of evolution.

While the effects of interference between selected mutations on patterns of molecular evolution at those sites have largely been ignored, the effects of selection acting at linked loci on patterns of neutral evolution and variation have received considerable attention. Selective sweeps of strongly beneficial mutations (Maynard Smith and Haigh 1974) and the rapid elimination of deleterious mutations (Charlesworth *et al.* 1993) can significantly reduce neutral variation in regions of low recombination. Selection on single weakly selected mutations where the product of the selection coefficient, s , and the effective population size, N_e , is of the order of 1, in contrast, has a largely negligible effect on variation at linked neutral loci (Golding 1997; Neuhauser and Krone 1997; Przeworski *et al.* 1999).

This focus on the effects of selection acting on linked neutral variation stems from the assumption that the majority of molecular evolution proceeds through events of little or no selective importance (Kimura 1983). There is, however, considerable reason to doubt this view of molecular evolution. In particular, synonymous codon positions (those at which some or all mutations have no effect on the amino acid encoded) appear to be under selection acting at the level of translation in many bacteria (Sharp and Li 1986) and eukaryotes, including yeast (Sharp *et al.* 1986), *Drosophila* (Shields *et al.* 1988), *Caenorhabditis* (Stenico *et al.* 1994), and *Arabidopsis* (Chiapello *et al.* 1998; Duret and Mouchiroud 1999). In addition, noncoding regulatory sequences can extend over large genomic regions, and there is indirect evidence that transcribed but untranslated regions of genes can experience stronger selection pressures than synonymous sites (Li and Graur 1991, p. 73; Bauer and Aquadro 1997).

If the majority of sites in genes, and perhaps in genomes as a whole, are not completely neutral but actually experience weak selection, then the predictions of the neutral theory are not applicable to the study of DNA sequence evolution. This raises two critical questions: How does weak selection acting at a single locus affect patterns of evolution relative to those expected under neutrality? And how does interference between selected loci affect the dynamics of molecular evolution relative to that expected under independence? The answer to the first question can be achieved through the application of diffusion theory, as for the neutral case (Kimura 1983). Simple models that assume indepen-

Corresponding author: Gilean McVean, ICAPB, Ashworth Laboratories, King's Bldgs., West Mains Rd., Edinburgh EH9 3JT, United Kingdom. E-mail: g.mcvan@ed.ac.uk

dence between sites can be manipulated to make predictions about the extent of codon bias and patterns of polymorphism and divergence (Bulmer 1991; Sawyer and Hartl 1992; McVean and Charlesworth 1999). Coalescent methods, adapted to incorporate selection (Krone and Neuhauser 1997; Neuhauser and Krone 1997), have also been used to investigate genealogies associated with weakly selected sites (Przeworski *et al.* 1999).

The second problem, that of interference, has received much less attention (Hill and Robertson 1966; Felsenstein 1974; Birky and Walsh 1988; Otto and Barton 1997). Strongly beneficial mutations are unlikely to interfere with each other, as these are probably rare and spread rapidly through the population. Likewise, strongly deleterious mutations at or near mutation-selection equilibrium will also generate little mutual interference as they exist at very low frequencies (though interference is an integral feature of the mutation accumulation process known as Muller's ratchet: Felsenstein 1974). In contrast, weakly selected mutations are probably more numerous and may segregate for sufficient time periods, at sufficiently high frequencies, for interference to be important.

Interference involving weakly selected sites may be further broken down into asymmetrical interference involving weakly and strongly selected mutations and interference simply between weakly selected mutations. In cases where the selection coefficients are asymmetrical, the dynamics of evolution at weakly selected sites may well be strongly affected, but the reverse is unlikely to be true. For example, while the effects of recurrent deleterious mutation (background selection: Charlesworth *et al.* 1993) on single, weakly selected, linked variants can largely be treated as a simple reduction in N_e , just as for neutral sites (Stephan *et al.* 1999) there is no detectable effect of the weakly selected site on the dynamics of the strongly selected site, assuming $Ns \geq 10$ for this site (G. McVean, unpublished results).

The extent to which interference between weakly selected mutations alone can affect patterns of molecular evolution is not known. Previous work (Li 1987; Comeron *et al.* 1999) has suggested that such interference can substantially reduce the expected level of codon bias in regions of low recombination, but there has been no attempt to provide a general description of the effects of interference on other aspects of molecular evolution, such as patterns of polymorphism and divergence. In this article, we present the results of series of simulations that investigate the extent to which interference between weakly selected mutations is dependent on the strength of selection, the recombination rate, and the number of sites under selection. Using a simple reversible mutation model (as in the evolution of synonymous codon usage: Bulmer 1991) we assess the magnitude of the effects of interference on the level of molecular adaptation (codon bias), the rate of sequence divergence,

levels of nucleotide site diversity, and the frequency spectrum of polymorphic sites.

We also consider whether interference between weakly selected mutations may explain a number of observations concerning synonymous codon usage that do not fit the predictions of models that assume independence between sites. In particular, we discuss its implications for the remarkable similarity in levels of codon bias observed across species of very different census population sizes (Powell and Moriyama 1997), interactions between recombination rate, codon bias, and synonymous-site diversity in the genome of *Drosophila melanogaster* (Begun and Aquadro 1992; Kliman and Hey 1993; Comeron *et al.* 1999), and the evolution of nonrecombining genomes.

DEFINITIONS AND MODELS

We first wish to make explicit our definitions of interference and the Hill-Robertson effect in terms of the fate of selected alleles at single loci embedded in genetic backgrounds of different fitnesses. Consider the fate of an allele with a selective advantage, s , over the wild-type allele, which is at frequency x in a haploid population. If there is variation in the fitness of individuals due to other loci and the average fitness of the population is scaled to 1, then the expected change in frequency of the advantageous allele due to selection is

$$M_{\delta x} = sx(1 - x) + \text{Cov}(x, w^*), \quad (1)$$

where $\text{Cov}(x, w^*)$ is the covariance between the allele and the fitness of the backgrounds on which it is found (Price 1970). That is, the expected change in allele frequency is the expected change in the absence of any interactions, plus the combined effects of other loci. It is this second (deterministic) term that we call interference. The expected change in allele frequency can also be written in a manner more applicable to explicit genetic models

$$M_{\delta x} = sx(1 - x) + \sum_y w_y D_{xy}, \quad (2)$$

where the second term is the sum, over all alleles (y) at other loci, of the linkage disequilibrium (D_{xy}) between the allele of interest and the second allele, times the relative fitness (w_y) of the second allele. When there is no linkage disequilibrium, the dynamics of the selected allele are the same as if there were no variation in genetic background. However, in finite populations, the stochastic nature of sampling and mutation generates linkage disequilibrium between the selected allele and genetic backgrounds, which then interferes with selection acting at the single locus (Felsenstein 1974). Disequilibrium is broken down by recombination (and mutation in the genetic background), while selection on genetic backgrounds tends to accentuate associations between the allele and the backgrounds. At the ex-

tremes, positive fitness covariance corresponds to hitchhiking (Maynard Smith and Haigh 1974) and negative fitness covariance corresponds to the elimination of alleles linked to strongly deleterious mutations (Fisher 1930, p. 136).

The Hill-Robertson effect (Hill and Robertson 1966) is a statement about the average effects of interference. Consider a large collection of identical populations into which single novel mutations are introduced at random onto backgrounds with different fitnesses. Associations made in the first generation are broken down by recombination and mutation, but new associations are generated stochastically through sampling and mutation. Thus, in any one generation the effect of selection at other loci is to increase the variance in allele frequency change across populations over that due to gamete sampling. This is equivalent to a reduction in the (variance) effective population size (N_e) and hence implies a decrease in the efficacy of selection. Under certain conditions, the long-term fate of neutral alleles (*e.g.*, the fixation time) can be predicted by a single, asymptotic reduction in N_e , where the effects are summed over generations and weighted by the fraction of associations remaining (Robertson 1961; Santiago and Caballero 1995, 1998). Because the efficacy of selection is determined by the effective population size, the net result of interference is to impede the fixation of favorable mutations and increase the fixation probability of deleterious alleles (Hill and Robertson 1966). This reduction in the efficacy of selection is commonly termed the Hill-Robertson effect (Felsenstein 1974).

One feature of interference as described by Hill and Robertson (1966) is of particular note: a buildup of negative linkage disequilibrium between selectively beneficial alleles. When beneficial mutations (or deleterious ones) are in positive linkage disequilibrium, selection is highly efficient and such haplotypes are rapidly fixed. In contrast, when beneficial mutations are in negative linkage disequilibrium, so that they are found in the same haplotype less frequently than expected from the product of the allele frequencies, then the efficacy of selection is weakened. This is because negative linkage disequilibrium reduces fitness variance and hence, by Fisher's fundamental theorem (Fisher 1930), the rate of adaptation. Such repulsion haplotypes tend to persist in populations (since they are neutral with respect to each other), leading to an excess of negative fitness disequilibrium between selectively favorable alleles. In an equilibrium system, this leads to a reduction in mean population fitness relative to the case of free recombination.

In this article, we consider two explicit genetic models to assess the impact of Hill-Robertson interference on patterns of molecular evolution. First, we consider a pair of completely linked loci at which one locus is segregating for a pair of alleles at some given frequency and novel mutations appear at the second as single

copies. This case has already received some attention (Hill and Robertson 1966; Barton 1995; Otto and Barton 1997); however, here we consider weaker selection and a wider range of evolutionary statistics than previously. We then investigate the case of a population where each chromosome consists of a large number of sites, each of which conforms to a two-allele, reversible-mutation model with genic selection. This is the simplest model for the evolution of synonymous codon usage (Li 1987; Bulmer 1991; McVean and Charlesworth 1999), in which the rates of forward and reverse mutation are equal.

A TWO-LOCUS MODEL OF INTERFERENCE

First consider the simplest case of two completely linked, biallelic sites with the same selection coefficient at each. If one of the loci is segregating, and a single novel mutation occurs at the second site, the fate of this new mutation is altered relative to that expected in the absence of interference (Fisher 1930; Barton 1995). Variants that arise in association with the beneficial allele will have an increased probability of fixation due to the hitchhiking effect, while those that arise in repulsion will have a reduced chance of becoming established in the population. We consider the extent to which such events alter the fixation probability, the contribution to heterozygosity, and the average time to loss or fixation for both advantageous and deleterious mutations.

The fate of single novel mutants (10^7 replicates) in a population of 100 diploids was followed until absorption (loss or fixation) for different starting frequencies of the favored allele at the linked locus. Linkage is complete and selection is genic (*i.e.*, heterozygotes are of intermediate fitness and the differences in fitness between homozygotes is $2s$). Selection across loci is also additive. For each replicate, the time to absorption (total sojourn time) and the contribution to heterozygosity [the sum of $2x(1-x)$ over generations, where x is the allele frequency] were recorded, as in Charlesworth *et al.* (1995). Pseudomultinomial sampling was used to select individuals for the next generation (Press *et al.* 1992). Three cases were examined: $4N_e s = 1, 2,$ and 4 for both advantageous and deleterious alleles. This analysis differs from that of Barton (1995) in that alleles at both loci are subject to drift.

The results of these simulations are shown in Figure 1. As previously shown (Hill and Robertson 1966; Barton 1995), interference tends to reduce the average fixation probability of beneficial mutations and increase the average fixation probability of deleterious mutations. Conditioning on cosegregation interference is stronger with greater selection coefficients; for example, the maximal reduction in average fixation probability for beneficial alleles with $4N_e s = 1$ is $\sim 1\%$ while it is 15% for beneficial alleles with $4N_e s = 4$. This is not to say that such weakly selected sites have no effect on

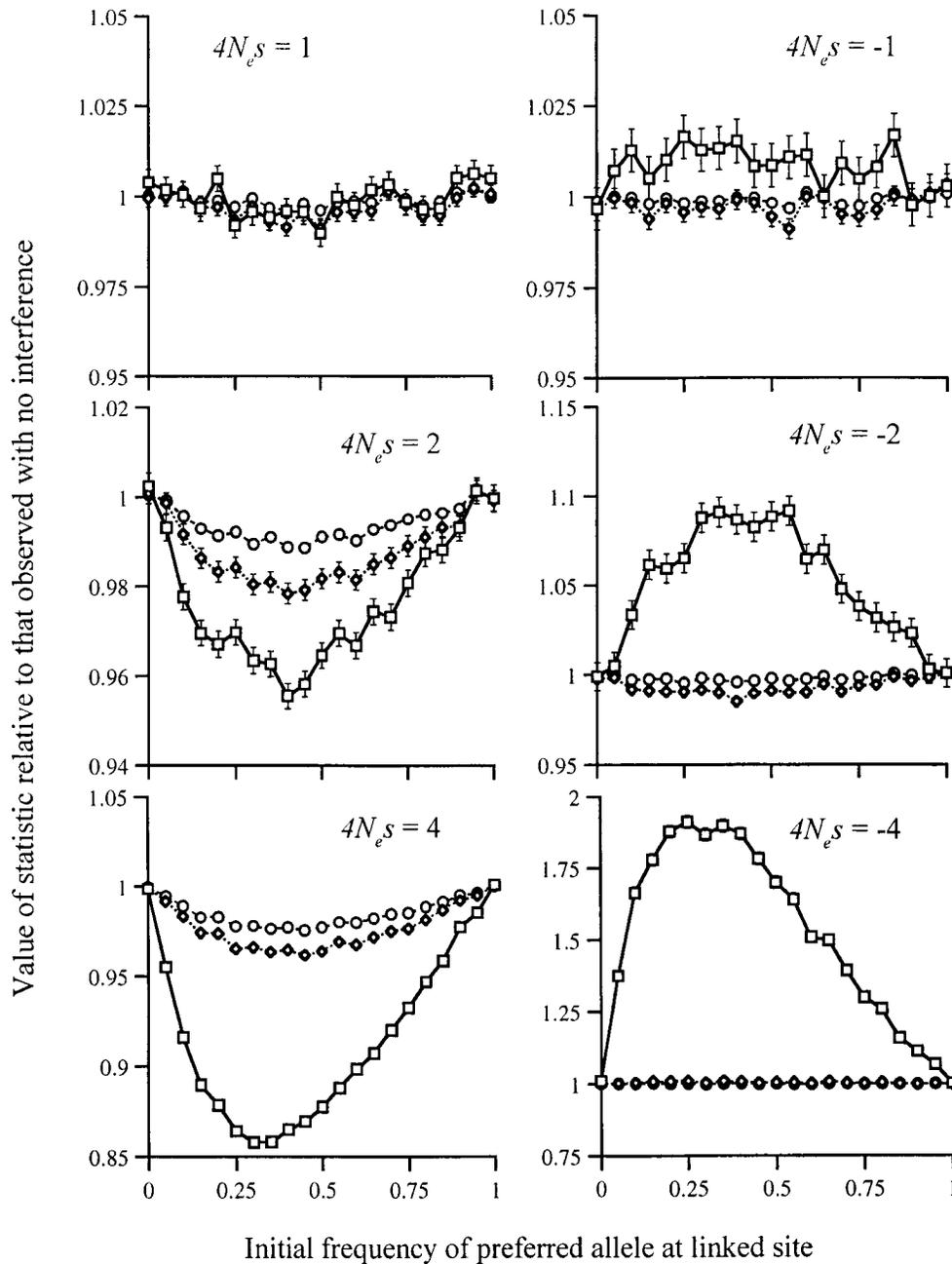


Figure 1.—The effects of interference on fixation probability (squares), average contribution to heterozygosity (diamonds), and average time to loss or fixation (circles) in the two-locus model.

the fixation probabilities of linked alleles, only that the average effect is negligible. Figure 2 shows the ratio of the average values when alleles arise in association with the beneficial allele to those when they arise in repulsion. Even with $4N_e s = 1$ there are strong effects on the fixation probabilities of linked alleles, though interestingly this seems to be largely independent of the frequency of the currently segregating allele.

In contrast, interference has much less effect on the contribution to heterozygosity of novel mutations and even less impact on the average total sojourn time. For deleterious mutations, interference appears to have essentially no overall effect on either of these statistics. This is because the average values of these statistics

(unlike fixation probabilities) are little affected by coupling or repulsion between selected alleles (Figure 2).

In sum, from the two-locus model of interference we can conclude: (i) The overall effect of interference is to reduce the efficacy of selection at linked loci. (ii) This results in reduced fixation probabilities for beneficial alleles and higher fixation probabilities for deleterious alleles, the relative effects being stronger for more strongly selected alleles. (iii) Interference decreases the contribution to heterozygosity and time until absorption for beneficial alleles by a small degree and has almost no effect on either property for deleterious mutations.

How do these results relate to equilibrium patterns of molecular evolution? Clearly we expect the level of

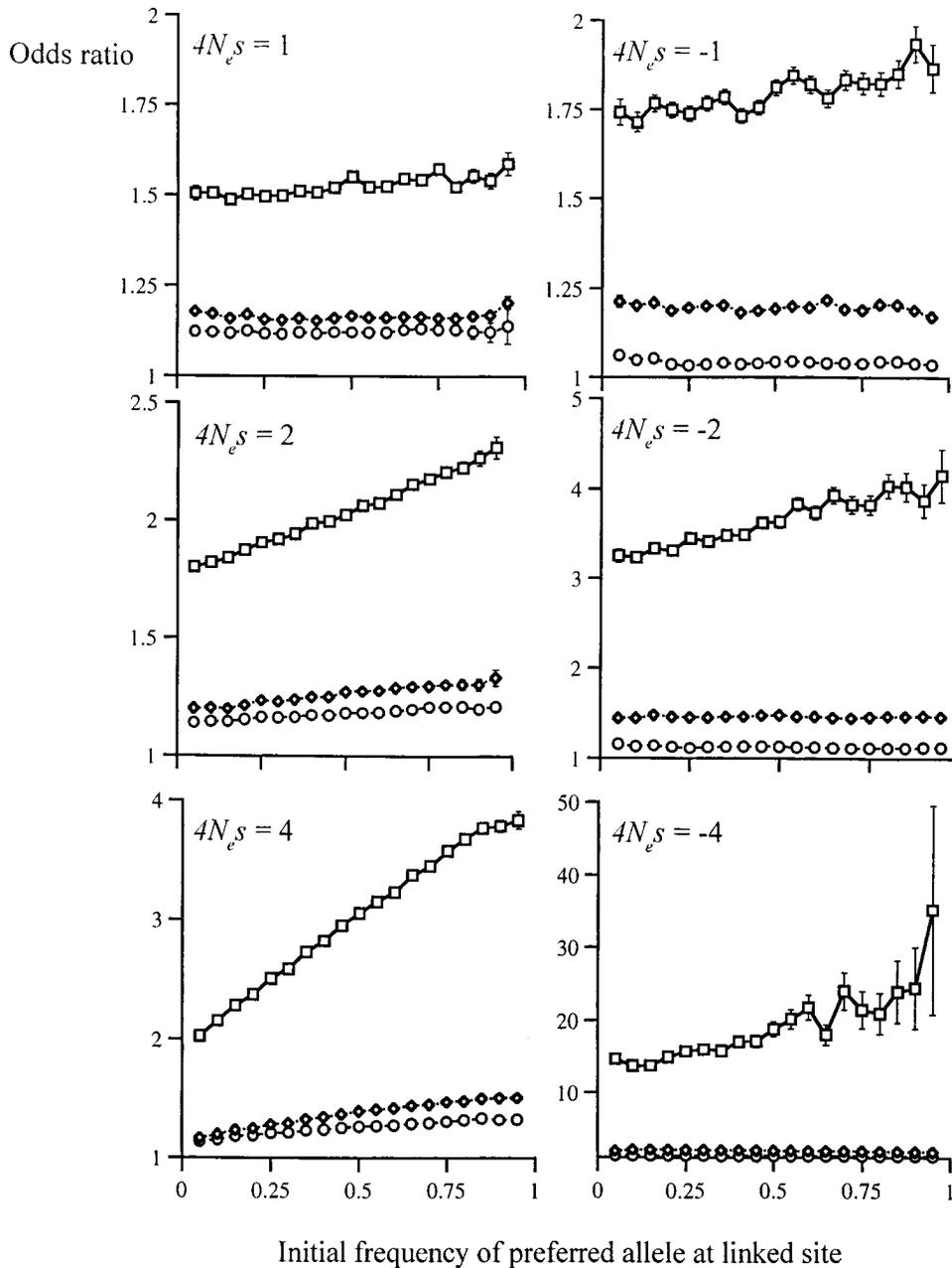


Figure 2.—The odds ratio of the fixation probability (squares), average contribution to heterozygosity (diamonds), and average time to loss or fixation (circles) for single novel mutants arising in association with a selectively favorable allele to that for mutants arising in repulsion.

adaptation (including codon bias) to be reduced in regions of low or absent recombination, while the level of nucleotide diversity (which is proportional to the heterozygosity contributed by a new mutation) should be less affected, and the number of segregating sites in the population (derived from the sojourn time of new mutations; Ewens 1979, p. 238) to be even less so. But it is difficult to extrapolate from these simulations to the equilibrium case with multiple loci. First, to predict the effects of interference on equilibrium statistics for even two loci requires that the above results be weighted by the probability that a novel mutation appears while

the other site is segregating and by the allele frequency distribution in such an event. So, for example, while the effects of interference (conditioning on cosegregation) increase with the selection coefficient, the probability of cosegregation decreases with increasing selection coefficient, and the allele frequency distribution becomes skewed toward rare variants that generate little interference. Second, we may expect that the effects of multiple loci are not simply the multiplicative effects of all pairwise interactions (unlike the case of background selection; M. Nordborg, unpublished results), because higher-order disequilibria will tend to affect interfer-

ence between any given pair of sites. For these reasons we have conducted an extensive series of multiple-locus simulations.

A MULTILOCUS MODEL OF INTERFERENCE

We consider a Wright-Fisher model of a diploid population of 100 individuals, each of which has a single pair of chromosomes consisting of G selected sites ($200 \leq G \leq 5000$). Simulations were carried out using a modified version of a program written by Richard R. Hudson. Each site represents a biallelic, reversible-mutation locus modeling the evolution of synonymous codon usage with equal rates of forward and reverse mutation (Bulmer 1991). Mutations occur at a frequency of μ per site per generation and can occur at sites that are currently segregating. Our results should therefore be compared with the allele-frequency distribution under selection and reversible mutation of Wright (1931) rather than with those based on the infinite-sites assumption (Kimura 1971), although this makes little difference for $N_e\mu \ll 1$. As previously, selection at each locus is genic with selection coefficient s . Selection across sites is multiplicative and all sites have the same selection coefficient. Recombination occurs with frequency r between adjacent sites per generation with a maximum of one crossover event per chromosome per gamete. Both larger population sizes (500 and 1000 individuals) and haploid populations were also investigated, but no differences were found in any case (*i.e.*, all results appear to depend on the scaled parameters $N_e\mu$, $N_e s$, and $N_e r$, as expected from diffusion theory; Ewens 1979).

Each statistic is the average over at least four runs, each consisting of 50 samples of 25 alleles taken every $2/\mu$ generations after an initial period of $2/\mu$ generations for the system to reach equilibrium. Such a time period between samples minimizes the risk of evolutionary nonindependence between samples, although each run was checked for autocorrelation and multiple independent runs were carried out. In all cases, except where stated, the scaled mutation rate per site (both forward and backward) was $4N_e\mu = 0.04$. While this is larger than the mean value of 0.013 for synonymous sites in *D. melanogaster* (Moriyama and Powell 1996), we expect the extent of interference to scale with the product of G and the scaled mutation rate. We consider scaled recombination rates between adjacent sites of $4N_e r = 0, 0.01, 0.1, \text{ and } 1$ (though $4N_e r = 1$ cannot be considered for the case of 5000 sites due to the limit of 1 crossover per gamete). For comparison with data on synonymous codon usage, r corresponds roughly to three times the per nucleotide site recombination rate. Statistics from simulated data sets are compared with those predicted from the sampling properties (Ewens 1979) of Wright's

(1931) model of selection and reversible mutation, assuming the independent evolution of different sites.

RESULTS

The effects of interference on the level of molecular adaptation: Interference between weakly selected mutations can have a considerable effect on the level of molecular adaptation, as measured here by the average frequency of preferred alleles at each locus among gametes sampled from the population. In terms of the codon bias model, we measure codon bias as the deviation from equal usage of alleles, or $2x - 1$, where x is the average frequency of preferred alleles over loci. Relative codon bias is the ratio of the observed bias to that expected from Wright's distribution of allele frequencies with selection and reversible mutation, assuming independence between sites.

Figure 3 shows how relative codon bias is sensitive to the selection coefficient acting on codon usage and the rate of recombination between adjacent sites for 500, 1000, and 5000 linked sites. Over the range of selection coefficients considered, the impact of interference on relative codon bias is similar for all values of $N_e s$, although the magnitude of change in the frequency of the preferred codon is greater for stronger selection. For example, with $4N_e s = 4$ and 5000 completely linked sites, the mean frequency of preferred alleles is reduced to 0.73, compared to a frequency of 0.97 in the absence of interference, while with $4N_e s = 1$, the respective decrease in the frequency of preferred alleles is from 0.71 to 0.60.

Recombination reduces the effects of weak selection Hill-Robertson (wsHR) interference on codon bias (Figure 3). An increase in codon bias is observed over the entire range of recombination rates for all selection coefficients and numbers of sites (although the effects are of greater magnitude for sites under stronger selection). Furthermore, for the numbers of sites considered, $4N_e r = 1$ appears to be sufficient to remove most of the effects of interference.

Levels of nucleotide diversity and the frequency distribution of segregating sites: The level of nucleotide-site diversity (average pairwise difference between alleles) is affected by interference in both quantitatively and qualitatively different ways from codon bias (Figure 3). Diversity tends to be reduced by interference (but see below, for the case of stronger selection coefficients). However, the proportional decrease in diversity is less than the proportional decrease in codon bias. This implies that the effects of wsHR interference on adaptation and polymorphism cannot be treated as a single decrease in N_e , since in the absence of mutation bias single-site models predict that a decrease in N_e causes the same proportional decrease in diversity and codon bias (McVean and Charlesworth 1999). In addition, inter-

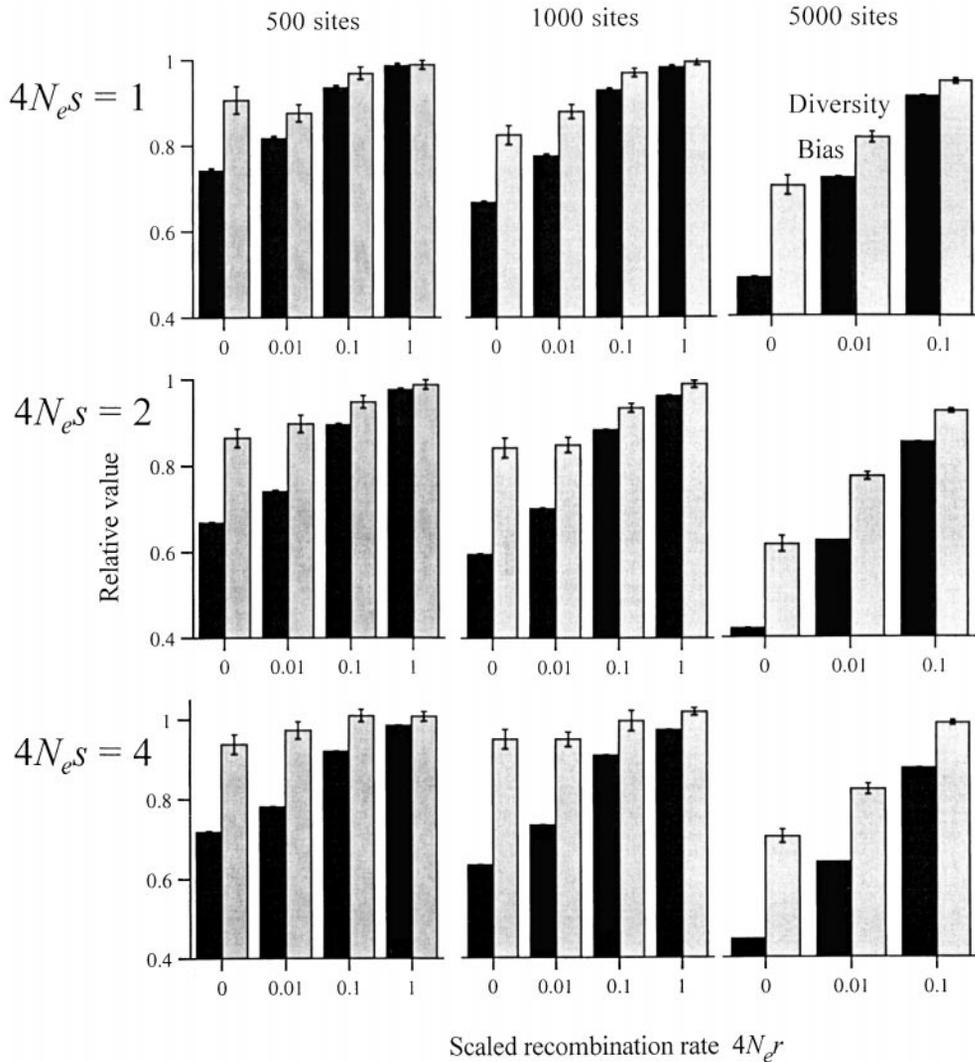


Figure 3.—The effects of wsHR interference on relative codon bias (solid) and nucleotide site diversity (shaded); $4N_e\mu = 0.04$.

ference has little effect on diversity for 1000 or fewer sites, but does decrease diversity for 5000 and more linked sites. To a large extent, this pattern is repeated in the number of sites observed to be segregating in a sample (data not shown), although the effects are weaker than on levels of nucleotide site diversity.

Interference affects the frequency distribution of segregating sites in two ways (Figure 4). First, it generates high-frequency deleterious mutations such that the distribution resembles that expected from single-site dynamics with smaller selection coefficients. Second, it increases the proportion of rare variants over that expected from single-site dynamics. We can assess the second effect by considering statistics describing the deviation of the frequency distribution from neutral expectations, such as those of Tajima (1989) and Fu and Li (1993). Under the infinite-sites, standard neutral model, the expectations of the statistics π (average pairwise differences), η/a_n (number of segregating sites in a sample of size $n/\sum_{i=1}^{n-1} 1/i$) and η_s (the number of external-branch mutations) are all equal to the scaled

mutation rate, $4N_e\mu$ (Watterson 1975; Tajima 1989; Fu and Li 1993). The difference between any two estimators of $4N_e\mu$ (scaled by the expected variance in the difference) describes particular aspects of deviations from neutrality. Tajima's D -statistic compares the estimates of $4N_e\mu$ from π and the number of segregating sites, while Fu and Li's D^* -statistic compares estimates from π and the number of external-branch mutations. Negative values of either statistic indicate an excess of low-frequency variants.

Table 1 shows the average value of Tajima's D and Fu and Li's D^* -test statistics and the proportion of samples showing significant deviation from neutrality for the cases of 1000 linked sites with no recombination and with $4N_e r = 1$, $4N_e\mu = 0.04$, a sample size of 25 sequences, and $4N_e s$ in the range 0–20. For $4N_e s \leq 4$, while selection causes a skew toward rare variants, and wsHR interference increases the skew, there is essentially no power to detect the action of selection (see also Akashi 1999) or interference. For stronger selection coefficients ($4N_e s \geq 10$), Tajima's D -statistic has some

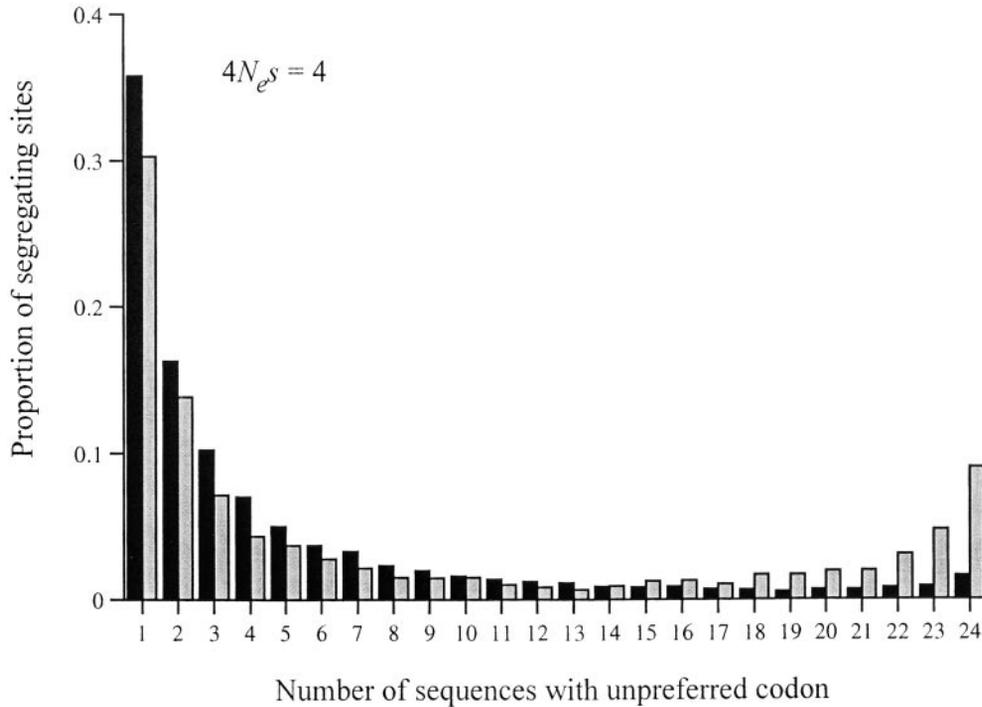


Figure 4.—The average frequency distribution of segregating sites in samples of 25 alleles with no recombination (shaded) and with $4N_e r = 1$ (solid) for 1000 sites and $4N_e \mu = 0.04$.

power to detect selection acting on codon usage, but interference reduces both the skew toward rare variants and the proportion of significant test statistics. In short, there appears to be little power in either of these tests to detect weak selection, such as that experienced by synonymous-site mutations, or the action of wsHR interference.

Estimating the strength of selection on allelic variants:

The bias introduced by wsHR interference into methods of estimating the selection coefficient acting on allelic variants is of considerable interest. Methods of estimating the strength of selection (*e.g.*, that acting on codon usage) are derived from fits of Wright’s (1931) distribution of allele frequencies under selection and reversible mutation to various sample statistics, assuming independence between sites. For the case of a two-allele model with no mutation bias, the expected proportion of preferred codons at fixed sites (in the population), at mutation-selection-drift equilibrium, is given by

$$E(x) = [1 + e^{-4N_e s}]^{-1} \tag{3}$$

(Li 1987; Bulmer 1991). Hence an approximate value of $4N_e s$ can be estimated from the relative frequency of sites fixed for preferred and unpreferred codons in a sample. Previous work has also shown that this formula accurately predicts the average proportion of preferred codons in a single sequence chosen at random from the population (McVean and Charlesworth 1999); $4N_e s$ can thus also be estimated from the average frequency of preferred codons in a sample of sequences. Finally, the value of $4N_e s$ can be estimated from the frequency distribution of segregating sites by using maximum likelihood to fit the expected sample configura-

tion to that observed (Sawyer and Hartl 1992; Hartl *et al.* 1994) and assuming independence between sites. While the assumption is valid only for the case of free recombination, our interest is in considering the bias caused by wsHR interference.

Under the infinite-sites model (which is the limiting case of the present model as $\mu \rightarrow 0$) at mutation-selection-drift equilibrium in a diploid population, the diffusion theory solution for the frequency distribution of preferred alleles at segregating sites in a two-allele model with genic selection and reversible mutation is

$$\phi(y) = \frac{C e^{4N_e s y}}{y(1-y)} \tag{4a}$$

(Wright 1931; McVean and Charlesworth 1999). The log likelihood of observing a given sample configuration (conditioning on segregation and assuming independence) is

$$L(X_1, X_2, \dots, X_{n-1}) = \sum_{i=1}^{n-1} L(X_i), \tag{4b}$$

where the likelihood of observing each count is given by the multinomial (or Poisson) distribution, such that

$$L(X_i) \propto X_i \ln \left[\frac{\binom{n}{i} \int_0^1 y^{i-1} (1-y)^{n-i-1} e^{4N_e s y} dy}{\int_0^1 [1 - y^n - (1-y)^n] e^{4N_e s y} y^{-1} (1-y)^{-1} dy} \right], \tag{4c}$$

where X_i is the number of segregating sites at which the preferred allele occurs in i of n sequences. Numerical analysis can be used to find the value of $4N_e s$, which maximizes the likelihood of observing the sample. This

TABLE 1
The effects of interference on sample statistics

$4N_e s$	$4N_e r$	Tajima's D: average (% $P < 0.05$)	Fu and Li's D*: average (% $P < 0.05$)	Average $4N_e s \pm SD$ estimated from codon bias	Average $4N_e s \pm SD$ estimated from fixed sites ^a	Average $4N_e s \pm SD$ estimated from segregating sites	Average $\sigma^2/E[\sigma^2] \pm SD$
0	0	-0.09 (0.013)	0.05 (0.018)	0.00 \pm 0.06	0.00 \pm 0.07	0.02 \pm 0.23	0.97 \pm 0.85
	1	-0.09 (0)	0.07 (0)	0.00 \pm 0.06	0.00 \pm 0.07	0.02 \pm 0.23	1.02 \pm 0.36
1	0	-0.51 (0.015)	-0.37 (0.025)	0.59 \pm 0.05	0.63 \pm 0.06	0.65 \pm 0.29	0.93 \pm 0.70
	1	-0.17 (0)	-0.01 (0)	0.90 \pm 0.07	0.98 \pm 0.08	0.94 \pm 0.26	1.03 \pm 0.38
2	0	-0.62 (0.015)	-0.41 (0.010)	0.92 \pm 0.06	0.97 \pm 0.07	0.99 \pm 0.32	0.83 \pm 0.58
	1	-0.39 (0)	-0.20 (0)	1.71 \pm 0.08	1.86 \pm 0.09	1.79 \pm 0.33	0.99 \pm 0.30
4	0	-0.99 (0.070)	-0.83 (0.025)	1.37 \pm 0.06	1.45 \pm 0.08	1.49 \pm 0.46	0.84 \pm 0.56
	1	-0.84 (0)	-0.66 (0)	3.09 \pm 0.14	3.49 \pm 0.19	3.48 \pm 0.53	0.99 \pm 0.34
10	0	-1.48 (0.220)	-1.53 (0.085)	2.38 \pm 0.09	2.48 \pm 0.10	2.92 \pm 0.80	0.76 \pm 0.39
	1	-1.71 (0.350)	-1.70 (0.05)	5.32 \pm 0.21	—	8.26 \pm 1.60	0.98 \pm 0.31
20	0	-1.83 (0.615)	-2.21 (0.300)	3.94 \pm 0.18	4.17 \pm 0.24	6.33 \pm 3.71	0.84 \pm 0.38
	1	-2.13 (0.965)	-2.44 (0.435)	6.17 \pm 0.24	—	15.94 \pm 4.26	1.01 \pm 0.31

All analyses are based on simulations with 1000 sites, $4N_e \mu = 0.04$, and a sample size of 25 sequences.

^a For the cases of $4N_e s = 10$ and 20 with $4N_e r = 1$, samples rarely contain sites fixed for unpreferred codons.

approach differs from that of Sawyer and Hartl (1992) and Hartl *et al.* (1994) in that it assumes reversible, as opposed to irreversible, mutation and considers only segregating sites. Values of $4N_e s$ estimated by each method were obtained for each sample taken from the simulations.

Table 1 shows the average estimated values of $4N_e s$ (\pm one standard deviation). Estimates from average bias are lower than from fixed sites alone, which again are lower than from the frequency distribution (which for $4N_e r = 1$ are close to the true value). As the selection coefficient increases, the discrepancy between estimates increases, such that for $4N_e s \geq 10$, estimates based on the average codon bias are considerable underestimates with $4N_e r = 1$. wsHR interference reduces estimates of the strength of selection by all methods, and for $4N_e s \leq 4$, the proportion by which estimates are reduced relative to the case of $4N_e r = 1$, is approximately the same for each method. That is, wsHR interference shifts the frequency distribution of segregating sites toward that expected for more weakly selected alleles, in a manner compatible with the reduction in efficacy of selection as estimated from fixed sites or average codon bias. For stronger selection coefficients, $4N_e s \geq 10$, wsHR interference has a greater effect on estimates of $4N_e s$ from the frequency distribution of segregating sites than from average codon bias.

These results suggest that while wsHR interference can have a large effect on estimates of the strength of selection, the reduction in efficacy of selection caused by interference is difficult to distinguish from a simple reduction in the selection coefficient, at least for weak selection ($4N_e s \leq 4$).

The rate of sequence divergence: In the absence of mutation bias, the rate of substitution of new mutations is expected to decrease with increasing $N_e s$ (Shields *et al.* 1988; Eyre-Walker and Bulmer 1995; McVean and Charlesworth 1999), and so we expect a decrease in the efficacy of selection (N_e) to increase the rate of substitution. For the case of 1000 sites with $4N_e s = 2$, Figure 5 shows how the amount of sequence divergence between two species decreases with increased recombination. Differences in the extent of sequence divergence are only detectable at considerable degrees of divergence and, to a large extent, this is simply the result of different equilibrium levels of codon bias. This is the same problem identified in the previous selection; wsHR can have a large effect on patterns of evolution and variation, but a reduced efficacy of selection is difficult to distinguish from a reduced selection coefficient.

Patterns of linkage disequilibrium: On average, interference between selected alleles reduces the efficacy of selection and leads to reduced codon bias. While we have demonstrated that interactions between synonymous-site mutations can have large effects on codon bias, we have also shown that these effects are difficult to distinguish from a simple reduction in the selection coefficient by means of standard tests. In addition, there

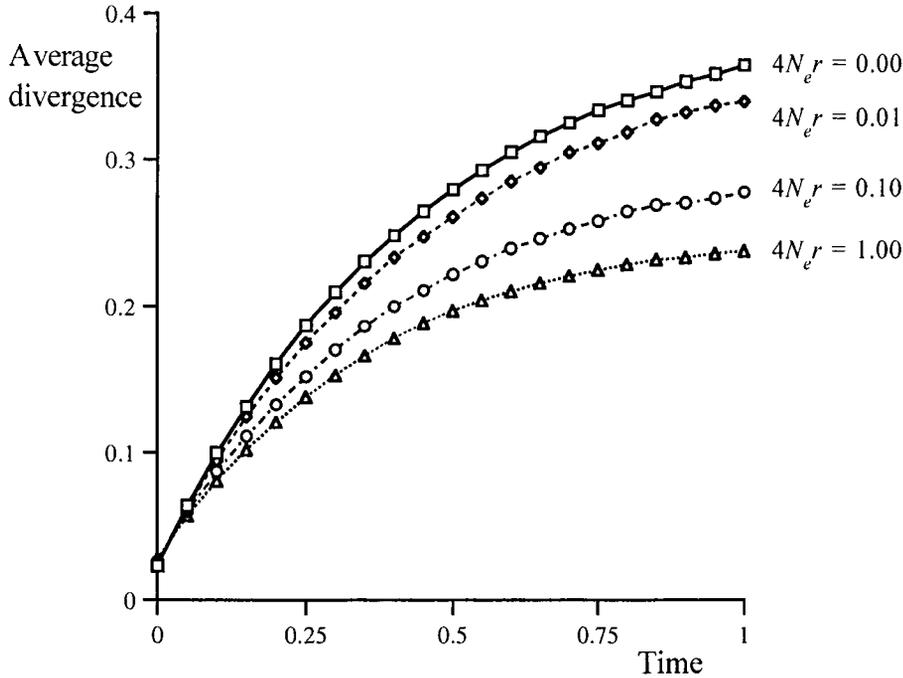


Figure 5.—The effect of interference on the average amount of sequence divergence for 1000 selected sites with $4N_e s = 2$ and $4N_e \mu = 0.04$. Time is measured in terms of the expected number of neutral substitutions per site (*i.e.*, $1/\mu$ generations).

may be other explanations for reduced codon bias in regions or genomes with low recombination rates, such as background selection or hitchhiking (Kliman and Hey 1993). This raises the question of whether there are specific properties of interference between weakly selected mutations that can be used to test for its influence.

One possibility derives from the observation that Hill-Robertson interference tends to lead to a buildup of negative linkage disequilibrium between selectively favorable alleles (Hill and Robertson 1966). We expect excess repulsion between preferred codons and a dearth in variation in fitness relative to that expected from the allele frequencies alone. We can therefore test two predictions: (i) the observed variance in the number of beneficial alleles among individuals should be less than that expected from the sum of the contributions across loci,

$$E[\sigma^2] = \sum_i x_i(1 - x_i), \quad (5)$$

where x_i is the frequency of the preferred allele at site i . And, as a direct corollary, (ii) the average value of pairwise linkage disequilibrium between preferred codons should be negative for tightly linked sites and tend to zero as recombination increases. We tested these predictions with our simulation results for 1000 selected sites and $4N_e s$ in the range 0–20. For prediction (ii) we consider only the case of 1000 sites with $4N_e s = 4$ and $4N_e r = 0.01$.

In the absence of recombination, wsHR interference generates an excess of negative linkage disequilibrium between preferred alleles (Table 1), so that the average ratio of observed-to-expected variance in the number

of preferred alleles is less than one. However, the distribution of the statistic $\sigma^2/E[\sigma^2]$ has high variance and is skewed toward high values, such that for neutral sites in the absence of recombination, the ratio is less than one in 67% of samples (although the expectation is one). Hence while this test can be used to detect wsHR interference between synonymous sites, it should be applied with caution.

Figure 6 shows the result of the second test. For pairs of segregating sites in close proximity, the average value of D between preferred alleles is less than zero, while the average absolute value $|D|$ decreases monotonically with distance. For more distant pairs, the average value of D tends to zero as expected. However, the properties of linkage disequilibrium are such that there is little power in this test. Furthermore, complications such as variable selection coefficients across sites will also reduce the power of any test based on linkage disequilibrium between putatively “preferred” alleles.

Stronger selection coefficients: We can compare the results obtained here with those known under certain limiting cases. With irreversible mutation, strong selection, and an infinite population size, the distribution of the number of deleterious mutations across individuals follows a Poisson distribution and is independent of the level of recombination (Kimura and Maruyama 1966; Haigh 1978; Maynard Smith 1978, p. 35). Hence for strong selection coefficients we should expect to see fewer effects of interference. Figure 7 shows the magnitude of the effects of interference, as defined by the ratio of the value of statistics (codon bias, mean population fitness, and average diversity) obtained with no recombination, to those with $4N_e r = 1$ (for 1000 sites). For all statistics we find a maximum for the effects of interfer-

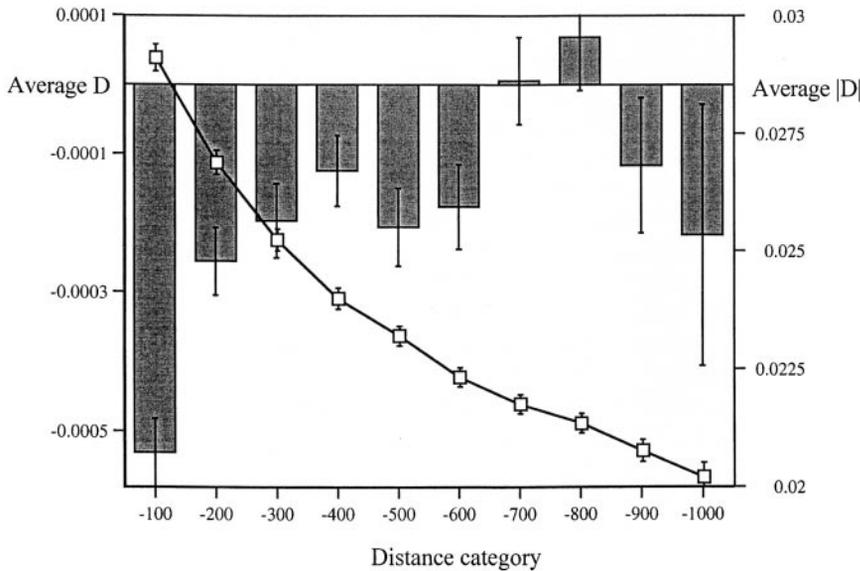


Figure 6.—Average linkage disequilibrium (columns) and average absolute linkage disequilibrium (line) between preferred alleles in a sample of 25 alleles as a function of the distance between segregating sites for 1000 sites with $4N_e s = 4$, $4N_e r = 0.01$, and $4N_e \mu = 0.04$.

ence at an intermediate selection coefficient; however, the value of $4N_e s$ at which the maximum occurs is dependent on the statistic. For codon bias, the relative effects of interference are strongest for $4N_e s = 2$, while the maximal decrease in mean population fitness is at $4N_e s = 10$. Interference decreases diversity for low selection coefficients (with a minimum at $4N_e s = 1$), while diversity is actually increased over that expected from single-site dynamics for $4N_e s \geq 10$.

This last result highlights two important features of the effects of interference on polymorphism. Linked heritable variation in fitness shifts the behavior of selected alleles toward that of neutral ones by increasing the variance in reproductive success. This both increases polymorphism by reducing the efficacy of selection (McVean and Charlesworth 1999) and decreases variation by reducing the fraction of the population that

will contribute to future generations. For very weakly selected sites, it is the latter effect that dominates, while for more strongly selected sites, it is the shift toward neutrality. For completely neutral loci, wsHR interference acting at linked sites will always reduce variability.

The effects on mean population fitness are also of particular note. With 1000 sites and $4N_e s = 10$, there is a maximal 75% reduction in mean fitness relative to high recombination rates (though for larger N_e and fixed $N_e s$ the reduction would be smaller). With a greater number of sites the effects are even greater. As has been noted before (Kondrashov 1995), the accumulation of weakly deleterious mutations presents a potential paradox in terms of genetic load. Reduced recombination rates can increase the magnitude of this effect considerably. Previous analytical work on models of interference involving a few loci has found only a small advantage

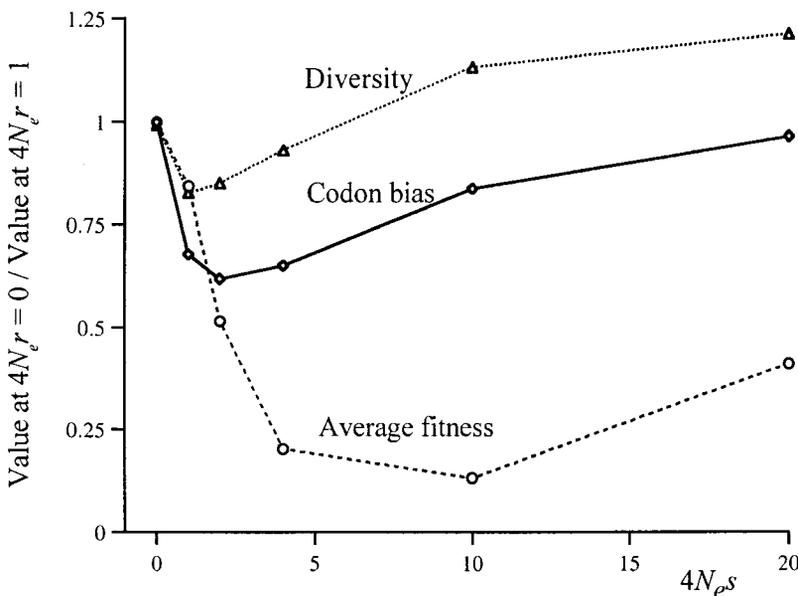


Figure 7.—The maximal effects of wsHR interference on codon bias, nucleotide diversity, and average fitness, as measured by the ratio of the average value observed under no recombination to that observed with $4N_e r = 1$ for 1000 sites and $4N_e \mu = 0.04$.

to modifiers that increase recombination rates (Otto and Barton 1997). Our results suggest that in multiple-locus systems with weak selection such modifiers may provide a much greater advantage.

Multiplicative vs. additive selection: A previous study found little difference between multiplicative and additive selection across loci on the magnitude of interference between weakly selected mutations (Li 1987). However, we find that under certain circumstances there can be considerable differences in the degree of codon bias observed under the two models. In particular, when the product of the selection coefficient (s) and the number of sites under selection (G) is large ($Gs \approx 1$), then additive selection is more efficient at eliminating deleterious mutations under low recombination rates (data not shown). This is because the genetic load imposed by a given number of mutations is greater in the case of additive selection across loci (*i.e.*, there is synergistic epistasis between deleterious mutations), so selection is more effective for additive selection, leading to higher codon bias and mean population fitness. The parameter values considered by Li (1987) were chosen to minimize the difference in fitness between multiplicative and additive selection across loci, *i.e.*, the approximation $(1 - s)^G \approx 1 - Gs$ holds. The product Gs is therefore critical in determining the importance of the epistatic nature of selection (Kondrashov 1995), and values greater than one can lead to considerable differences in fitness between multiplicative and epistatic interactions for a given number of deleterious mutations. For *Drosophila* it seems likely that Gs exceeds one for synonymous codon positions (Kondrashov 1995); hence synergistic epistasis between the fitness effects of unpreferred codons could place a limit on the extent to which interference can decrease adaptation.

DISCUSSION

Previous work has shown that interference between cosegregating selected alleles can reduce the efficacy of selection considerably (Hill and Robertson 1966; Felsenstein 1974; Li 1987; Barton 1995; Comeron *et al.* 1999). This article is the first comprehensive analysis of the extent to which interactions between weakly selected mutations, such as those occurring at synonymous sites in many organisms, can affect patterns of molecular adaptation, polymorphism, and divergence. We suggest that wsHR interference may be an important influence on genome evolution and have proposed ways of detecting its action. In addition, we wish to consider whether such interference may be important in specific cases relating to empirical data.

wsHR interference reduces the sensitivity of levels of codon bias to population size: One of the most notable features of synonymous codon usage is that similar degrees of codon bias are observed in species of very different census population sizes (Powell and Moriyama

1997). For example, codon bias in *Drosophila* is, if anything, greater than in *Escherichia coli* (Powell and Moriyama 1997). This is remarkable, given that there is only a very narrow range of values of $N_e s$ under which a balance between mutation, selection, and drift is expected under single-site models (Li 1987; Bulmer 1991). There are four possible explanations for this paradox: (A) the selective value of alternative codons differs considerably between species and is negatively correlated with population size; (B) there is some form of threshold selection on codon bias that puts an upper limit on the power of selection to increase codon bias; (C) the effective population sizes of all such species are very similar and within the appropriate range; or (D) interference between synonymous codons is sufficiently strong to maintain moderate codon bias over several orders of magnitude of population size.

Explanation A seems highly unlikely, and one might in fact expect that unicellular organisms under strong replication-rate selection should have stronger translation-rate-mediated selection than multicellular organisms. Neither is there evidence or theoretical reason to suspect strong synergistic fitness effects (such as truncation selection) of unpreferred codons (explanation B). In contrast, there are many factors that reduce N_e , such as background selection, hitchhiking, and population subdivision with nonconservative migration (explanation C). The problem is to understand why the factor by which N_e is reduced should be such that $N_e s$ remains within the narrow range (0.1–1.0) required to maintain a balance between mutation, selection, and drift.

Explanation D does not face the same problem, as selection acting at synonymous sites both generates and limits codon bias. The question is whether realistic parameters can generate sufficient “selective drag” to maintain intermediate codon bias over several orders of magnitude in population size. Figure 8 shows how codon bias increases across three orders of magnitude change in haploid population size for a given per-site mutation rate (forward and backward rates of 2.5×10^{-6} per site per generation) and selection coefficient ($s = 10^{-3}$) when there are 10^4 and 10^5 completely linked sites. As suggested, wsHR interference maintains intermediate levels of codon bias when single-site models would predict complete fixation of preferred codons. For example, with $2N_e s = 20$ and $2N_e \mu = 0.05$, the average frequency of unpreferred codons is expected to be 2.5×10^{-3} (the deterministic equilibrium), while it is actually >0.3 for 10^5 sites. These results are for complete linkage, but as seen from Figure 3, wsHR interference can have a large effect even with moderate recombination rates.

One final prediction is worth noting. In bacteria there is evidence for both selection on synonymous codon usage (Sharp and Li 1986) and variation between species in the genome-wide level of recombination (Maynard Smith *et al.* 1993). If interference (either wsHR

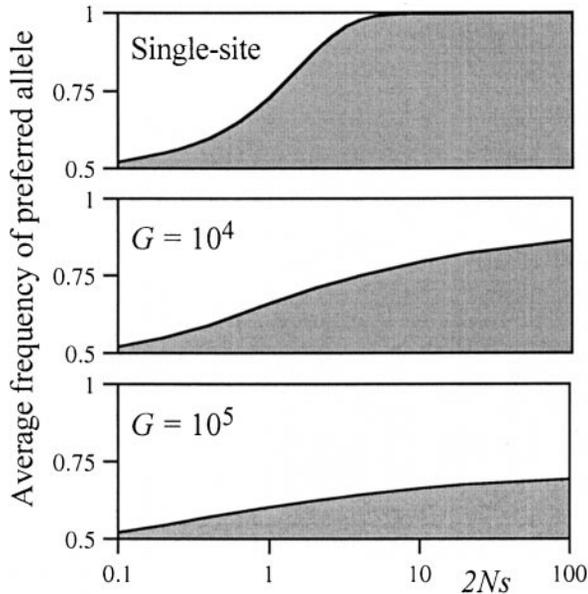


Figure 8.—The increase in the frequency of preferred codons expected from single-site dynamics and that observed for 10^4 and 10^5 completely linked sites (G) across three orders of haploid population size for a fixed per-site mutation rate (2.5×10^{-6}) and selection coefficient (10^{-3}).

or background selection) is responsible for the constancy of codon bias across species, then we predict that, on average, species with higher recombination rates should demonstrate higher levels of codon bias.

Detecting the influence of wsHR interference on patterns of codon bias and polymorphism in *Drosophila*: While patterns of codon bias in *Drosophila* demonstrate clear evidence for selection on synonymous codon usage mediated at the level of translational efficiency (Shields *et al.* 1988; Akashi and Schaeffer 1997; Duret and Mouchiroud 1999), there are several features that are not predicted by simple single-site models (Kliman and Hey 1993; Comeron *et al.* 1999; McVean and Charlesworth 1999). These include the observations that codon bias and silent-site diversity are reduced in regions of low recombination (Begun and Aquadro 1992; Kliman and Hey 1993) and that codon bias and synonymous polymorphism are correlated even when genes in regions of no recombination are excluded (Moriyama and Powell 1996; Comeron *et al.* 1999). To what extent can wsHR interference provide explanations for these phenomena?

The inverse relationship between recombination and codon bias (Kliman and Hey 1993) and diversity (Begun and Aquadro 1992) is typically interpreted in terms of the reduced efficacy of natural selection in regions of low recombination, resulting from the hitchhiking effect of beneficial mutations (Maynard Smith and Haigh 1974) or from elimination of strongly deleterious mutations (background selection; Charlesworth *et al.* 1993; Hudson and Kaplan 1994; Stephan

et al. 1999). We have shown that wsHR interference can also reduce codon bias and synonymous site diversity considerably over a range of recombination rates. The challenge is to identify predictions that distinguish between these explanations.

The major differences between hitchhiking, background selection, and wsHR interference lie in the consequences for patterns of polymorphism; wsHR interference, like background selection (Hudson and Kaplan 1994), and unlike hitchhiking (Braverman *et al.* 1995), has little effect on the frequency distribution of segregating sites beyond a shift toward that expected for sites under weaker selection. However, as shown here, it can generate specific patterns of linkage disequilibrium between preferred codons. The problem is that when there is low polymorphism, there is often little power to detect patterns in the frequency distribution of segregating sites.

One important difference between wsHR interference and background selection is that, for a given decrease in equilibrium codon bias, the expected reduction in nucleotide diversity is less for interference than under background selection (McVean and Charlesworth 1999); as yet, there has been no detailed analysis of the effects of hitchhiking on patterns of codon usage. For example, with $4Ns = 2$, $4N\mu = 0.04$, 5000 sites, and no recombination, wsHR interference causes a decrease in bias of $\sim 60\%$ and a decrease in diversity of 40% (Figure 3). To achieve the same reduction in bias through background selection, a decrease in diversity of $\sim 60\%$ is expected (McVean and Charlesworth 1999). For stronger selection coefficients ($4Ns \geq 10$) interference can actually increase diversity relative to the case of free recombination, whereas with no mutation bias, the relationship between recombination rate and diversity is expected to be monotonic under background selection (McVean and Charlesworth 1999).

The extremely low levels of polymorphism observed within regions of essentially no recombination within the *D. melanogaster* genome (Begun and Aquadro 1992; Moriyama and Powell 1996) are therefore more compatible with background selection or hitchhiking than wsHR interference. In addition, even within regions of very low recombination (where the per-site recombination rate, c , is $< 10^{-9}$; Comeron *et al.* 1999) there is still considerable variation in codon bias between genes, and this is hard to reconcile with the level of background selection required for little or no diversity. Very low variation, but nonzero codon bias, suggests recent hitchhiking events or recent changes in local recombination rate (Comeron *et al.* 1999). Genome rearrangements that change the recombination environment of genes will affect levels of diversity more rapidly than levels of bias. However, for regions that have experienced zero recombination for the order of $1/\mu$ years, background selection and recurrent hitchhiking models predict neither diversity nor codon bias. It is therefore interesting

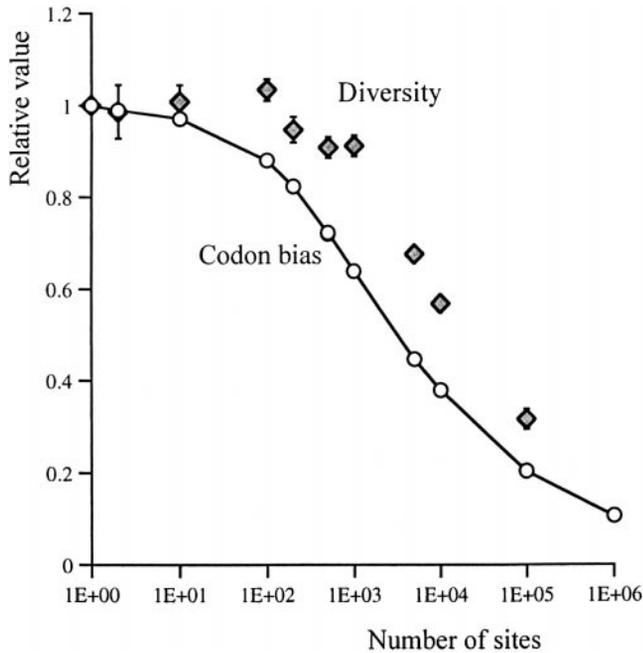


Figure 9.—The effect of the number of completely linked sites on relative codon bias (circles) and diversity (diamonds) for $4N_e s = 4$ and $4N_e \mu = 0.04$.

to note that codon bias on the nonrecombining fourth chromosome of *D. melanogaster* is close to that expected from mutational biases alone (Kliman and Hey 1993; Comeron *et al.* 1999). Assuming there are ~ 200 genes of average length (500 codons) on this chromosome (10^5 weakly selected sites), wsHR interference predicts low but nonzero bias and diversity (see also Figure 9).

These considerations also identify a possible interaction between wsHR interference and stronger forces such as background selection and hitchhiking. Because deleterious mutations tend to persist at very low frequencies and beneficial mutations are likely to be rare, background selection and hitchhiking will tend to reduce the efficacy of selection on codon usage only in regions of essentially no recombination. Under such circumstances, wsHR interference will tend to have little effect, as the reduction in diversity caused by other forces reduces the potential for wsHR interference. When background selection and hitchhiking are less strong, synonymous site diversity is expected to increase, which generates wsHR interference. At moderate recombination rates, we expect wsHR interference to become the major factor limiting selection on codon usage. wsHR interference may therefore provide an explanation for the observed relationship between codon bias and diversity for genes in *D. melanogaster* in regions of nonzero recombination (Moriyama and Powell 1998; Comeron *et al.* 1999).

The impact of wsHR interference on the evolution of nonrecombining genomes: There is considerable evidence, both theoretical and empirical, that nonrecom-

binning genomes are subject to the accumulation of deleterious mutations. Factors such as Muller's ratchet (Muller 1964; Felsenstein 1974), hitchhiking (Maynard Smith and Haigh 1974), and the elimination of deleterious mutations (Fisher 1930; Charlesworth 1994) have been invoked to explain patterns of evolution of Y chromosomes (Graves 1994; Charlesworth 1996; Rice 1996), organelle genomes (Lynch 1997; Lynch and Blanchard 1998), and bacterial endosymbionts (Moran 1996). To this list we must add wsHR interference as a process that limits the efficacy of selection in the absence of recombination. While interference is an integral part of Muller's ratchet (Felsenstein 1974), the difference between the wsHR interference model considered here and Muller's ratchet is that with back mutation interference causes genomes to reach a lower equilibrium fitness, rather than to decay continuously. In addition, the selection coefficients considered here are generally much weaker than those considered in models of Muller's ratchet, but the cumulative effects on fitness can be considerable. Because a large number of sites in genomes are likely to be under weak selection, we suggest that wsHR interference may be an important feature of the evolution of nonrecombining genomes.

Figure 9 shows the effects of wsHR interference on equilibrium codon bias and nucleotide diversity for the range of 2 to 10^6 completely linked sites in a haploid population of 1000 individuals with $2N_e s = 4$ and $2N_e \mu = 0.04$. Two features are of note. First, for 10^5 linked sites, wsHR interference reduces the efficacy of selection to the extent that codon bias is only 20% of that expected under no interference. Polymorphism is reduced to 30% of that expected with no interference. These results resemble patterns of synonymous site polymorphism and divergence in the mitochondrial genome of *D. melanogaster*. The ratio of silent-site polymorphism to divergence is about half that expected from patterns in autosomal genes (data in Moriyama and Powell 1996; Nachman 1997), suggesting that there are factors reducing polymorphism in mitochondria. Likewise, patterns of codon usage indicate that selection on synonymous positions is very weak (Ballard and Kreitman 1994). Of course, these data may also be explicable by background selection, Muller's ratchet, or hitchhiking. The point is that interference between weakly selected mutations can cause low levels of diversity and the accumulation of weakly deleterious mutations in nonrecombining genomes in a manner similar to the effects of much more strongly selected mutations.

The second point of note is that the reduction in efficacy of selection scales in a much less than multiplicative way with the number of sites under selection. The ratio N_e/N caused by interference, as inferred from the level of codon bias, decreases more on the scale of the log of the number of sites under selection than the absolute number. This also holds if the effect is scaled by the observed number of segregating sites rather than

the total number of sites. We have no explanation for this pattern.

We thank Philip Awadalla, Nick Barton, Adam Eyre-Walker, Sally Otto, and Molly Przeworski for discussion and comments on the manuscript, and Richard Hudson and Molly Przeworski for the use of their simulation program. G.M. is funded by the Natural Environment Research Council and B.C. is a Royal Society Research Professor.

LITERATURE CITED

- Akashi, H., 1999 Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. *Genetics* **151**: 221–238.
- Akashi, H., and S. W. Schaeffer, 1997 Natural selection and the frequency distributions of “silent” DNA polymorphism in *Drosophila*. *Genetics* **146**: 295–307.
- Ballard, J. W., and M. Kreitman, 1994 Unraveling selection in the mitochondrial genome of *Drosophila*. *Genetics* **138**: 757–772.
- Barton, N. H., 1995 Linkage and the limits to natural selection. *Genetics* **140**: 821–841.
- Bauer, V. L., and C. F. Aquadro, 1997 Rates of DNA sequence evolution are not sex biased in *Drosophila melanogaster* and *D. simulans*. *Mol. Biol. Evol.* **14**: 1252–1257.
- 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.
- Birky, C. W., and J. B. Walsh, 1988 Effects of linkage on rates of molecular evolution. *Proc. Natl. Acad. Sci. USA* **85**: 6414–6418.
- 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 polymorphism. *Genetics* **140**: 783–796.
- Bulmer, M. G., 1991 The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**: 897–907.
- Charlesworth, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**: 213–227.
- Charlesworth, B., 1996 The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**: 149–162.
- Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Charlesworth, D., B. Charlesworth and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. *Genetics* **141**: 1619–1632.
- Chiappello, H., F. Lisacek, M. Caboche and A. Henaut, 1998 Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene* **209**: GC1–GC38.
- Comeron, J., M. Kreitman and M. Aguadé, 1999 Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* **151**: 239–249.
- Duret, L., and D. Mouchiroud, 1999 Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila* and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **96**: 4482–4487.
- Ewens, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, Berlin.
- Eyre-Walker, A., and M. Bulmer, 1995 Synonymous substitution rates in enterobacteria. *Genetics* **140**: 1407–1412.
- Felsenstein, J., 1974 The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- 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.
- Golding, G. B., 1997 The effect of purifying selection on genealogies, pp. 271–285 in *Progress in Population Genetics and Human Evolution*, edited by P. Donnelly and S. Tavaré. Springer-Verlag, New York.
- Graves, J. A. M., 1994 Evolution of the mammalian Y-chromosome. *Cytogenet. Cell Genet.* **67**: 401.
- Haigh, J., 1978 The accumulation of deleterious genes in a population—Muller’s ratchet. *Theor. Popul. Biol.* **14**: 251–267.
- Hartl, D. L., E. N. Moriyama and S. A. Sawyer, 1994 Selection intensity for codon bias. *Genetics* **138**: 227–234.
- Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. *Genet. Res.* **8**: 269–294.
- Hudson, R. R., and N. L. Kaplan, 1994 Gene trees with background selection, pp. 140–153 in *Non-neutral Evolution*, edited by B. Golding. Chapman and Hall, New York.
- Kimura, M., 1971 Theoretical foundation of population genetics at the molecular level. *Theor. Popul. Biol.* **2**: 174–208.
- Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, United Kingdom.
- Kimura, M., and T. Maruyama, 1966 The mutational load with epistatic gene interactions in fitness. *Genetics* **54**: 1337–1351.
- Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**: 1239–1258.
- Kondrashov, A., 1995 Contamination of the genome by very slightly deleterious mutations: Why have we not died 100 times over? *J. Theor. Biol.* **175**: 583–594.
- Krone, S. M., and C. Neuhäuser, 1997 Ancestral processes with selection. *Theor. Popul. Biol.* **51**: 210–237.
- Li, W.-H., 1987 Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J. Mol. Evol.* **24**: 337–345.
- Li, W.-H., and D. Graur, 1991 *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- Lynch, M., 1997 Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol. Biol. Evol.* **14**: 914–925.
- Lynch, M., and J. L. Blanchard, 1998 Deleterious mutation accumulation in organelle genomes. *Genetica* **102/103**: 29–39.
- Maynard Smith, J., 1978 *The Evolution of Sex*. Cambridge University Press, Cambridge, United Kingdom.
- Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- Maynard Smith, J., N. H. Smith, M. O’Rourke and B. G. Spratt, 1993 How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**: 4384–4388.
- McVean, G. A. T., and B. Charlesworth, 1999 A population genetic model for the evolution of synonymous codon usage: patterns and predictions. *Genet. Res.* **74**: 145–158.
- Moran, N., 1996 Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **93**: 2873–2878.
- Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**: 261–277.
- Moriyama, E., and J. Powell, 1998 Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res.* **26**: 3188–3193.
- Muller, H. J., 1964 The relation of recombination to mutational advance. *Mutat. Res.* **1**: 2–9.
- Nachman, M. W., 1997 Deleterious mutations in animal mitochondrial DNA. *Genetica* **102/103**: 61–69.
- Neuhäuser, C., and S. M. Krone, 1997 The genealogy of samples in models with selection. *Genetics* **145**: 519–534.
- Otto, S. P., and N. Barton, 1997 The evolution of recombination: removing the limits to natural selection. *Genetics* **147**: 879–906.
- Powell, J. R., and E. N. Moriyama, 1997 Evolution of codon usage bias in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **94**: 7784–7790.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992 *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, United Kingdom.
- Price, G. R., 1970 Selection and covariance. *Nature* **227**: 520–521.
- Przeworski, M., B. Charlesworth and J. D. Wall, 1999 Genealogies and weak purifying selection. *Mol. Biol. Evol.* **16**: 246–252.
- Rice, W. R., 1996 Evolution of the Y sex chromosome in animals. *BioScience* **46**: 331–343.
- Robertson, A., 1961 Inbreeding in artificial selection programmes. *Genet. Res.* **2**: 189–194.
- Santiago, E., and A. Caballero, 1995 Effective size of populations under selection. *Genetics* **139**: 1013–1030.
- Santiago, E., and A. Caballero, 1998 Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics* **149**: 2105–2117.
- Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. *Genetics* **132**: 1161–1176.
- Sharp, P. M., and W.-H. Li, 1986 An evolutionary perspective on

- synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**: 28–38.
- Sharp, P. M., T. M. E. Tuohy and K. R. Mosurski, 1986 Codon usage in yeast—cluster-analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* **14**: 5125–5143.
- Shields, D. C., P. M. Sharp, D. G. Higgins and F. Wright, 1988 “Silent” sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5**: 704–716.
- Stenico, M., A. T. Lloyd and P. M. Sharp, 1994 Codon usage in *Caenorhabditis elegans*—delineation of translational selection and mutational biases. *Nucleic Acids Res.* **22**: 2437–2446.
- Stephan, W., B. Charlesworth and G. McVean, 1999 The effect of background selection at a single locus on weakly selected, linked variants. *Genet. Res.* **73**: 133–146.
- Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- Wright, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Communicating editor: N. Takahata