Inferring the Fitness Effects of DNA Mutations From Polymorphism and Divergence Data: Statistical Power to Detect Directional Selection Under Stationarity and Free Recombination

Hiroshi Akashi

Section of Evolution and Ecology, University of California, Davis, California 95616

Manuscript received May 12, 1998
Accepted for publication September 23, 1998

ABSTRACT

The fitness effects of classes of DNA mutations can be inferred from patterns of nucleotide variation. A number of studies have attributed differences in levels of polymorphism and divergence between silent and replacement mutations to the action of natural selection. Here, I investigate the statistical power to detect directional selection through contrasts of DNA variation among functional categories of mutations. A variety of statistical approaches are applied to DNA data simulated under Sawyer and Hartl's Poisson random field model. Under assumptions of free recombination and stationarity, comparisons that include both the frequency distributions of mutations segregating *within* populations and the numbers of mutations fixed *between* populations have substantial power to detect even very weak selection. Frequency distribution and divergence tests are applied to silent and replacement mutations among five alleles of each of eight *Drosophila simulans* genes. Putatively "preferred" silent mutations segregate at higher frequencies and are more often fixed between species than "unpreferred" silent changes, suggesting fitness differences among synonymous codons. Amino acid changes tend to be either rare polymorphisms or fixed differences, consistent with a combination of deleterious and adaptive protein evolution. In these data, a substantial fraction of both silent and replacement DNA mutations appear to affect fitness.

THE evolutionary fate of a DNA sequence mutation is governed by genetic drift, demographic processes, and natural selection acting directly on the mutation or indirectly through its effect on closely linked mutations. Distinguishing among the roles of each of these factors in patterning within- and between-species genetic variation is a central goal of population genetics (Lewontin 1974). In particular, the shape of the distribution of fitness effects of mutations remains a contentious issue (Kimura 1983; Gillespie 1991; Ohta 1992; Takahata 1996).

A number of approaches attempt to infer evolutionary processes by comparing patterns of DNA variation from a given genetic region to those predicted under a specified evolutionary model (*i.e.*, Watterson 1978; Strobek 1987; Tajima 1989; Hudson *et al.* 1992, 1994; Fu and Li 1993; Braverman *et al.* 1995; Simonsen *et al.* 1995; Fu 1996, 1997; Kelly 1997). The most common null model assumes an infinite number of mutable sites, no natural selection, no recombination, a stationary frequency distribution of segregating mutations, and a Wright-Fisher demographic model. Rejection of the null hypothesis can be caused by a number of departures from these assumptions including changes in population size, population subdivision, genetic linkage to adaptive and deleterious mutations, and selection on the mutations themselves (Strobek 1987; Hudson *et al.* 1992; Braverman *et al.* 1995; Charlesworth *et al.* 1995; Simonsen *et al.* 1995; Fu 1996, 1997). However, distinguishing between the contributions of demographic history and natural selection to a given departure from the null can be difficult. For example, patterns of neutral DNA sequence variation closely linked to a site that has undergone a recent adaptive substitution or "selective sweep" are similar to those in an expanding population (Simonsen *et al.* 1995). Alternatively, patterns of neutral DNA variation linked to a site at which a polymorphism is maintained by balancing selection can be similar to sequence variation sampled from subdivided populations (Hudson 1990).

A second class of approaches compares patterns of DNA variation between two or more genetic regions (Hudson *et al.* 1987; McDonald 1996, 1998; Hey 1997). Such comparisons attempt to distinguish between the effects of demographic history, which should have a roughly equal impact throughout the genome, and natural selection, whose effect may be more localized. Although the statistical power of this approach to detecting particular scenarios of evolution has not been investigated, regional differences in levels of polymorphism, or in the frequency distributions of mutations, could result from either balancing selection elevating linked neutral variation or directional selection reduc-

Address for correspondence: Department of Ecology and Evolutionary Biology, Haworth Hall, University of Kansas, Lawrence, KS 66045-2106. E-mail: hakashi@falcon.cc.ukans.edu

ing neutral polymorphism. However, between-region comparisons neither identify the particular site(s) under selection nor address what fraction of segregating or fixed mutations affects fitness.

Comparisons of evolutionary patterns between categories of mutations interspersed *within* a genetic region attempt to identify the direct action of natural selection. If the classes of mutations (such as replacement and silent changes) are randomly interspersed within a genetic region, then population level effects and selection at linked sites are expected to have a roughly equivalent impact on mutations in the two classes (Hudson 1993). Thus, differences in the frequency distributions of polymorphic mutations (Sawyer *et al.* 1987) and in the ratios of polymorphic and fixed mutations (McDonald and Kreitman 1991; Templeton 1996; Akashi 1997a) should reflect differences in the fitness effects of the mutations. A number of claims of adaptive (McDonald and Kreitman 1991; Eanes *et al.* 1993; Long and Langley 1993; Karotam *et al.* 1995; King 1998), deleterious (Sawyer *et al.* 1987; Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Akashi 1996; Templeton 1996; Wise *et al.* 1998), and balancing (Wayne *et al.* 1996) selection on amino acid variants, mutation-selection-drift at silent sites (Ballard and Kreitman 1994; Akashi 1995, 1997a; Akashi and Schaeffer 1997), and deleterious effects of transposable element insertions (Golding *et al.* 1986) rely on such comparisons.

Although a growing number of studies are inferring evolutionary processes from comparisons among interspersed mutations, the sensitivity and robustness of this approach to detecting selection have not been examined. Here, I investigate the statistical power to detect directional selection through comparisons of patterns of variation between putative fitness classes of DNA mutations. All results are obtained under Sawyer and Hartl's Poisson random field model (Sawyer and Hartl 1992; Hartl *et al.* 1994) under assumptions of stationarity, free recombination, and independent fitness effects (results under departures from these assumptions will be addressed in a separate study). The performance of a number of statistical tests is compared over a wide range of selection intensities and sample sizes of DNA sequences. Such tests are applied to DNA sequence data from eight *Drosophila simulans* genes to determine the contribution of natural selection in silent and protein evolution.

NATURAL SELECTION AND THE EXPECTED CONFIGURATIONS OF MUTATIONS

Kimura and Ohta treat gene frequency changes within populations and the accumulation of fixed differences between populations as two facets of an underlying process of evolution under relatively constant mutation rates, effective population sizes, and (for some mutations) directional selection (Kimura and Ohta 1971; Kimura 1983). Positive selection increases the probability that a mutation will rise in frequency in each generation, whereas negative selection has the opposite effect. However, except for very strong selection, the time scale of the process is on the order of N_e, the effective population size, generations. Inferring the fitness effects of mutations by measuring the evolutionary trajectories (frequency changes) of individual mutations in laboratory or natural populations requires relatively strong deterministic forces ($s > 0.001$; Dykhuizen and Hartl 1983). In large populations, much weaker selection ($s > 1/N_e$) can have an important impact in long-term evolution (Fisher 1930; Wright 1931; Kimura 1962; Ohta 1973) but cannot be measured directly. For such mutations, the trajectories of the mutations, and thus their fitness effects, can be inferred by sampling two aspects of the evolutionary process. A "snapshot" of evolution can be obtained by comparing alleles from within a population; the data consist of the number of segregating sites and their frequencies in the sample. The numbers of mutations "fixed" between an outgroup and the most recent common ancestor of the population sample can be inferred by examining sequences from closely related species.

Consider an aligned set of DNA sequences from m individuals from a population and at least one sequence from an outgroup. Assume an infinite number of mutable sites in these sequences so that all mutations occur at unique sites. Assume first that ancestral and derived nucleotides can be determined at sites that vary in the sample and at sites that differ between the sample and the outgroup. At a given variable site, the nonancestral nucleotide will be found in $r = 1$ to m of the sequences. The distribution of nonancestral nucleotides falling into the r frequency classes will be referred to as the "configuration" of mutations [to distinguish the pattern from the "frequency distribution" of mutations that is often used to describe polymorphic mutations ($r = 1$ to $m - 1$)]. Mutations in frequency class m will be referred to as "fixed" between the sample and the outgroup. In the absence of information about the ancestral and derived nucleotides at variable sites, the configuration can be "folded-over" by pooling each pair of frequency classes $r = i$ and $r = m - i$ for all integers, $1 \leq i \leq m/2$.

Figure 1 illustrates the quantitative effects of directional selection on the expected configurations of mutations. Positive directional selection skews the configuration toward a larger proportion of mutations at high frequencies within the population or fixed in the sample (Figure 1b). Negative directional selection has the opposite effect: a greater proportion of nonancestral nucleotides segregate at low frequencies (Figure 1a). However, note that under positive selection, the total expected number of variable sites in the sample increases as a

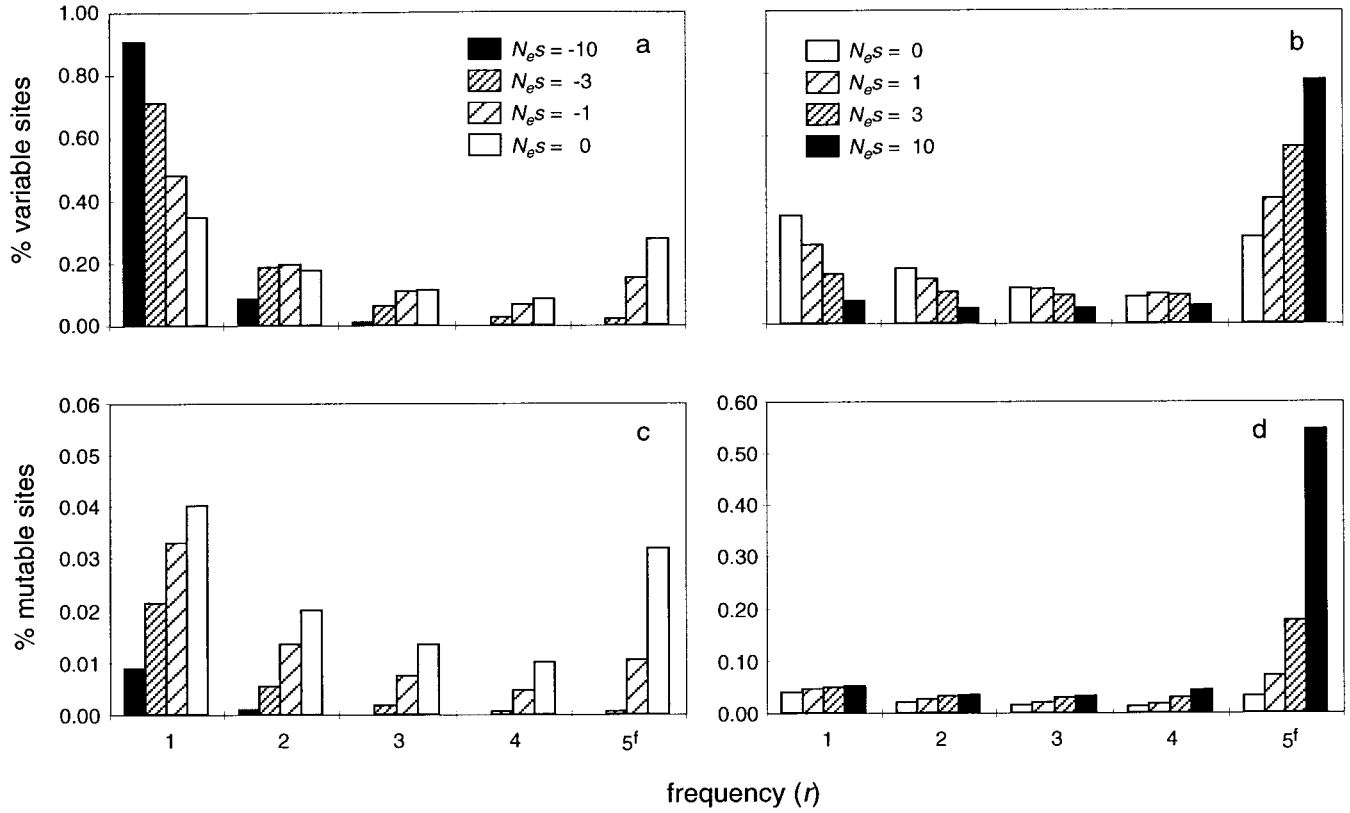

Figure 1.—Expected configurations of neutral and selected mutations. The expected numbers of newly arisen mutations at frequency classes $r = 1$ to m in a sample of sequences were calculated according to Sawyer and Hartl (1992) and Hartl *et al.* (1994). Data are shown for $m = 5$ sequences and $t_{div} = 0.6$. a and b show the expected proportion of *variable* sites in the sample at different frequencies under negative and positive selection, respectively. c and d show the proportion of *mutable* sites at which variants are expected to be segregating at different frequencies or fixed in the sample under negative and positive selection, respectively. Note that the scales for the *y*-axis differ for c and d. For $m = 5$, pooling classes $r = 1$ and $r = 4$, and $r = 2$ and $r = 3$ would give a folded-over distribution with three frequency classes. Superscript f denotes "fixed" difference class ($r = m$). In this figure, and in the following figures, the scales of unlabeled *x*- and *y*-axes are equivalent to those of graphs in the same columns and rows, respectively.

function of $N_e s$ (Figure 1d), whereas under negative selection, the opposite is true (Figure 1c).

Golding *et al.* (1986) and Sawyer *et al.* (1987) were the first to infer differences in the average fitness effects of mutations by contrasting within- and between-species variation among functional classes of DNA changes. Sawyer *et al.* (1987) suggested a simple comparison of the frequency distributions of silent and replacement polymorphisms in a 2×2 contingency table. Their data did not include an outgroup sequence, so the analyses were confined to polymorphism data with unknown ancestral and derived states (folded configurations). They divided segregating mutations into two frequency classes, "singletons" ($r = 1$ and $r = m - 1$) and other frequency classes ($1 < r < m - 1$). A departure from homogeneity in these classes among silent and replacement changes was interpreted as evidence for differences in the fitness effects of the mutations.

McDonald and Kreitman (1991) included between-species variation in a similar test of homogeneity among frequency classes for silent and replacement mutations. Their 2×2 contingency table compares the numbers of polymorphic mutations, pooled across frequency classes ($1 \leq r < m$), and the numbers of fixed differences ($r = m$). Because directional selection has a strong impact on the fixed differences class (Figure 1), including this information is likely to increase the sensitivity of the statistical approach to detect selection. However, by pooling all polymorphic mutations into a single category, McDonald and Kreitman's test sacrifices information from the frequency distribution of segregating mutations.

Templeton (1996) combined the approaches of Sawyer *et al.* (1987) and McDonald and Kreitman (1991) by expanding the contrast to a test of homogeneity across three frequency classes. The numbers of singleton polymorphisms ($r = 1$ and $r = m - 1$), polymorphisms at intermediate frequencies ($1 < r < m - 1$), and fixed differences ($r = m$) were compared between silent and replacement mutations. The statistical test examines information from both the folded frequency distribution of segregating mutations and the numbers of fixed differences, but information for polymorphic

mutations segregating at frequencies greater than one is lost by pooling such variants into a single category.

The statistical power to detect selection through these approaches could, in principle, be enhanced by including more information from the sample and by employing a statistical test that is more sensitive to deviations caused by the particular alternative hypotheses of interest (Akashi 1997a). Because even very weak selection affects the expected proportion of mutations in each frequency class, treating each class ($1 \leq r < m$) as a distinct category may increase the sensitivity of the approach. In addition, outgroup sequences can be used to infer ancestral and derived states at variable positions so that $r = i$ and $r = m - i$ classes do not have to be pooled. Finally, under the Poisson random field model, directional selection has a strong effect on the means of the configurations of mutations (Figure 1). Statistical comparisons that are more sensitive to differences in the location of distributions may be more powerful for detecting fitness effects of mutations than tests of homogeneity.

CONFIGURATION TESTS BETWEEN NEUTRAL AND SELECTED MUTATIONS

Sawyer *et al.* (1987) identified the fitness effects of amino acid mutations by comparing the configurations of silent and replacement mutations within coding regions in DNA. Assuming neutral evolution at silent sites, a configuration of amino acid mutations skewed toward an excess of rare polymorphisms reflects deleterious amino acid mutations, whereas an excess of common or fixed amino acid differences supports adaptive protein evolution. The analyses below investigate the statistical power to detect selection through such comparisons between neutral and selected mutations.

Sawyer and Hartl's Poisson random field model allows relatively straightforward simulation of DNA variation data under directional selection. The model assumes a Wright-Fisher population of haploid individuals, an infinite number of mutable sites, a stationary frequency distribution of segregating mutations, and independent evolution at all sites (free recombination and independent fitness effects of mutations). Under these assumptions, the numbers of mutations in each frequency class in the configuration ($r = 1$ to m) are independent Poisson random variables whose means can be calculated according to the equations of Sawyer and Hartl (1992) and Hartl *et al.* (1994). These means are a function of five parameters: N_e, the species effective population size; u, the expected number of mutations per nucleotide site, per generation; l, the number of aligned nucleotide sites; m, the number of alleles sampled from a given population; t_{div}, the time of divergence between the population sampled and the outgroup (scaled to N_e generations); and s, the selective effect of mutations. Note that number of "alleles" refers to the number of chromosomes sampled from a population regardless of whether any pairs of the DNA sequences are identical or not.

In the following power tests, N_e and u were fixed and the other parameters were varied over a range of interest. $N_e = 10^6$ (Kreitman 1983) and $u = 10^{-9}$ (Moriyama 1987; Rowan and Hunt 1991) correspond to rough estimates for these parameters in Drosophila. Statistical power was examined for $m = 5$, 10, 25, and 50 alleles and for $l = 500$, 1000, 2500, and 5000 mutable sites. Selection coefficients were varied between $-100 \leq N_e s \leq 100$, and the time of divergence between the sampled alleles and the outgroup was varied between $t_{div} = 0.6$, 1.2, 2.4, and 4.8. The lower t_{div} value was that estimated from intron polymorphism and divergence data in *D. simulans* since its split with its sister species, *D. melanogaster* (see Akashi and Schaeffer 1997). The upper value corresponds to ~10% expected divergence at neutral sites. For a given set of parameter values, mutations at half of the sites were neutral and the other half were selected. Simulations were also conducted for 20% neutral and 80% selected sites and for 80% neutral and 20% selected sites. In all simulations, selection coefficients were uniform for all mutations within each category.

For each set of five parameters, the expected values of the numbers of neutral and selected mutations in each frequency class in the configuration were calculated according to the equations of Table 2 of Sawyer and Hartl (1992) and Equation 2 of Hartl *et al.* (1994). The algorithm of Press *et al.* (1992, p. 293) was used to generate 1000 simulated data sets by sampling integers from Poisson distributions with these expected values as their means. All simulations were written in the C computer language and run on Macintosh and Pentium desktop computers.

A variety of statistical tests was applied to each of the simulated data sets. For polymorphism (frequency distribution) data, the tests examined were as follows: Sawyer *et al.*'s (1987) 2×2 test of independence with frequency classes $r = 1$ and $1 < r < m$, a $2 \times (m - 1)$ test of independence for all the frequency classes of polymorphic mutations and a Mann-Whitney *U*-test (MWU) for all the frequency classes (see Table 1 for abbreviations). For unfolded distributions, ancestral and derived states were assumed to be inferred without error.

For the tests of homogeneity, the probability of the data under the null hypothesis of independence was estimated through a Monte Carlo approach. For each simulated $2 \times n$ table, the product of each cell value and its natural logarithm was summed across all cells to give a test statistic. The test statistic was also calculated for 1000 randomized tables. In a generalization of the Fisher exact test, the joint probability of cell values was assumed to be the joint hypergeometric probabilities of the cells under homogeneity for the same marginal values as the simulated table (see appendix). For each simulated table, 1000 random tables were generated

TABLE 1

Statistical tests comparing the configurations of mutations

Abbreviation	Statistical test	Frequency classes	Reference
fd2MCH	2×2 Monte Carlo test of homogeneity	$r = 1, 1 < r < m$	Sawyer *et al.* (1987)
fdMCH	$2 \times (m - 1)$ Monte Carlo test of homogeneity	$r = 1, r = 2, \ldots, r = m - 1$	
fdMWU	Mann-Whitney *U*-test	$r = 1, r = 2, \ldots, r = m - 1$	Akashi and Schaeffer (1997)
TD	Tajima's *D*-test	Not applicable	Tajima (1989)
pdMCH	2×2 Monte Carlo test of homogeneity	$1 \leqq r < m, r = m$	McDonald and Kreitman (1991)
sidMCH	2×3 Monte Carlo test of homogeneity	$r = 1, 1 < r < m, r = m$	Templeton (1996)
fddMCH	$2 \times m$ Monte Carlo test of homogeneity	$r = 1, r = 2, \ldots, r = m$	
fddMWU	Mann-Whitney *U*-test	$r = 1, r = 2, \ldots, r = m$	Akashi (1997a)

Frequency classes assume that ancestral and derived states are inferred at variable nucleotide sites (unfolded distributions). The original tests suggested by Sawyer *et al.* (1987), McDonald and Kreitman (1991), and Templeton (1996) were performed on folded distributions. Monte Carlo tests of homogeneity were employed in lieu of *G*-tests, Fisher's exact tests, and χ^2 tests employed in the references.

from these joint hypergeometric probabilities. The fraction of these random tables with a test statistic equal to, or greater than, that observed in the sample was used as the estimate of the two-tailed probability of the observed data under the null hypothesis of homogeneity in their configurations (the procedure follows that of B. Engels, personal communication). In these simulations, the first column of the table is generated under the assumption of selective neutrality. Rejection of the null hypothesis indicates that the test has detected significant selective effects on the distribution of cell counts in the second column of the table.

Tajima's (1989) *D*-test was also applied to the simulated nonneutral class of variation. This procedure compares polymorphism data to expectations under an equilibrium, neutral, no recombination model. The critical values of the test statistic were taken from Tajima's (1989) tables. The *D*-test differs from the other statistical tests described above because the test compares a single class of mutations to a null model and because the critical values that are commonly employed assume no recombination rather than free recombination. Tajima's test was included for comparison because it is often applied to data from recombining regions of DNA (*i.e.*, Tajima 1989; Moriyama and Powell 1996) and because its power to detect the fitness effects of mutations has not been addressed.

For the tests restricted to polymorphism data, the statistical power to detect both negative and positive selection is shown in Figures 2 and 3. The power to reject the null hypothesis generally increases as a function of the absolute value of $N_e s$, but decreases for large negative values (Figure 3). The cause of this pattern is apparent from Figure 1; although the location of the distribution of mutations continues to change as selection becomes stronger, the sample size of nonneutral polymorphisms decreases to zero. For 25 alleles of 1250 neutral and selected sites, however, the power to detect even very strong negative selection is considerable. Among tests of homogeneity, the 2×2 test is more sensitive to negative selection, whereas the $2 \times (m - 1)$ test is generally more powerful for $N_e s > 0$. This appears to reflect the lack of information in the higher frequency classes for mutations under negative selection (Figure 1). Tajima's *D*-test performs poorly for small numbers of alleles, but is quite sensitive to negative directional selection when the number of sampled alleles is large. Positive directional selection has a smaller impact on the frequency distribution of mutations (Figure 1) and was not detectable by the Tajima *D*-test under any of the parameter values examined.

The polymorphism tests show different sensitivities to changes in the examined numbers of alleles and numbers of sites. Increasing the number of sites has a larger impact on the power of tests of independence and fdMWU tests, whereas the Tajima test gains considerably from increasing the number of sampled alleles. For the parameter ranges considered, the fdMWU test is at least as powerful, and is often considerably more powerful, than the other polymorphism tests for detecting both positive and negative directional selection.

The expected proportion of variable sites in the fixed differences class ($r = m$) is very sensitive to selection (Figure 1). Thus, adding divergence data to comparisons of the configurations of mutations is likely to add substantial power to detect natural selection. Four such tests were applied to simulated data. Three different tests of independence were employed, the 2×2 polymorphism ($1 \leq r < m$) and divergence ($r = m$) test of McDonald and Kreitman (1991), the 2×3 singleton ($r = 1$), intermediate frequency ($1 < r < m$), and divergence ($r = m$) test of Templeton (1996), and the $2 \times m$ frequency distribution and divergence test for unpooled frequency classes. These tests were performed as Monte Carlo tests of homogeneity (MCH) as described above. Mann-Whitney *U*-tests were also applied to the *m* frequency classes (see Table 1 for abbreviations for tests). The equivalent tests and the fdMWU test were also per-

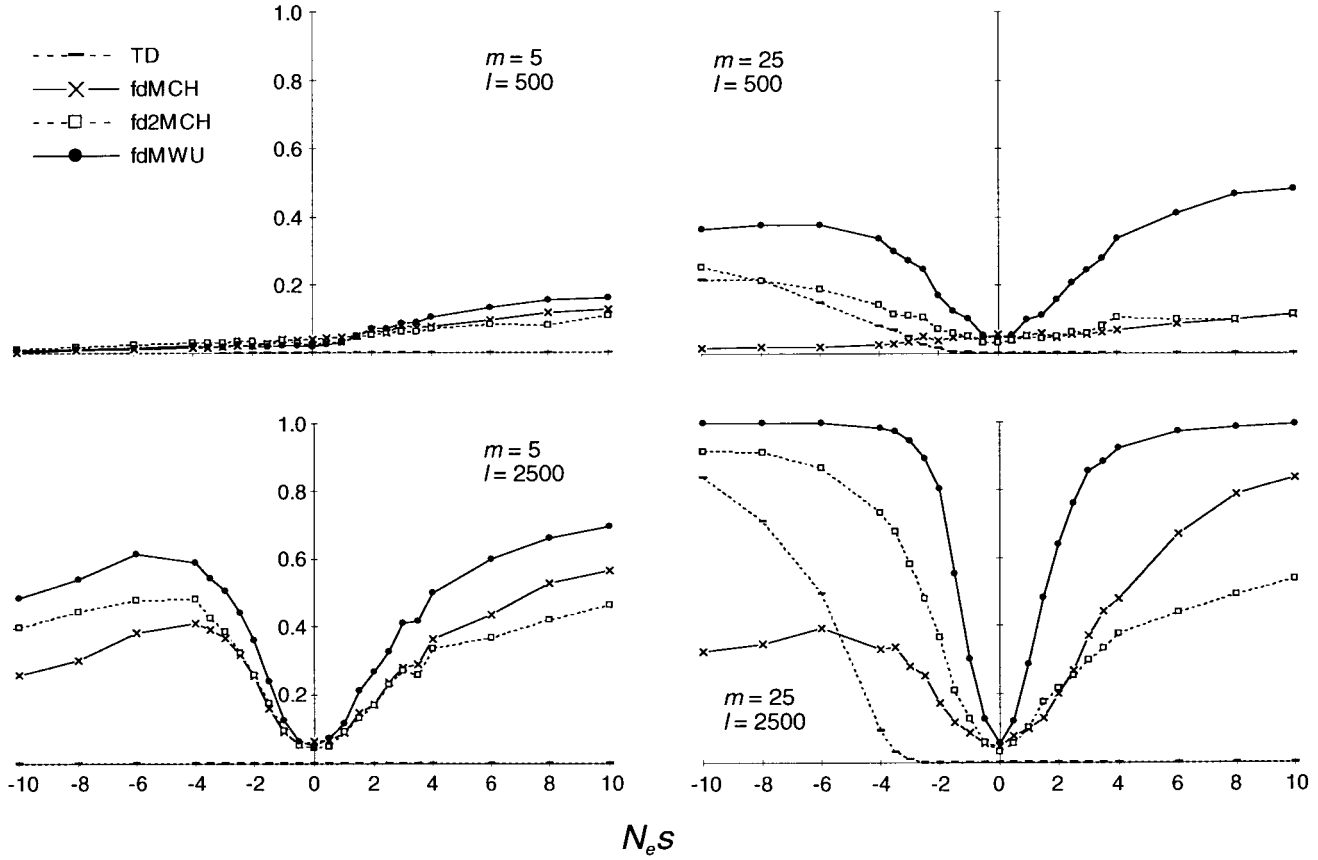

Figure 2.—Power of polymorphism configuration tests between neutral and weakly selected mutations. The *y*-axis plots the proportion of tests that reject fitness equivalence, $P < 0.05$, among 1000 simulated data sets for each value of $N_e s$. See Table 1 for abbreviations for tests. Because each point on the graph reflects a proportion from a random binomial sample of size $n = 1000$, the 95% confidence intervals for the true values are $\hat{p} \pm 2[\hat{p}(1 - \hat{p})/1000]^{1/2}$. The relative order of the power of these tests (and those of Figures 3 through 6) was similar for the simulations under unequal numbers of neutral and selected sites described in the text.

formed on folded distributions (ancestral and derived states not inferred). The power of statistical tests that include divergence data is shown in Figures 4 and 5.

As expected, inclusion of the fixed difference class adds a great deal of power to detect natural selection, especially for $N_e s > 0$. Among the tests of independence, the 2×2 test is relatively insensitive to deleterious evolution but is roughly equivalent to the 2×3 test for positive selection. The $2 \times m$ test of homogeneity was considerably less powerful than the 2×3 test over almost the entire range of parameters examined. Apparently, the higher frequency cells increase the degrees of freedom in the statistical test but contribute little to the test statistic both because the expected and observed values are not sufficiently different and because the values in the cells are small (Figure 1). For a small number of sites and a large number of alleles, the fdMWU test can be more sensitive to negative selection than tests of independence that include divergence. However, the fddMWU test is either equivalent to, or more powerful than, all the other tests over all parameter values examined. The difference in power is most notable when the number of alleles is large and when selection coefficients are small. For tests that include divergence comparisons the number of sampled sites generally has a much larger impact on statistical power than increasing the number of sampled alleles (Figures 4 and 5). Examining the unfolded distributions of newly arisen mutations can have a substantial effect on power when the numbers of sampled alleles is small and when selection is weak (Figure 6).

It is important to note, however, that the results above hold only for the given model of evolution under the parameters examined. Although these findings probably hold for parameters within the range investigated, the superior power of the MWU test over the homogeneity tests is due to the particular deviation from the null investigated here. Under uniform selection, the location of the configuration of mutations undergoes a unilateral shift as a function of selection intensity (Figure 1). However, other alternative models may show differences in the configurations of mutations that may result in smaller deviations in their means. The choice of tests, therefore, depends on the particular alternatives under consideration. If a model predicts differences in the locations of distributions, then MWU tests may provide

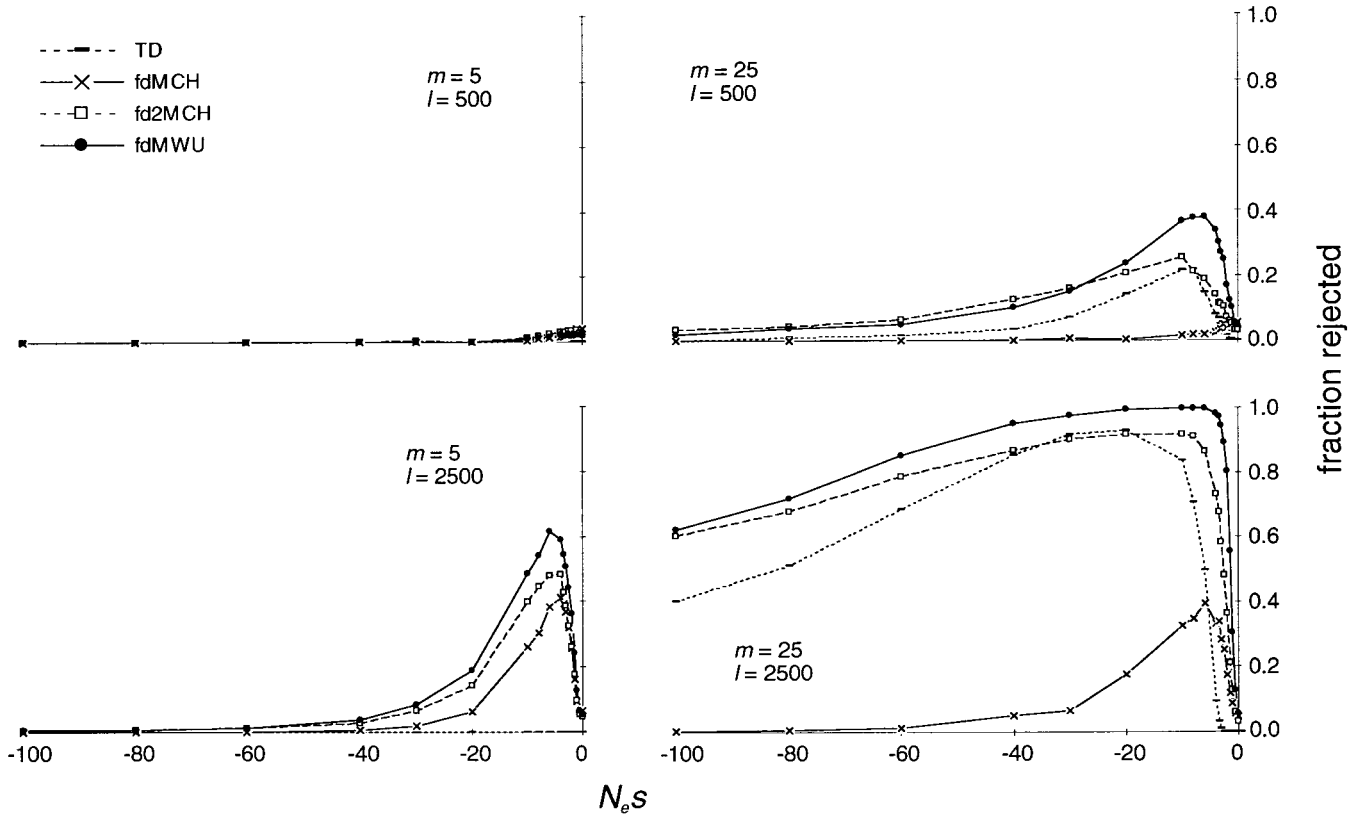

Figure 3.—Power of polymorphism configuration tests between neutral and deleterious mutations. Plots are equivalent to those of Figure 2 for strongly deleterious mutations.

the greatest statistical power to detect selection. In the absence of a particular alternative hypothesis, Templeton's sidMCH test is a general method to test for departures from the null hypothesis of equivalent configurations among classes of mutations. (Similar comparisons that combine the frequency distribution of polymorphic mutations and the number of fixed differences could also enhance the sensitivity of between-region comparisons of DNA variation, but the power of such tests has not been investigated.)

The evolutionary distance, t_{div}, between the alleles sampled from within a population and the outgroup sequence can have a large impact on the power to detect selection. Figure 7 shows the effect of times of divergence on the pdMCH and fddMWU tests. Increasing t_{div} increases the sample sizes of fixed differences resulting in an increase in the power of all the tests that include this information and a decrease in the differences among these tests. However, these results assume accurate counting of the numbers of substitutions (under the infinite sites model, the number of diverged sites equals the number of substitutions). In practice, the number of substitutions is inferred given the number of differences between extant sequences and an evolutionary model that determines the appropriate correction for the number of sites that have undergone multiple substitutions. At higher levels of divergence, estimation of substitution rates can be quite sensitive to the assumed model of evolution (Nei and Gojobori 1986; Goldman and Yang 1994; Rzhetsky and Nei 1995; Ina 1996; Muse 1996). Accurate estimation requires knowledge of transition probabilities among different nucleotides (or codons) and of how these probabilities vary among sites and over time. In the absence of such knowledge, shorter divergence times allow more reliable inference of the numbers of evolutionary fixations at the expense of some statistical power. In addition, ancestral and derived states at variable nucleotide positions can be inferred with greater confidence when levels of evolutionary divergence among the sequences are low (Collins *et al.* 1994; Frumhoff and Reeve 1994; Yang *et al.* 1995; Schluter *et al.* 1997; Zhang and Nei 1997).

These simulation data suggest that, under the assumptions of the Sawyer-Hartl model, comparisons of the configurations of neutral and selected mutations have considerable statistical power to detect even very weak positive and negative selection. This inference of selection is dependent on two steps. A difference in the configurations of two categories of mutations suggests different distributions of their fitness effects. If one of the categories of mutations evolves neutrally, then, under a constant directional selection model, the *sign* of the fitness effects of the second class of mutations can be inferred from the location of its distribution relative to

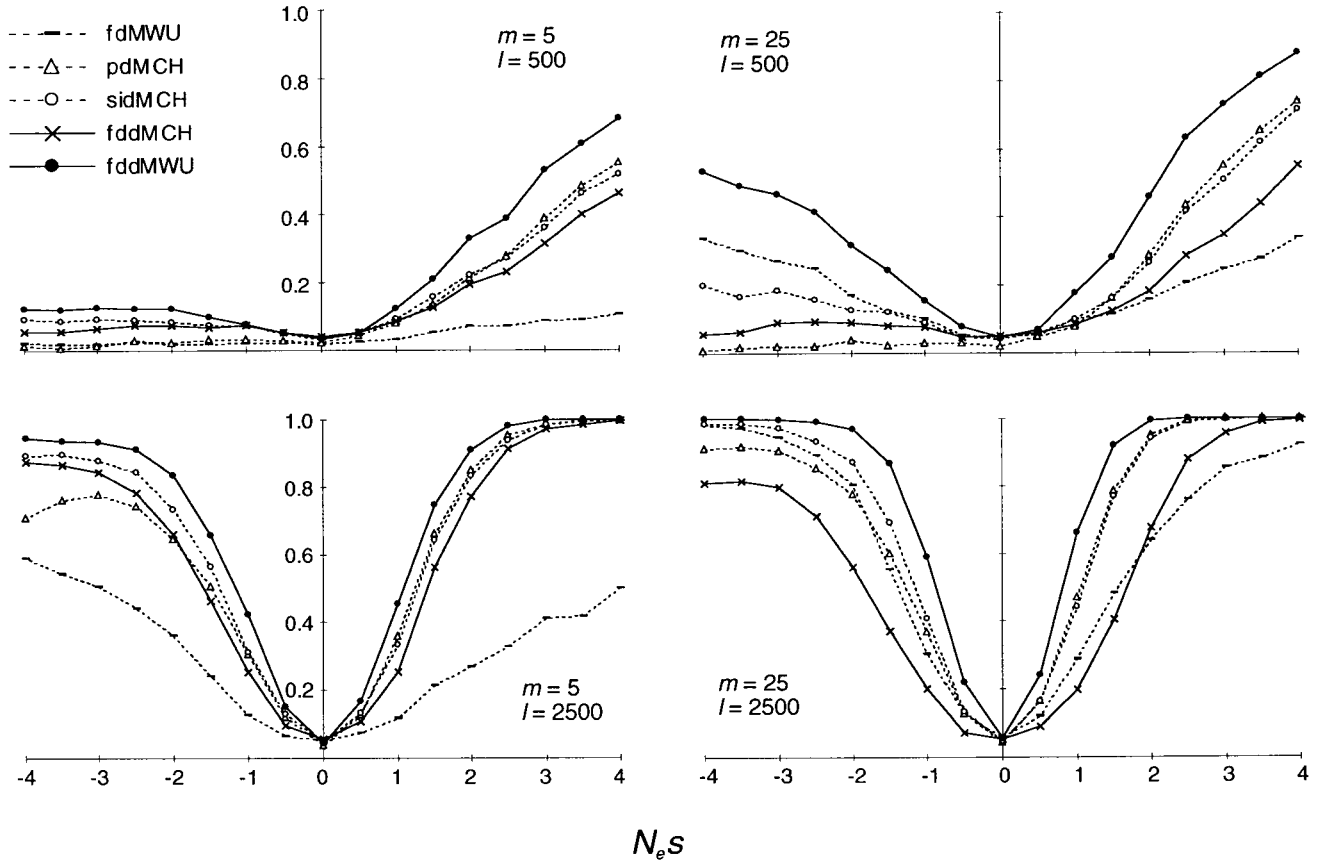

Figure 4.—Power of polymorphism and divergence configuration tests between neutral and weakly selected mutations. The *y*-axis plots the proportion of tests that reject fitness equivalence, $P < 0.05$, among 1000 simulated data sets for each value of $N_e s$. See Table 1 for abbreviations for tests. Templeton (1996) also suggested a "young *vs.* old" mutations 2×2 test of independence between singletons ($r = 1$ and $r = m - 1$) and all other frequency classes ($1 < r < m - 1$ and $r = m$). The power of this test was examined but the results are not shown. This test is sensitive to strong deleterious selection, especially for small sample sizes, but has little power to detect positive selection coefficients. For some parameters, the power of this test decreases with increasing numbers of alleles.

that for the neutral class. An excess of rare polymorphisms, relative to the neutral class, suggests negative selection coefficients, whereas too many fixed differences suggest adaptive evolution. The assumption of neutrality at silent sites in coding regions is critical to such inferences of selection in protein evolution. The following section employs simulation data to examine the statistical power to detect a particular model of selection at silent sites and compares the configurations of putative fitness classes of silent DNA mutations in *D. simulans*.

CONFIGURATION TESTS OF MUTATION-SELECTION-DRIFT

Patterns of codon usage in a number of organisms are consistent with natural selection discriminating among synonymous codons to enhance the efficiency and/or the accuracy of protein synthesis (reviewed in Ikemura 1985; Andersson and Kurland 1990; Sharp *et al.* 1995). Under "major codon preference," codon usage bias is maintained by a balance among the forces of mutation pressure, genetic drift, and natural selection favoring translationally superior major codons (Sharp and Li 1986; Li 1987; Bulmer 1988, 1991). The simplest evolutionary model of this scenario considers twofold redundant codons in a haploid organism (Li 1987; Bulmer 1991). Mutations occur at rates v from nonmajor codons to major codons and u in the opposite direction. Major codons confer selective advantage s. This scenario is depicted below:

$$\underset{+s}{\text{major}} \underset{v}{\overset{u}{\rightleftarrows}} \underset{-s}{\text{minor}}$$

Consider a locus consisting of a number of such sites. The proportion of major codons at the locus is determined by u/v, the ratio of the mutation rates, and $N_e s$, the product of effective population size and selection coefficient. If these parameters remain relatively constant, then the proportion of major codons at the locus will reach a steady state (*i.e.*, equal numbers of forward and backward substitutions).

Major codon preference predicts two fitness classes

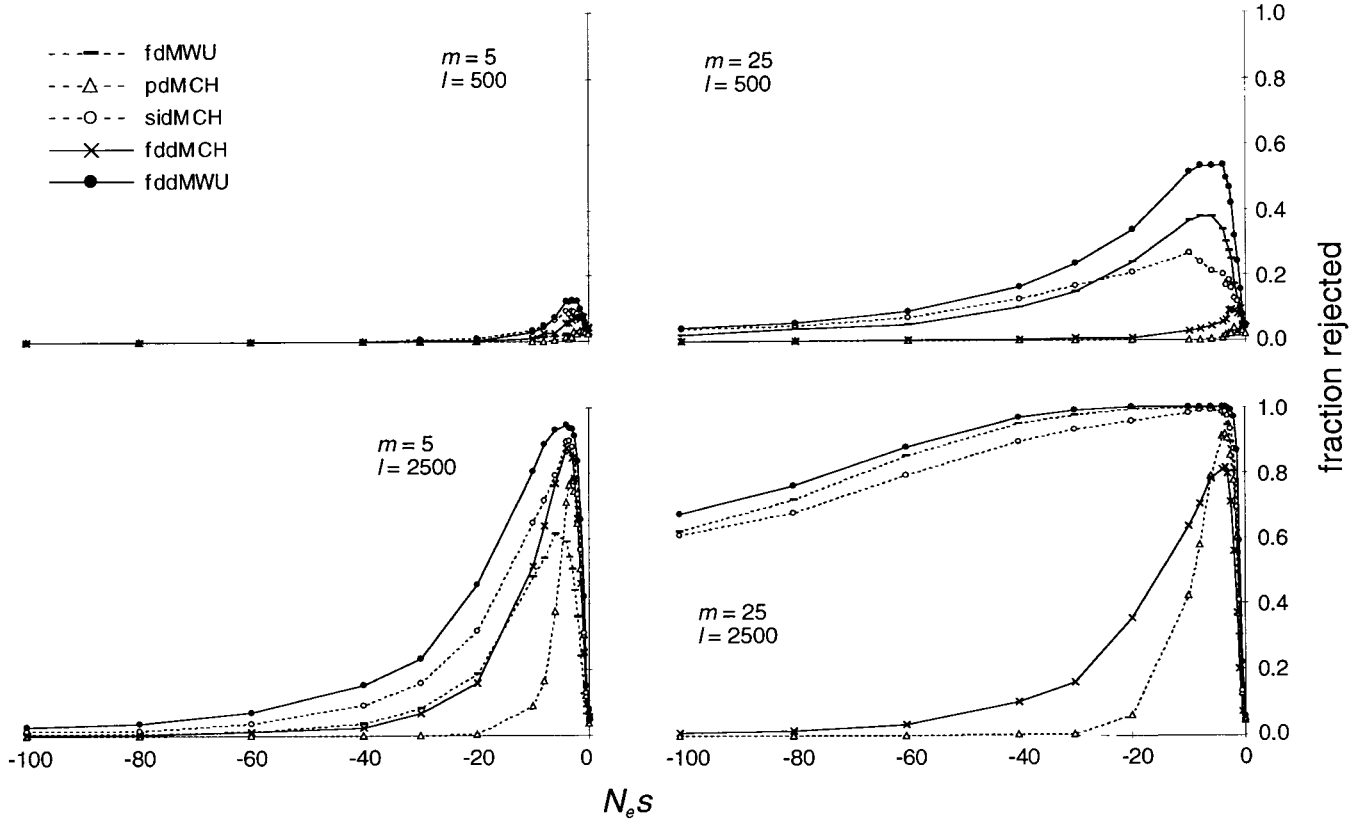

Figure 5.—Power of polymorphism and divergence configuration tests between neutral and deleterious mutations. The plots are equivalent to those of Figure 4 for strongly deleterious mutations.

of silent mutations, "preferred" mutations from nonmajor to major codons and "unpreferred" mutations in the opposite direction (Akashi 1995). A purely mutational model of codon bias requires differences in the forward and backward mutations rates (Freese 1962; Sueoka 1962, 1988), but does not predict differences in the evolutionary configurations of mutations in the two directions, because both are neutral.

The statistical power to detect major codon preference at silent sites can be examined under the Sawyer-Hartl model. Li (1987) and Bulmer (1991) give expressions for the steady-state proportion of major codons in a given gene under assumptions of constant $N_e s$ and independent evolution among sites. This proportion is determined by four parameters: u and v, the forward and backward mutation rates; s, the selective advantage

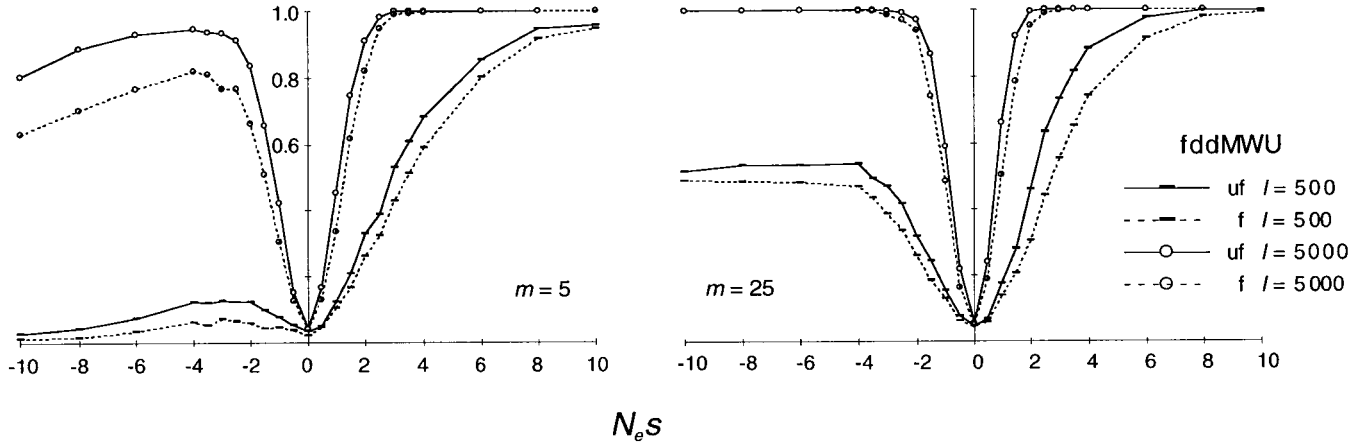

Figure 6.—Configuration tests between neutral and selected mutations for folded and unfolded distributions. The y-axis plots the proportion of tests that reject fitness equivalence, $P < 0.05$, among 1000 simulated data sets for each value of $N_e s$. The statistical power of the fddMWU test for folded (f) and unfolded (uf) configurations of mutations is shown. For unfolded configurations, ancestral and derived states are assumed to be inferred with complete accuracy. See Table 1 for abbreviations for tests.

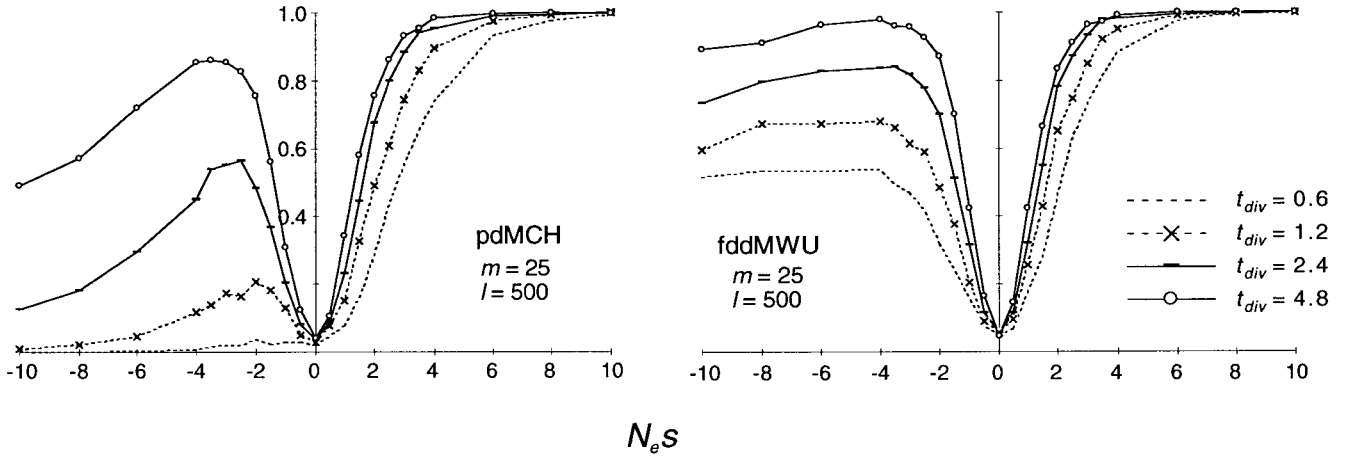

Figure 7.—The effect of t_{div} on the power of configuration tests between neutral and selected mutations. The *y*-axis plots the proportion of tests that reject fitness equivalence, $P < 0.05$, among 1000 simulated data sets for each value of $N_e s$. Results for the pdMCH and fddMWU tests are shown for $t_{div} = 0.6$, 1.2, 2.4, and 4.8. See Table 1 for abbreviations for tests.

of major codons; and N_e, the species effective population size. For these simulations, per-site mutation rates were set to $u = 1.2 \times 10^{-9}$ and $v = 0.8 \times 10^{-9}$. The ratio of the mutation rates, $u/v = 1.5$, gives an equilibrium mutational base composition of 60% A + T, the average base composition of putatively neutrally evolving introns in *D. melanogaster* (Shields *et al.* 1988; Moriyama and Hartl 1993). This base composition is also consistent with substitution patterns in presumably "dead-on-arrival" non-LTR transposable elements in Drosophila (D. Petrov, personal communication). The proportion of sites encoding major codons was calculated from Equation 6 of Bulmer (1991) given the parameters, u, v, N_e, and s. The numbers of major and nonmajor codons were determined by this proportion and the number of sites in the locus, l, and were assumed fixed (no stochastic variance) for a given set of parameter values. Per-locus mutation rates to unpreferred and preferred mutations are the product of per-site mutation rates and the numbers of major and nonmajor codons, respectively. Note that, under this model, per-locus preferred and unpreferred mutation rates are a function of the strength of selection.

Figure 8 shows the expected configurations of unpreferred and preferred mutations under major codon preference. The numbers of mutations in each frequency class in the configuration, $r = 1$ to m, can be determined from Sawyer and Hartl's sampling equations given m, the number of alleles examined, t_{div}, the time of divergence between the species sampled and the outgroup, and the five parameters discussed above. Even very weak selection can skew the configurations of the two classes of silent mutations. As selection increases the proportion of major codons in a given locus, differences in the configurations of the *proportion* of mutations in the frequency classes become more pronounced (Figure 8, a–c) but the expected *numbers* of preferred mutations decrease (Figure 8, d–f). This decrease in the per-locus preferred mutation rate will result in a loss of statistical power to detect differences between the configurations of preferred and unpreferred mutations. However, under major codon preference, observed levels of codon bias in Drosophila require selection coefficients in the range of $\sim 0 < |N_e s| < 3$. The analyses below examine the statistical power to detect differences in the configurations of the two classes of silent mutations under such a parameter range.

DNA variation data were simulated for preferred and unpreferred mutations under the Sawyer-Hartl Poisson random field model. Assuming stationary frequency distributions and independent evolution at all sites, the numbers of sampled preferred and unpreferred mutations in each frequency class are independent Poisson random variables. Simulations were conducted for the parameters described above and $l = 500$, 1000, 2500, and 5000 mutable sites and $m = 5$, 10, 25, and 50 alleles. Selection coefficients between major and nonmajor codons were varied between $0 \leq N_e s \leq 6$, and the time of divergence between the sampled alleles and the outgroup was varied between $t_{div} = 0.6$, 1.2, 2.4, and 4.8. A total of 1000 sample configurations of preferred and unpreferred mutations were simulated for each set of parameters. The pdMCH, sidMCH, and fddMCH tests of independence, and the fdMWU and fddMWU tests were applied to each simulated data set.

Figure 9 compares the statistical power of these five methods to detect mutation-selection-drift. For all tests, the power to detect selection increases initially with $N_e s$ but falls off as major codon usage reaches 100%. Each of the statistical methods shows some power to detect weak selection. The relative power of the different tests is similar to that for the neutral *vs.* selected mutations tests. The frequency distribution test is generally less powerful than tests that include divergence data. Among the latter category, 2×3 tests of independence are considerably more powerful than both 2×2 and $2 \times$

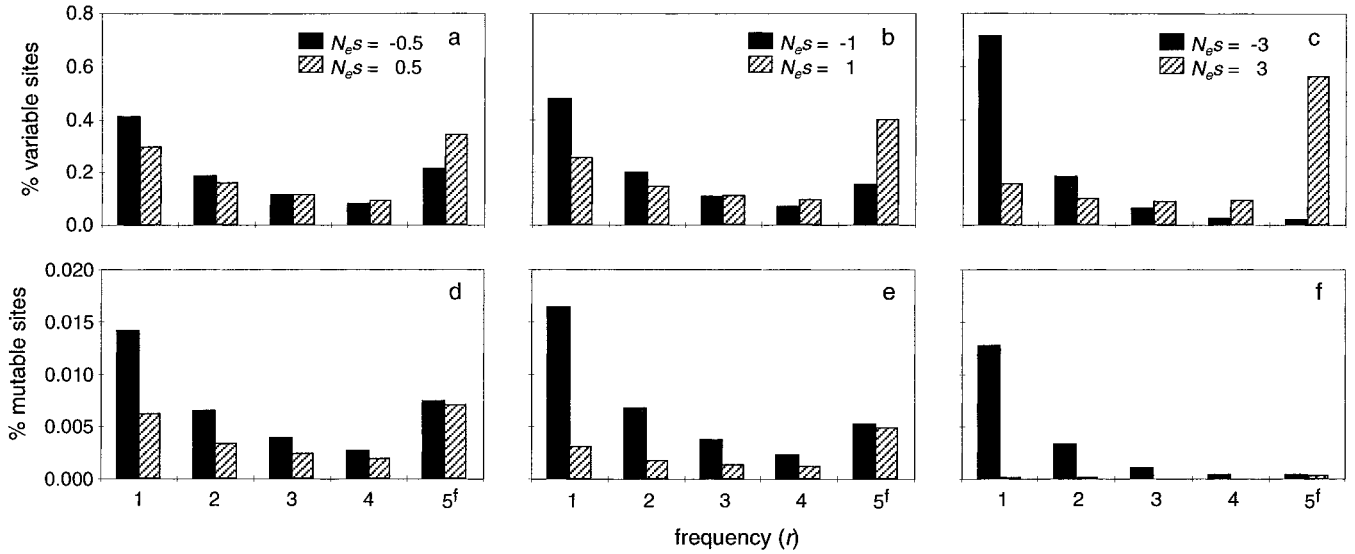

Figure 8.—Expected configurations of preferred and unpreferred mutations under major codon preference. The expected numbers of newly arisen mutations at frequency classes $r = 1$ to m in a sample of sequences were calculated according to Sawyer and Hartl (1992) and Hartl *et al.* (1994). Data are shown for $m = 5$ sequences and $t_{div} = 0.6$. a, b, and c show the expected proportion of *variable* sites in the sample at different frequencies under $N_es = \pm 0.5$, ± 1.0, and ± 3.0, which correspond to major codon usages of 65, 85, and 100%, respectively. c, d, and e show the proportion of sites at which variants are expected to be segregating at different frequencies or fixed in the sample. Superscript f denotes "fixed" difference class ($r = m$).

m tests. Overall, however, the fddMWU test is either indistinguishable from, or more powerful than, all the other tests over the parameter ranges considered. The gain in power is greatest when the number of sampled alleles is large. For the same total number of aligned nucleotides, increasing the number of sampled sites has a greater impact on statistical power than increasing the numbers of alleles.

These power analyses suggest that configuration comparisons, given enough mutations, can detect natural selection near its limit of efficacy. The configurations of preferred and unpreferred synonymous codons have been compared in DNA sequence data from *D. simulans* (Akashi 1997a) and the results are reiterated in Figure 10. Major codon preference predicts frequency distributions and divergence skewed toward higher values for advantageous preferred silent mutations than for deleterious unpreferred mutations. Equivalent configurations constitute the null hypothesis in this comparison and neutrality of both classes of mutations (the purely mutational model) is a subset of this null. Figure 10 shows the configurations of preferred and unpreferred mutations pooled from five alleles from each of eight *D. simulans* genes found in the literature or in GenBank (Table 2). Methods to identify major codons and infer ancestral and derived states for silent mutations are given in Akashi (1995). Equivalent data from *D. melanogaster* are not shown because other lines of evidence indicate a reduction in the efficacy of selection at silent sites in this lineage (Akashi 1995, 1996). Although only five alleles were analyzed in *D. simulans*, close to 2500 silent sites were examined across the eight genes in Table 2. If codon bias is maintained under mutation-selection-drift in this lineage, then frequency distribution and divergence comparisons should have a high probability of rejecting fitness equivalence between preferred and unpreferred mutations.

In *D. simulans*, the configurations of preferred and unpreferred mutations are similar to those expected under weak selection (Figure 8b). The 37 preferred mutations are segregating at higher frequencies and are more often fixed than the 101 unpreferred changes (Mann-Whitney *U*-test, $z = 3.12$, $P = 0.0009$, one-tailed). The other statistical tests were also significant at the 5% level; frequency distributions are skewed toward higher values (Mann-Whitney *U*-test, $z = 1.71$, $P = 0.044$), ratios of polymorphism to divergence are lower (Fisher's exact test, $P = 0.007$), and the ratios of singleton, intermediate frequency, and fixed differences are skewed toward higher values for preferred than for unpreferred mutations (Monte Carlo homogeneity test, $P = 0.015$). These patterns are both consistent with major codon preference and difficult to explain in the absence of selection (Akashi 1997a).

NONNEUTRAL SILENT SITES AND TESTS OF NATURAL SELECTION IN PROTEIN EVOLUTION

Sawyer *et al.* (1987) and McDonald and Kreitman (1991) proposed comparisons of configurations between a putatively neutrally evolving (silent) and a potentially selected (replacement) class of mutations. A number of claims of adaptive protein evolution depend on such an assumption of neutral evolution of synonymous mu-

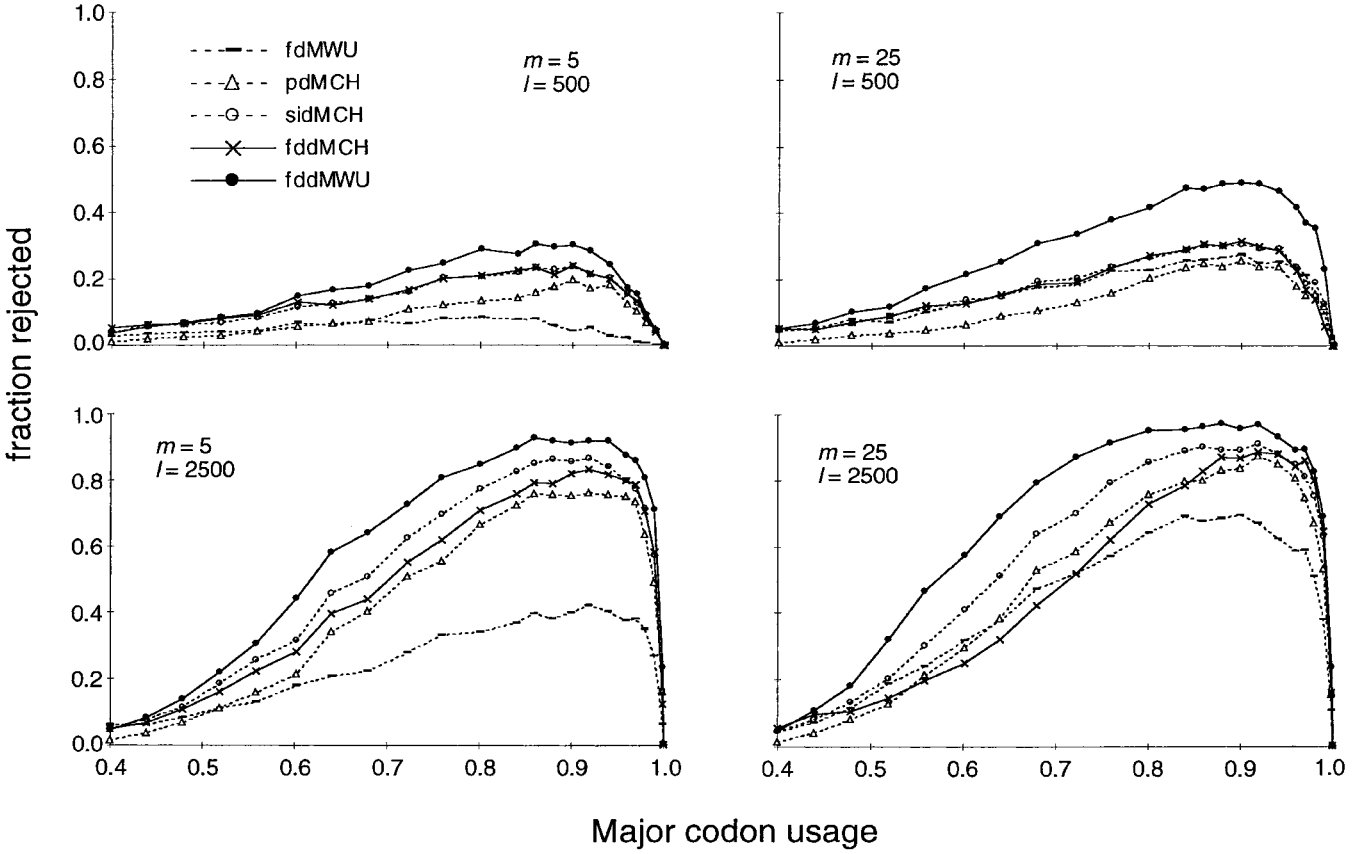

Figure 9.—Power of polymorphism and divergence configuration tests between preferred and unpreferred mutations. The *y*-axis plots the proportion of tests that reject fitness equivalence, $P < 0.05$, among 1000 simulated data sets for each value of the proportion of major codons at mutable sites. Because the direction of the deviations in the configurations of preferred and unpreferred mutations are predicted under major codon preference, one-tailed probabilities were calculated for these tests. For Monte Carlo homogeneity tests, the proportion of randomized tables with both a higher average frequency of preferred than unpreferred codons and an equal or larger test statistic than the data was taken as the probability of a given configuration. The five tests are those from Figure 6. The analyses are similar to those shown in Akashi (1997a, Figure 5) but the per-site mutation rates are more conservative (lower), the $2 \times m$ test of independence is included, Monte Carlo homogeneity tests replace *G*-tests, and power to reject the null is plotted as a function of levels of major codon usage rather than selection intensity.

tations (McDonald and Kreitman 1991; Eanes *et al.* 1993; Long and Langley 1993; Karotam *et al.* 1995, King 1998). Configuration tests, however, support weak selection at silent sites in Drosophila (Akashi 1995, 1997a; Akashi and Schaeffer 1997), and a number of other patterns of silent DNA evolution in Drosophila are consistent with mutation-selection-drift (Shields *et al.* 1988; Sharp and Li 1989; Kliman and Hey 1993, 1994; Moriyama and Hartl 1993; Akashi 1994; Moriyama and Powell 1997; Powell and Moriyama 1997). Although major codon preference could, in principle, affect interpretations of comparisons between silent and replacement mutations (Akashi 1995), the magnitude of such effects has not been evaluated.

Computer simulations were conducted to determine whether neutral protein evolution and selection at silent sites could mimic patterns that have been attributed to adaptive amino acid substitutions. DNA variation data were generated under a combination of the two scenarios described in the sections above. A given locus consists of two categories of mutations; half of the sites evolve under major codon preference and the other half evolve neutrally. Preferred and unpreferred mutations at major codon preference sites are pooled into a single category representing silent mutations and compared to the neutral class, representing replacement mutations. (This does not assume that all protein mutations are neutral. Amino acid positions at which mutations are strongly selected against do not contribute to variation and would not be counted in the number of mutable replacement sites in these simulations.) DNA variation data were generated as described above for the same mutational parameters, effective population size, times of divergence, and sample sizes. pdMCH tests of homogeneity and fdMWU and fddMWU tests were applied to 1000 simulated configurations of silent and replacement mutations for each set of parameters.

Figure 11 shows the fraction of statistical tests that reject equivalence of the configurations of nonneutral silent and neutral replacement mutations. Under the

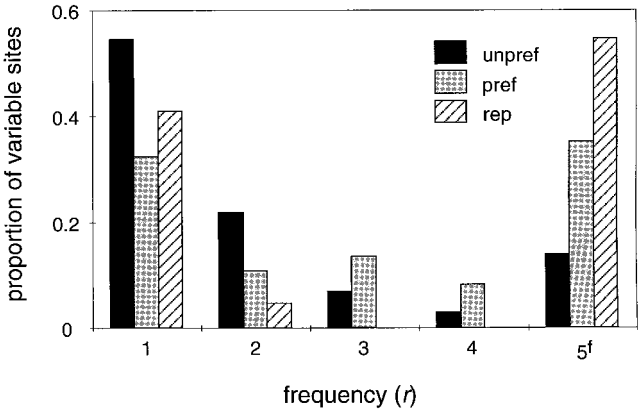

Figure 10.—The configurations of preferred, unpreferred, and replacement mutations in *D. simulans*. The proportions of 101 unpreferred (black), 37 preferred (gray), and 22 replacement (striped) mutations segregating at the given frequencies or fixed among five alleles of each of eight *D. simulans* genes are shown. Pooled data are from eight *D. simulans* genes from Table 2. Interestingly, 11 of the 12 amino acid "fixations" in *D. simulans* have occurred in the *Zw* gene, whereas the singleton polymorphisms are distributed more evenly among the eight genes; the distribution of selection coefficients may differ among genes. Superscript f denotes "fixed" difference class ($r = m$).

parameters examined, these tests show some power to reject the null when major codons reach a frequency of about 70 to 80% ($N_e s \approx 1$). At higher levels of major codon usage (stronger selection), the tests can be quite sensitive to mutation-selection-drift. At $t_{div} = 0.6$, the fdMWU test is more powerful than the pdMCH test, and the fddMWU test is generally most powerful. At higher levels of divergence, all tests become more sensitive to major codon preference and the pdMCH test outperforms the fdMWU test (data not shown).

The impact of mutation-selection-drift on silent/replacement configuration comparisons is very sensitive to the strength of selection. If major codon preference accurately describes silent evolution, then silent/replacement comparisons may be valid in low codon bias genes. However, when selection intensity (and the proportion of major codons) is high, the majority of silent mutations are deleterious unpreferred changes (Figure 8, d–f), and the means of the configurations of silent mutations tend to be lower than those of the neutral expectation. The null model can be rejected at a high rate, and the relative locations of the configurations are consistent with neutral evolution at silent sites and adaptive amino acid substitutions. For the data examined in this study, the average codon bias across the eight *D. simulans* genes is ~80% major codons; comparisons of replacement mutations to pooled silent mutations are difficult to interpret.

Separate contrasts of replacement mutations to preferred and unpreferred silent changes may shed some light on mechanisms of protein evolution (Akashi 1995). Under major codon preference, the configurations of preferred and unpreferred mutations reflect evolution under small positive and negative selection coefficients, respectively. Thus, configurations of replacement mutations skewed toward a larger fraction of high frequency variants and fixed differences than preferred mutations reflect adaptive protein evolution, whereas configurations skewed toward an excess of low frequency amino acid mutations relative to unpreferred silent mutations support deleterious protein evolution.

The configurations of replacement mutations as well as preferred and unpreferred silent mutations among the five alleles of eight *D. simulans* genes are shown in Figure 10. Surprisingly, roughly half of the variable replacement sites are singleton polymorphisms and the other half are fixed in the samples of five *D. simulans* alleles. Because no prediction had been made for the shape of the configuration of replacement mutations, Templeton's (1996) sidMCH test was applied to the data. The configuration of replacement mutations is significantly different from that of both preferred ($P = 0.028$, two-tailed) and unpreferred ($P < 0.001$) silent changes. Although the number of replacement mutations in these data is small, this configuration does not appear to conform to the predictions for protein evolution under uniform selection coefficients (including neutral evolution).

The excesses of rare amino acid polymorphisms and fixed differences can be explained by relaxing the assumption of uniform $N_e s$. One possibility is a combination of a large fraction of slightly deleterious amino acid changes and heterogeneity in effective population size over time. Lower effective population sizes in the past would have allowed slightly deleterious mutations to go to fixation, whereas more effective selection in larger current populations keeps deleterious polymorphisms at low frequencies. Such a nonequilibrium scenario was suggested by Ohta (1993) to explain lower ratios of

TABLE 2

Frequency distributions and divergence of DNA mutations in *D. simulans*

	n_r		
r	Unpref	Pref	Rep
1	55	12	9
2	22	4	1
3	7	5	0
4	3	3	0
5	14	13	12

The numbers of nonancestral mutations, n_r, segregating at frequency r in samples of five *D. simulans* sequences are shown for unpreferred (unpref) and preferred (pref) silent changes, and replacement (rep) mutations. Data were pooled across eight genes: *Adh*, *Adhr*, *boss*, *Mlc1*, *Rh3*, *per*, *Pgi*, and *Zw*. See Akashi (1997a) for GenBank accession numbers or references for these data.

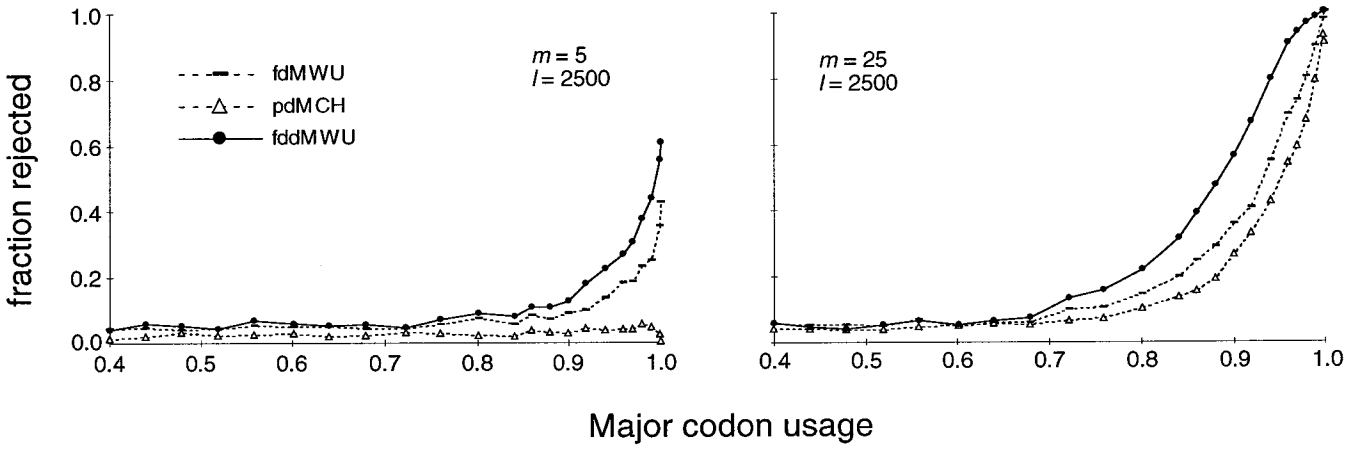

Figure 11.—Power of polymorphism and divergence configuration tests between neutral mutations and mutations under major codon preference. The *y*-axis plots the proportion of tests that reject fitness equivalence, $P < 0.05$, among 1000 simulated data sets for each value of major codon usage. See text for simulation parameters and Table 1 for abbreviations for tests.

polymorphism to divergence for replacement than silent evolution at the Drosophila Adh locus. However, because major codon preference is very sensitive to small changes in $N_e s$ (Akashi 1996), population-level phenomena should impact silent DNA mutations as well as amino acid changes. Roughly equal numbers of preferred and unpreferred silent fixations in the *D. simulans* lineage (Table 2) argue against large, or prolonged, fluctuations in effective population size.

Heterogeneity in selection coefficients, either across sites or across time, could also account for the configurations of amino acid mutations in *D. simulans*. One possibility is that selection coefficients vary among amino acid positions; low frequency polymorphisms are deleterious mutations that rarely go to fixation, whereas fixed differences in the sample reflect occasional adaptive amino acid substitutions. In this scenario, the polymorphic and fixed mutations in the sample are not a result of a single process of evolution under constant parameters but reflect a combination of the evolutionary dynamics of multiple fitness classes of mutations.

Selection coefficients varying across time, rather than among DNA sites, could also explain the deficiency of intermediate frequency amino acid polymorphisms (Hartl and Dykhuizen 1985). Most amino acid mutations are deleterious and are unlikely to reach appreciable frequencies within populations, but occasional environmental changes cause a subset of polymorphisms to become adaptive and rapidly go to fixation. Although the configuration of amino acid mutations in *D. simulans* is intriguing, more rigorous inference of mechanisms of protein evolution will require DNA sequence data both for a larger number of genes in this species and equivalent data from other lineages. In addition, interpreting these data may require predictions for the configurations of amino acid mutations under more complex models of evolution than those considered here.

DISCUSSION

Under the Sawyer-Hartl Poisson random field model, comparisons of the configurations of functional categories of DNA mutations can have considerable power to detect even very weak directional selection on classes of DNA mutations. These findings cannot be generalized beyond evolution under the parameter ranges considered and under the Sawyer-Hartl assumptions of stationarity, free recombination, and independent fitness effects of all mutations. Given these assumptions, configuration comparisons that include information from both frequency distributions of polymorphic mutations and numbers of fixed differences confer the greatest power to detect the fitness effects of mutations. The Mann-Whitney *U*-test, which is sensitive to differences in the locations of distributions, is a more powerful statistical approach to detect uniform selection coefficients than contingency tests of homogeneity. Accumulating DNA variation data for a large number of mutations with similar fitness effects is critical to the power of these tests. Configuration tests suggest that among eight *D. simulans* genes, a large fraction of both silent and replacement mutations affect fitness. Some limitations to this approach and these findings are discussed below.

Robustness of configuration comparisons: Under the Sawyer-Hartl assumptions, the numbers of observed mutations in each frequency class for each category of mutations are independent Poisson random numbers. Under these conditions, the test statistics of both Monte Carlo homogeneity tests and Mann-Whitney *U*-tests will be appropriately distributed under the null hypothesis of equivalent configurations of mutations. However, independent evolution at all sites, a stationary frequency distribution of mutations, and random sampling from a panmictic population are clearly not biologically realistic assumptions for many DNA sequence studies. One of the most appealing features of configuration compar-

isons is its claimed robustness to these assumptions. In the special case of no recombination, each of the alleles in a sample will be related by a single genealogy. If mutations have occurred at a constant (and low) rate on this genealogy, then the numbers of mutations from each category on each branch of the genealogy will be independent Poisson random numbers, regardless of whether the particular genealogy is sampled from an equilibrium, panmictic population (Sawyer *et al.* 1987; Hudson 1993). A similar argument has been made for intermediate levels of recombination. If the two classes of mutations are randomly interspersed in a genetic region (with respect to differences in evolutionary histories of subregions in the data), then configuration tests should be robust to departures from stationarity and panmixis (McDonald and Kreitman 1991; Hudson 1993). The effects of such departures from the Sawyer-Hartl assumptions on the distribution of the test statistics of configuration comparisons have not been confirmed. It is unclear whether genetic linkage, nonstationarity, or nonrandom population sampling can lead to false rejection of the null model under equivalent distributions of fitness effects (Type I error). The robustness and power of configuration tests under violations of the Sawyer-Hartl assumptions are not addressed here.

The analyses above have implicitly assumed that per-locus mutation rates have remained constant over the time periods examined. Particular scenarios of variable mutation rates can produce differences in the configurations of mutations identical to those resulting from natural selection (Eyre-Walker 1997). Consistent differences in configuration tests in independent lineages can distinguish the effects of mutational processes from those of selection (Akashi 1997b).

Interpreting departures from equivalent configurations of mutations: If configuration comparisons are robust to departures from the assumptions of the Poisson random field model, then such methods provide a general approach for inferring the distribution of fitness effects for various classes of mutations. However, the relationship between evolutionary configurations and the fitness effects of mutations must be treated with caution. The null hypothesis of these tests is equivalent configurations for the classes of mutations. This condition is satisfied when the distributions of selection coefficients for the two classes of mutations are equivalent (neutrality for both classes of mutations is one scenario that satisfies this null). However, the converse does not necessarily hold; it is possible that different distributions of selection coefficients can give rise to the same configuration of mutations. Thus, rejection of the null can be interpreted as evidence for differences in the fitness effects of mutations, but similar configurations do not necessarily imply similar fitness effects. "Tests of neutrality" is not an appropriate description of configuration comparisons because the null hypothesis includes, but is not limited to, neutral evolution for both classes of mutations.

Given a departure from equivalent configurations for two or more classes of mutations, further inference (of the sign and magnitude of selection coefficients) requires additional information or assumptions. Under a directional selection model with uniform selection coefficients, the relative locations of the configurations identify the relative magnitudes of the fitness effects for the two classes of mutations. For example, if one class of mutations is known to evolve neutrally, then configurations skewed toward an excess of high frequency and fixed variants for a second class suggest adaptive evolution. However, the same pattern could arise from weak deleterious effects of the first class of mutations and neutral evolution for the second class. The assumptions underlying such inferences should be made explicit.

A number of studies have attempted to infer the absolute intensity of selection (the magnitude of $N_e s$) from the configuration of mutations (Sawyer *et al.* 1987; Sawyer and Hartl 1992; Hartl *et al.* 1994; Akashi 1995; Akashi and Schaeffer 1997; Nachman 1998). These studies have found maximum-likelihood estimates for $N_e s$ given the observed ratios of polymorphic and fixed differences or the observed frequency distribution of polymorphic mutations. In addition to the Sawyer-Hartl assumptions of free recombination and stationarity, these studies have imposed an additional assumption of uniform $N_e s$ for all mutations in a given category. Surprisingly, each of the studies has found maximum-likelihood estimates of $|N_e s| \approx 1$. However, none of the studies has tested the fit of a distribution of selection coefficients to the data (such nested hypotheses can be tested through likelihood-ratio tests). For example, a number of studies have interpreted higher polymorphism/divergence ratios for replacement than for silent mutations as evidence for slightly deleterious protein mutations with uniform $N_e s \approx -1$. It is possible that a combination of relatively strongly deleterious and neutral (or even adaptive) mutations could explain the data equally well. Examination of each frequency class in the configuration of mutations under the Sawyer-Hartl maximum-likelihood method may help in distinguishing between such scenarios. The patterns of DNA variation in *D. simulans* suggest that distributions including both positive and negative fitness effects should be considered for both silent and protein variation.

Defining putative fitness classes of mutations: Configuration comparisons can be applied to any categories of interspersed DNA mutations. Fitness effects of mutations in introns and noncoding regions, insertion/deletion events, and mutations in regulatory elements can be assessed through such methods. However, the power of this approach depends critically on the ability to identify putative fitness classes of mutations. At silent sites, a simple model of selection for translational efficiency predicts differential fitness effects of forward and

backward DNA mutations; configuration comparisons provide a straightforward and powerful test of such a prediction. Comparison of pooled silent changes to a neutral class of variation has considerably less power to reveal selection (Figure 11). Unfortunately, defining putative fitness classes of protein variation is difficult; replacement mutations are often pooled into a single category. Thorne *et al.* (1996) and Templeton (1996) have attempted to subdivide amino acid positions by their role in protein structure, but such subdivisions correspond to functional categories, rather than putative fitness classes. A combination of biochemical and ecological studies (reviewed in Takahata 1996; Golding and Dean 1998) could allow more informative use of configuration comparisons. The relative lack of biological models that predict the fitness effects of particular mutations may be the strongest limitation in current applications of configuration tests to reveal mechanisms of protein evolution. However, one of the critical determinants of the power of configuration tests is the number of sites evolving under a particular model of evolution. If protein adaptation occurs by a small number of amino acid substitutions at a few key sites (Perutz 1983; Yokoyama 1997; Golding and Dean 1998), then tests of parallelism/convergence (Goldman 1993; Zhang and Kumar 1997) or the maximum-likelihood tests of Nielsen and Yang (1998) may be more powerful than configuration tests for identifying positive directional selection in protein evolution.

I am grateful to John Gillespie, Dave Cutler, Chuck Langley, and Laura Rose for many valuable discussions during the course of this work. Two anonymous reviewers contributed a great deal to improving this manuscript. H.A. was a National Science Foundation/Alfred P. Sloan Foundation Postdoctoral Fellow in Molecular Studies of Evolution.

LITERATURE CITED

Akashi, H., 1994 Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. Genetics **136:** 927–935.

Akashi, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. Genetics **139:** 1067–1076.

Akashi, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics **144:** 1297–1307.

Akashi, H., 1997a Codon bias evolution in Drosophila: population genetics of mutation-selection-drift. Gene **205:** 269–278.

Akashi, H., 1997b Distinguishing the effects of mutational biases and natural selection on DNA sequence variation [letter]. Genetics **147:** 1989–1991.

Akashi, H., and S. W. Schaeffer, 1997 Natural selection and the frequency distributions of "silent" DNA polymorphism in Drosophila. Genetics **146:** 295–307.

Andersson, S. G. E., and C. G. Kurland, 1990 Codon preferences in free-living microorganisms. Microbiol. Rev. **54:** 198–210.

Ballard, J. W., and M. Kreitman, 1994 Unraveling selection in the mitochondrial genome of Drosophila. Genetics **138:** 757–772.

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 plymorphisms. Genetics **140:** 783–796.

Bulmer, M., 1988 Are codon usage patterns in unicellular organisms determined by selection-mutation balance. J. Evol. Biol. **1:** 15–26.

Bulmer, M., 1991 The selection-mutation-drift theory of synonymous codon usage. Genetics **129:** 897–907.

Charlesworth, D., B. Charlesworth and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics **141:** 1619–1632.

Collins, T. M., P. H. Wimberger and G. J. P. Naylor, 1994 Compositional bias, character-state bias, and character-state reconstruction using parsimony. Syst. Biol. **43:** 482–496.

Dykhuizen, D. E., and D. L. Hartl, 1983 Functional effects of PGI allozymes in *Escherichia coli*. Genetics **105:** 1–18.

Eanes, W. F., M. Kirchner and J. Yoon, 1993 Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90**: 7475–7479.

Eyre-Walker, A., 1997 Differentiating between selection and mutational bias. Genetics **147:** 1983–1987.

Fisher, R. A., 1930 The distribution of gene ratios for rare mutations. Proc. R. Soc. Edinb. Sect. B **50:** 204–219.

Freese, E., 1962 On the evolution of base composition of DNA. J. Theor. Biol. **3:** 82–101.

Frumhoff, P. C., and H. K. Reeve, 1994 Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. Evolution **48:** 172–180.

Fu, Y.-X., 1996 New statistical tests of neutrality for DNA samples from a population. Genetics **143:** 557–570.

Fu, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. Genetics **147:** 915–925.

Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, New York.

Golding, B. G., and A. M. Dean, 1998 The structural basis of molecular adaptation. Mol. Biol. Evol. **15:** 355–369.

Goldman, N., 1993 Simple diagnostic statistical tests of models for DNA substitution. J. Mol. Evol. **37:** 650–661.

Goldman, N., and Z. Yang, 1994 A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11:** 725–736.

Golding, G. B., C. F. Aquadro and C. H. Langley, 1986 Sequence evolution within populations under multiple types of mutation. Proc. Natl. Acad. Sci. USA **83:** 427–431.

Hartl, D. L., and D. E. Dykhuizen, 1985 The neutral theory and the molecular basis of preadaptation, pp. 107–124 in *Population Genetics and Molecular Evolution*, edited by T. Ohta and K. Aoki. Japan Sci. Soc. Press, Tokyo.

Hartl, D. L., E. N. Moriyama and S. Sawyer, 1994 Selection intensity for codon bias. Genetics **138:** 227–234.

Hey, J., 1997 Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14:** 166–172.

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Series in Ecology and Evolution*, Vol. 7, edited by D. Futuyma and J. Antonovics. Oxford University Press, Oxford.

Hudson, R. R., 1993 Levels of DNA polymorphism and divergence yield important insights into evolutionary processes. Proc. Natl. Acad. Sci. USA **90:** 7425–7426.

Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Hudson, R. R., M. Slatkin and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

Hudson, R. R., K. Bailey, D. Skarecky, J. Kwiatowski and F. J. Ayala, 1994 Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

Ikemura, T., 1985 Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2:** 13–34.

Ina, Y., 1996 Pattern of synonymous and nonsynonymous substitutions: an indicator of mechanisms of molecular evolution. J. Genet. **75:** 91–115.

Karotam, J., T. M. Boyce and J. G. Oakeshott, 1995 Nucleotide variation at the hypervariable esterase 6 isozyme locus of Drosophila simulans. Mol. Biol. Evol. **12:** 113–122.

Kelly, J. K., 1997 A test of neutrality based on interlocus associations. Genetics **146:** 1197–1206.
Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics **47:** 713–719.
Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
Kimura, M., and T. Ohta, 1971 Protein polymorphism as a phase of molecular evolution. Nature **229:** 467–469.
King, L. M., 1998 The role of gene conversion in determining sequence variation and divergence in the *Est-5* gene family in *Drosophila pseudoobscura*. Genetics **148:** 305–315.
Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.
Kliman, R. M., and J. Hey, 1994 The effects of mutation and natural selection on codon bias in the genes of Drosophila. Genetics **137:** 1049–1056.
Kreitman, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.
Lewontin, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
Li, W.-H., 1987 Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. **24:** 337–345.
Long, M., and C. H. Langley, 1993 Natural selection and the origin of *jingwei*, a chimeric processed functional gene in Drosophila. Science **260:** 91–95.
McDonald, J. H., 1996 Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **13:** 253–260.
McDonald, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **15:** 377–384.
McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.
Moriyama, E. N., 1987 Higher rates of nucleotide substitution in Drosophila than in mammals. Jpn. J. Genet. **63:** 139–147.
Moriyama, E. N., and D. L. Hartl, 1993 Codon usage bias and base composition of nuclear genes in *Drosophila*. Genetics **134:** 847–858.
Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.
Moriyama, E. N., and J. R. Powell, 1997 Codon usage bias and tRNA abundance in *Drosophila*. J. Mol. Evol. **45:** 514–523.
Muse, S. V., 1996 Estimating synonymous and nonsynonymous substitution rates. Mol. Biol. Evol. **13:** 105–114.
Nachman, M. W., 1998 Deleterious mutations in animal mitochondrial DNA. Genetica **102/103:** 61–69.
Nachman, M. W., S. N. Boyer and C. F. Aquadro, 1994 Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc. Natl. Acad. Sci. USA **91:** 6364–6368.
Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.
Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.
Nielsen, R., and Z. Yang, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:** 929–936.
Ohta, T., 1973 Slightly deleterious mutant substitutions in evolution. Nature **246:** 96–98.
Ohta, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Sys. **23**: 263–286.
Ohta, T., 1993 Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. Proc. Natl. Acad. Sci. USA **90:** 4548–4551.
Perutz, M. F., 1983 Species adaptation in a protein molecule. Mol. Biol. Evol. **1:** 1–28.
Powell, J. R., and E. N. Moriyama, 1997 Evolution of codon usage bias in *Drosophila*. Proc. Natl. Acad. Sci. USA **95:** 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*, Ed. 2. Cambridge University Press, Cambridge.
Rand, D. M., M. Dorfsman and L. M. Kann, 1994 Neutral and nonneutral evolution of Drosophila mitochondrial DNA. Genetics **138:** 741–756.
Rowan, R. G., and J. A. Hunt, 1991 Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of Hawaiian Drosophila. Mol. Biol. Evol. **8:** 49–70.
Rzhetsky, A., and M. Nei, 1995 Tests of applicability of several substitution models for DNA sequence data. Mol. Biol. Evol. **12:** 131–151.
Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics **132:** 1161–1176.
Sawyer, S. A., D. E. Dykhuizen and D. L. Hartl, 1987 Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA **84:** 6225–6228.
Schluter, D., T. Price, A. O. Mooers and D. Ludwig, 1997 Likelihood of ancestral states in adaptive radiation. Evolution **51:** 1699–1711.
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., and W.-H. Li, 1989 On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Biol. **28:** 398–402.
Sharp, P. M., M. Averof, A. T. Lloyd, G. Matassi and J. F. Peden, 1995 DNA sequence evolution: the sounds of silence. Philos. Trans. R. Soc. Lond. B Biol. Sci. **349:** 241–247.
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.
Simonsen, K. L., G. A. Churchill and C. F. Aquadro, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.
Strobeck, C., 1987 Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics **117:** 149–153.
Sueoka, N., 1962 On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA **48:** 582–592.
Sueoka, N., 1988 Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85:** 2653–2657.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.
Takahata, N., 1996 Neutral theory of molecular evolution. Curr. Opin. Genet. Dev. **6:** 767–772.
Templeton, A. R., 1996 Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the Hominoid primates. Genetics **144:** 1263–1270.
Thorne, J. L., N. Goldman and D. T. Jones, 1996 Combining protein evolution and secondary structure. Mol. Biol. Evol. **13:** 666–673.
Watterson, G. A., 1978 The homozygosity test of neutrality. Genetics **88:** 405–417.
Wayne, M. L., D. Contamine and M. Kreitman, 1996 Molecular population genetics of *ref(2)P*, a locus which confers viral resistance in Drosophila. Mol. Biol. Evol. **13:** 191–199.
Wise, C. A., M. Sraml and S. Easteal, 1998 Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. Genetics **148:** 409–421.
Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.
Yang Z., S. Kumar and M. Nei, 1995 A new method of inference of ancestral nucleotide and amino acid sequences. Genetics **141:** 1641–1650.
Yokoyama, S., 1997 Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. Annu. Rev. Genet. **31:** 315–336.
Zhang, J., and S. Kumar, 1997 Detection of convergent and parallel evolution at the amino acid sequence level. Mol. Biol. Evol. **14:** 527–536.
Zhang, J., and M. Nei, 1997 Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. **44:** S139–S146.

Communicating editor: A. G. Clark

APPENDIX

Generating random 2 × *n* tables for Monte Carlo tests of homogeneity: Assume a 2 (columns) × *n* (rows)

contingency table with column sums c_1, c_2 and row sums $r_1, r_2, \ldots, r_n$. Into an array of length $l = c_1 + c_2$, randomly insert r_1 copies of the integer 1, r_2 copies of the integer 2, . . . , and r_n copies of integer n. The counts (the numbers of 1's, 2's, . . . , n's) for the first c_1 entries of this array form the first column of the $2 \times n$ table, and the counts for the remaining c_2 entries form the second column. Randomly permute the array of n integers and reconstitute the table. Such random tables will have the same row and column sums as the observed (or simulated) table and the correct joint hypergeometric distribution (Bill Engels, personal communication).