J Mol Evol (1999) 49:343–351

JOURNAL OF MOLECULAR EVOLUTION

Recent Selection on Synonymous Codon Usage in *Drosophila*

Richard M. Kliman*

Department of Biology, Radford University, Radford, VA 24142, USA

Received: 15 April 1998 / Accepted: 13 May 1999

Abstract. Evidence from a variety of sources indicates that selection has influenced synonymous codon usage in *Drosophila*. It has generally been difficult, however, to distinguish selection that acted in the distant past from ongoing selection. However, under a neutral model, polymorphisms usually reflect more recent mutations than fixed differences between species and may, therefore, be useful for inferring recent selection. If the ancestral state is preferred, selection should shift the frequency distribution of derived states/site toward lower values; if the ancestral is unpreferred, selection should increase the number of derived states/site. Polymorphisms were classified as ancestrally preferred or unpreferred for several genes of *D. simulans* and *D. melanogaster*. A computer simulation of coalescence was employed to derive the expected frequency distributions of derived states/site under various modifications of the Wright–Fisher neutral model, and distributions of test statistics (*t* and Mann–Whitney *U*) were derived by appropriate sampling. One-tailed tests were applied to transformed frequency data to assess whether the two frequency distributions deviated from neutral expectations in the direction predicted by selection on codon usage. Several genes from *D. simulans* appear to be subject to recent selection on synonymous codons, including one gene with low codon bias, *esterase-6*. Selection may also be acting in *D. melanogaster*.

Key words: Codon bias — *Drosophila* — Natural selection — *Esterase-6*

* *Present address:* Department of Biology, Kean University, 1000 Morris Ave., Union, NJ 07083, USA
e-mail: rkliman@turbo.kean.edu

Introduction

Several studies have suggested that synonymous codon usage in *Drosophila* reflects, in part, the influence of natural selection. Shields et al. (1988) first presented evidence for this, perhaps most notably that the disproportionally used codons end in either C or G, though the apparent mutational bias in *D. melanogaster* is toward A and T. Divergence between *D. melanogaster* and *D. pseudoobscura* at synonymous sites was found to correlate negatively with codon bias (Shields et al. 1988; Sharp and Li 1989; Moriyama and Gojobori 1992), as would be expected if most silent mutations produced an unpreferred codon. Various population genetics models predict that natural selection will be less effective in regions of low recombination (Hill and Robertson 1966; Felsenstein 1974; Charlesworth et al. 1993), and codon bias is significantly lower in regions presumed to have little or no recombination in *D. melanogaster* (Kliman and Hey 1993a). Akashi (1995) has found patterns of synonymous substitution in *Drosophila* consistent with directional selection on codon usage.

In several unicellular organisms, codon bias is clearly correlated with level of gene expression [e.g., *Escherichia coli* (Gouy and Gauthier 1982); *Saccharomyces cerevisiae* (Bennetzen and Hall 1982), and *Dictyostelium discoideum* (Sharp and Devine 1989)]. Defining "high expression" in a multicellular organism is difficult, though anecdotal evidence suggests that codon bias correlates with expression levels of paralogous members of multigene families (Shields et al. 1988). Codon bias might also, or instead, arise from selection on fidelity of translation (Akashi 1994).

Little of the evidence that supports a selective basis to codon bias in *D. melanogaster* and its close relatives can be used to infer recent selection on this character (but see Akashi and Schaeffer 1997). High codon bias might mainly reflect past selection, allowing the possibility that selection no longer has much effect. For example, it is possible that extant *D. melanogaster* arose from an ancestor with a substantially higher effective population size; if selection coefficients for differential codon usage are sufficiently low, it is possible that genetic drift subsequently overwhelmed the effect of selection (e.g., see Akashi 1995). On the other hand, if selection is still acting on synonymous codon usage, selective neutrality may be rejected by appropriate analyses.

When comparing orthologous DNA sequences, it is customary to distinguish fixed differences sampled among populations and polymorphisms sampled within populations. A fixed difference will be sampled when a derived character state rises to sufficiently high frequency (if not to fixation) in one population, such that all orthologous sites in gene copies sampled from one population differ from all those sampled from the other population. A polymorphism will be sampled when a newly derived character state rises to a frequency high enough to be sampled, but not so high as to present itself as a fixed difference. Under a neutral model, polymorphisms are thought to be due, on average, to more recent mutations, with the expectation that either the ancestral or the derived state will ultimately be fixed. Thus, analysis of patterns of polymorphism may be used to infer generally more recent evolutionary processes than analysis of fixed differences.

A number of statistical methods have been devised to compare observed patterns of DNA sequence polymorphism to those expected under neutral evolutionary models (Tajima 1989; Hudson et al. 1987; Fu and Li 1993). The test statistic devised by Tajima (1989) compares average number of pairwise sequence differences among gene copies per site to the value expected under a neutral model given the number of alleles sampled (Watterson 1975). If most mutations are deleterious, then purifying selection should result in fewer average pairwise differences per site than expected for the number of gene copies sampled.

A test of neutrality will be more powerful if polymorphic sites can be classified, a priori, as ancestrally preferred or unpreferred. In other words, the test is more powerful if it is known that selection should either increase or decrease the frequency of derived character states. If the common ancestor for all sampled gene copies had the preferred character state, then sites in the unpreferred state should be removed by purifying selection. If, on the other hand, the common ancestor had the unpreferred state, sites showing the preferred state should be found in a frequency greater than expected under a neutral model. Sawyer et al. (1987) were the first to apply a test comparing frequency distributions of two classes of mutations (synonymous and nonsynonymous), and Akashi and Schaeffer (1997) first used such an approach to compare distributions of ancestrally preferred and unpreferred synonymous mutations.

In this paper, a parametric test is used to compare observed frequency distributions of synonymous mutations to those expected under a Wright–Fisher neutral model. Evidence is found for recent selection having acted on codon usage of individual genes in *D. simulans;* that is, selection on codon usage seems to have influenced codon usage subsequent to the most recent common ancestor of the gene copies sampled. However, there is little evidence supporting recent selection in *D. melanogaster*.

Materials and Methods

Data Used in Analyses

Synonymous polymorphic nucleotide sites were identified in aligned sequences for each of several *Drosophila* genes. The genes were chosen because several gene copies had been sequenced in both *D. melanogaster* and *D. simulans.* Only polymorphic sites segregating two character states, one associated with a preferred codon and one associated with an unpreferred codon [based on Akashi's (1995) identification of preferred codons in *Drosophila*], were used in the analyses. The ancestral state for each *D. melanogaster* site was inferred from the character state in *D. simulans,* and vice versa; each site was then classified as being presumed ancestrally preferred or ancestrally unpreferred. Because *D. melanogaster* is generally less polymorphic than *D. simulans,* and because codon bias limits the potential number of ancestrally unpreferred polymorphic sites, some of the *D. melanogaster* genes lack ancestrally unpreferred sites and can not be used in the analyses.

For n homologous sequences, the number of derived states at a polymorphic site can range from 1 to $n - 1$. Tables 1 and 2 give the frequency distributions of ancestral:derived states for ancestrally preferred and ancestrally unpreferred polymorphic sites in *D. melanogaster* and *D. simulans,* respectively. The number of gene copies analyzed ranged from four (*est-6* in *D. simulans*) to 33 (*Zw* in *D. melanogaster*).

Statistical Tests

If selection has not been influencing the frequencies of derived character states in a population, then one should expect the same frequency distributions of derived states for sites classified in advance as either ancestrally preferred or ancestrally unpreferred. Frequency distributions of ancestrally preferred and unpreferred character states were compared using a parametric test (the t test) and a nonparametric test (the Mann–Whitney U test). To calculate the numerator of the t statistic (Sokal and Rohlf 1981, p. 226), the mean for data points from ancestrally preferred sites is subtracted from that for ancestrally unpreferred sites, i.e., the value of t is expected to be positive. Thus, statistical significance is always based on a one-tailed test.

Because the expected frequency distribution of derived states is not normally distributed, data points for t tests were transformed as described in the next section and in Table 3. The test statistic can be directly compared to the Student t distribution. However, because one of the assumptions of this test is equal variance in the two samples, it is useful to compare t_s, as well as U, to the distributions of these test statistics derived from simulated data sets; the simulated data sets, of

Table 1. Synonymous polymorphic sites in *D. melanogaster*[a]

Locus	Source of data	Ratio of ancestral:derived states	Ancestrally preferred sites	Ancestrally unpreferred sites
Adh	Laurie et al. (1991)	10:1	2	0
		9:2	3	1
		8:3	1	0
		6:5	2	0
		5:6	3	0
		4:7	0	1
		2:9	0	1
Zw	Eanes et al. (1993)	32:1	6	0
		31:2	2	0
		30:3	0	1
		28:5	1	0
		26:7	2	0
		22:11	1	0
		21:12	2	0
		15:18	1	0
		12:21	0	1
		10:23	1	0
		5:28	1	0
		1:32	0	1
est-6	Cooke & Oakeshott (1989)	12:1	9	2
		11:2	2	0
		10:3	1	1
		9:4	0	1
		7:6	1	0
		6:7	1	0
		5:8	1	0
		4:9	2	0
		3:10	2	0
		2:11	1	0

[a] For each gene, data are provided only for those sites that (i) could be categorized as ancestrally preferred or ancestrally unpreferred on the basis of comparison to an outgroup and (ii) were segregating one preferred and one unpreferred character state.

the same size as the actual data set, were constructed by sampling with replacement from the expected frequency distribution (see below). In general, *P* values obtained from the two methods are in close agreement (see Table 4), though the Monte Carlo approach should be more accurate.

An alternative to estimating statistical significance involves using a bootstrapping approach. Data from ancesetrally preferred and unpreffered sites were pooled to give the probability of obtaining each possible number of derived states; for example, the probability of observing two derived states at a site in *Adh* in *D. melanogaster* would be (3 + 1)/14 (see Table 1). Simulated data sets, of the same size as the actual data set, were produced by sampling with replacement using this probability distribution. The t_s statistics for the actual data and simulated data were calculated following data transformation, and statistical significance for *t* tests and *U* tests were estimated as described above.

It should be noted that, by sampling with replacement to produce all simulated data sets, tests using a Monte Carlo approach implicitly assume independent assortment of polymorphic sites. The tests are automatically conservative to tests that assume linkage, as linkage would lead to covariance between the number of derived states per site of ancestrally preferred and that of unpreferred sites without changing the mean.

Expected Frequency Distribution of Derived Character States Under a Neutral Model

The statistical tests described here (other than those using a bootrapping approach) takes into consideration the expected frequency distribution of derived states under a Wright–Fisher neutral model, assuming constant population size. Let *n* be the number of homologous gene copies from which polymorphisms are identified. A rooted, bifurcating genealogy for *n* homologous nucleotide positions can be described by *n*-1 coalescent events (i.e., *n* lineages descending to *n*-1 ancestral lineages, *n*-1 lineages then descending to *n*-2 lineages, . . . , 2 remaining lineages descending to a single ancestor). Under a Wright–Fisher neutral model with a constant population size *N* (of gene copies), if *N* is substantially larger than *n*, the probability that no two gene copies share a common ancestor in the previous generation is very nearly $1 - [n(n - 1)/2N]$ (Hudson 1983). The topology of the rooted genealogy is independent of the timing of coalescent events.

For any number of *n* sequences, the probability of a site having *d* derived states (where *d* can range from 1 to *n*-1) under a constant-*N* Wright–Fisher model is given by

$$P\langle d\rangle = \left(d \sum_{i=1}^{n-1} \frac{1}{i} \right)^{-1} \quad (1)$$

[see Eq. 22 of Fu (1995)], as long as mutations are Poisson-distributed and appear at a rate proportional to the total branch length of the genealogy.

Alternatively, if it assumed that every polymorphic site is produced by a single forward mutation and no back mutation, then the expected frequency distribution of new mutations can be derived by simulation. When simulating coalescence, the frequency of the derived state depends on the location of the mutation on the genealogy for the nucleo-

Table 2. Synonymous polymorphic sites in *D. simulans*[a]

Locus	Source of data	Ratio of ancestral: derived states	Ancestrally preferred sites	Ancestrally unpreferred sites
per	Kliman & Hey (1993b)	5:1	18	1
		4:2	3	0
		3:3	3	0
		2:4	2	1
		1:5	1	2
Adh	McDonald & Kreitman (1991)	5:1	6	0
		4:2	1	0
		3:3	0	2
Pgi	McDonald & Kreitman[b]	5:1	5	2
		2:4	0	1
Zw	Eanes et al. (1993)	11:1	1	0
		10:2	2	0
		8:4	1	1
		7:5	1	1
		6:6	0	1
		4:8	0	1
boss	Ayala & Hartl (1993)	4:1	12	5
		3:2	4	0
		2:3	3	0
		1:4	0	2
est-6	Karotam et al. (1995)	3:1	11	10
		2:2	1	6
		1:3	1	5

[a] See Table 1, footnote a, for details.
[b] Unpublished *Pgi* sequences can be found in GenBank (accession numbers L27547–L27552).

Table 3. Example of data transformation[a]

d[b]	$P(d)$[c]	$\sum_{i=1}^{d-1} P(i)$	$P(x)$	Transformed d [x from Eq. (2)]
1	0.4379562	0.0000000	0.2189781	−0.7754292
2	0.2189781	0.4379562	0.5474453	0.1189394
3	0.1459854	0.6569343	0.7299270	0.6122469
4	0.1094891	0.8029197	0.8576643	1.0699090
5	0.0875912	0.9124088	0.9562044	1.7086217

[a] The transformation for analyses involving six gene copies is shown.
[b] The variable d represents the number of derived states.
[c] $P(d)$ is the expected frequency of d based on Eq. (1).

tide site. Each random genealogy is constructed by first establishing the times (in units of N generations) separating successive coalescent events (Hudson 1990). Pairs of lineages are randomly chosen at each coalescent event to establish the branching pattern (i.e., topology) of the genealogy. Once the genealogy is constructed, a mutation is randomly placed onto it, the probability of it appearing on a particular branch being proportional to its length. In all of the analyses described, 10^7 iterations of the simulation were performed to estimate the expected frequency distribution of derived states.

Data Transformation

From Eq. (1), it should be clear that the different numbers of derived states at a site are not equally likely; the probability decreases as d goes from 1 to n-1. The same is true when no back mutation is assumed. To perform a t test on the distributions of d, values of d are normalized by considering the values of $P(d)$ for n gene copies under a given model. Hastings (1955, p. 192) provides a formula that estimates quite well the area under the normal distribution, $P(x)$, covered by the range $-\infty$ to x, where x is measured in units of standard deviation from the mean:

$$x = \frac{c_0 + c_1 a + c_2 a^2}{1 + d_1 a + d_2 a^2 + d_3 a^3} - a \qquad (2)$$

where

$$a = \sqrt{\frac{1}{P\langle x\rangle}}$$

$$c_0 = 2.515517, \quad c_1 = 0.802853, \quad c_2 = 0.010328$$

$$d_1 = 1.432788, \quad d_2 = 0.189269, \quad d_3 = 0.001308$$

Here the normal distribution is divided vertically into n-1 sections equivalent in size to $P(d)$, the expected frequency of d. The transformed value of d, $f(d)$, is calculated from Eq. (2) by $f(d) = x$, where

$$P\langle x\rangle = \sum_{i=1}^{d-1} P\langle i\rangle + \frac{P\langle d\rangle}{2}$$

An example of this transformation for six gene copies is given in Table 3. For values of $P(x)$ greater than 0.5, the value of 1-$P(x)$ is used in place of $P(x)$ and the resulting value is multiplied by −1.

Table 4. Proportion of test statistic values exceeding critical values[a]

4 df				20 df			
Critical value	*t* distribution	Simulated 3 anecestrally preferred: 3 ancestrally unpreferred	Simulated 4 anecestrally preferred: 2 ancestrally unpreferred	Critical value	*t* distribution	Simulated 11 ancestrally preferred: 11 ancestrally unpreferred	Simulated 20 ancestrally preferred: 2 ancestrally unpreferred
7.173	0.001	0.0042	0.0089	3.552	0.001	0.0013	0.0022
3.747	0.01	0.0149	0.0199	2.528	0.01	0.0106	0.0163
2.132	0.05	0.0401	0.0562	1.725	0.05	0.0501	0.0635
1.533	0.1	0.1023	0.1199	1.325	0.1	0.0992	0.1126
0.941	0.2	0.2127	0.1989	0.860	0.2	0.1989	0.2019
0.000	0.5	0.5341	0.5014	0.000	0.5	0.5026	0.4740

[a] For unequal sample sizes, the simulated distribution is based on the larger sample having the larger mean. At higher (i.e., more significant) critical values, the percentage exceeding the critical value is lower when the larger sample has the smaller mean.

Results

The results of tests are given in Table 5. The analyses indicate that selection has influenced codon usage in the relatively recent past, especially in *D. simulans.* The one-tailed significance values were below 0.05 in four of six *D. simulans* genes (*per, Adh, est-6,* and *Zw*) regardless of the test statistic used or method of estimating statistical significance. The one-tailed tests based on expected frequency distributions were significant for the *Zw* locus, but not for the other two loci, in *D. melanogaster;* the Mann–Whitney *U* test [as employed by Akashi and Schaeffer (1997)] was not significant for *D. melanogaster Zw*. There is little effect, in general, of changing the assumption regarding back mutation on significance of the *t* tests. In some cases, statistical significance was greater using the bootstrapping approach (particularly for the *U* tests); however, the findings were not qualitatively different.

It is possible that a significant test for any of these genes (i.e., a test that gives $P < 0.05$) is spurious, especially given the fact that multiple tests were performed. The probability that one or more of nine tests would be significant, by chance alone, is 0.299. There is a 0.063 chance of performing two or more spuriously significant tests. However, while only the test on *D. simulans Adh* can be declared significant after correction using the sequential Bonferroni method (Rice 1989), it is reasonable to suggest that at least three of the tests are nonspuriously significant.

Applying Fisher's test for combining probabilities of independent tests (Sokal and Rohlf 1981, p. 780), the data from *D. simulans* strongly reject the Wright–Fisher neutral model. Applying the test to one-tailed probabilities from simulation tests using the expected frequency distribution based on Eq. (1), $\chi^2 = 43.289$ (12 df, $P = 2.0 \times 10^{-5}$). Highly significant results were also obtained using the probabilities based on the alternative tests (e.g., tests using frequency distributions in the absence of back mutation and tests using the bootstrapping approach); the least significant combined test used the *U* tests for simulated data sets based on Eq. (1) ($\chi^2 = 36.862$, 12 df, $P = 2.36 \times 10^{-4}$). Thus, the *D. simulans* data, overall, indicate strong departure from the neutral model employed.

The two data sets previously compiled by Akashi and Schaeffer (1997) were also analyzed, and results of the tests are given in Table 5. One of these is a composite data set of five gene copies for several *D. simulans* loci (111 polymorphic sites in all). Consistent with their analyses, the null hypothesis is rejected using the *t* test, regardless of the sampling algorithm. The second data set used by Akashi and Schaeffer was composed of polymorphic sites from 99 copies of the *Adh* and *Adhr* loci of *D. pseudoobscura* (93 polymorphic sites in all). Again, the *t* test is consistent with their findings, as the neutral hypothesis is rejected by the *t* test. For both data sets, however, the significant Mann–Whitney *U* tests reported by Akashi and Schaeffer may be misleading. When the *U* distribution was derived using simulated data sets, neither test was significant at the $P < 0.05$ level, though they were nearly significant (Tables 5A and 5B). However, the *U* tests on both composite data sets were significant when the bootstrapping approach was used (Table 5C).

Discussion

The analyses reported here are consistent with a history of natural selection on synonymous codon usage of multiple genes in *D. simulans.* There is also evidence for possible selection on codon usage at *Zw* in *D. melanogaster.* At polymorphic sites segregating a preferred and an unpreferred codon, the average frequency of the derived preferred codons is greater than that for derived unpreferred codons. Because polymorphisms represent mutations that have occurred since the time of the most recent common ancestor of gene copies sampled, the

Table 5. Results of one-tailed tests[a]

			Mann–Whitney *U* test			*t* test	
Locus	Species	df	*P* (*U* distribution)	*P* (simulation)	*t*	*P* (*t* distribution)	*P* (simulation)
			A				
per	*D. simulans*	29	0.0132	0.0175	2.8144	0.0044	0.0069
Adh	*D. simulans*	7	0.0072	0.0095	5.0210	0.0008	0.0030
pgi	*D. simulans*	6	0.0984	0.2463	1.3693	0.1100	0.1132
est-6	*D. simulans*	32	0.0205	0.0233	2.0560	0.0240	0.0228
Zw	*D. simulans*	7	0.0236	0.0221	2.3842	0.0243	0.0271
bos	*D. simulans*	24	0.0992	0.4695	0.5838	0.2824	0.2773
Adh	*D. melanogaster*	12	0.0769	0.0838	1.4311	0.0890	0.0941
est-6	*D. melanogaster*	22	0.7705	0.7764	−0.7346	0.7352	0.7586
Zw	*D. melanogaster*	18	0.0597	0.0671	1.9302	0.0348	0.0397
—	*D. pseudoobscura*[b]	91	0.0429	0.0540	1.7373	0.0432	0.0438
—	*D. simulans*[b]	109	0.0436	0.0540	2.0076	0.0236	0.0256
			B				
per	*D. simulans*	29	0.0132	0.0160	2.7926	0.0005	0.0076
Adh	*D. simulans*	7	0.0072	0.0092	4.9802	0.0008	0.0037
pgi	*D. simulans*	6	0.0984	0.2447	1.3693	0.1110	0.1160
est-6	*D. simulans*	32	0.0205	0.0217	2.0639	0.0236	0.0224
Zw	*D. simulans*	7	0.0236	0.0215	2.3749	0.0246	0.0274
bos	*D. simulans*	24	0.0992	0.4690	0.5611	0.2900	0.2858
Adh	*D. melanogaster*	12	0.0769	0.0830	1.4178	0.0909	0.0963
Est-6	*D. melanogaster*	22	0.7705	0.7756	−0.7260	0.7378	0.7554
Zw	*D. melanogaster*	18	0.0597	0.0667	1.9172	0.0356	0.0408
—	*D. pseudoobscura*[b]	91	0.0429	0.0540	1.7482	0.0419	0.0429
—	*D. simulans*[b]	109	0.0436	0.0520	1.9938	0.0244	0.0267
			C				
per	*D. simulans*	29	0.0132	0.0087	2.7425	0.0052	0.0106
Adh	*D. simulans*	7	0.0072	0.0062	5.2637	0.0006	0.0099
pgi	*D. simulans*	6	0.0984	0.1332	1.3693	0.1100	0.1700
est-6	*D. simulans*	32	0.0205	0.0194	2.0831	0.0227	0.0211
Zw	*D. simulans*	7	0.0236	0.0225	2.5368	0.0194	0.0210
bos	*D. simulans*	24	0.0992	0.4673	0.4772	0.3188	0.3107
Adh	*D. melanogaster*	12	0.0769	0.0843	1.8156	0.0472	0.0489
est-6	*D. melanogaster*	22	0.7705	0.7694	−0.6895	0.7511	0.7409
Zw	*D. melanogaster*	18	0.0597	0.0658	1.8934	0.0372	0.0426
—	*D. pseudoobscura*[b]	91	0.0429	0.0428	1.7368	0.0429	0.0454
—	*D. simulans*[b]	109	0.0436	0.0430	1.9727	0.0255	0.0289

[a] **A** Expected frequency distribution from Eq. (1). **B** Expected frequency distribution derived from 10,000,000 simulations of coalescence assuming no opportunity for back mutation. **C** Expected frequency distribution based on observed frequency distribution after combining ancestrally preferred and unpreferred sites (i.e., bootstrapping).

[b] Combined data for multiple genes (Akashi and Schaeffer 1997).

results indicate (on an evolutionary time scale) fairly recent selection. One should not conclude that selection is *only* recent (i.e., that selection was not in effect before the time of the common ancestor). The tests provide no information regarding the influence of selection prior to the time of the common ancestor of gene copies sampled.

There are two main reasons that one might expect that those genes with high codon bias (i.e., with high usage of preferred codons) would be those for which the statistical tests would reject the neutral model. First, one might assume that recent selection on codon usage, inferred from these statistical analyses, might mirror past selection inferred from overall codon usage bias. Second, as Akashi (1999) has shown, the power of various tests, including the Mann–Whitney *U* test, to detect selection increases as preferred codon usage increases (i.e., until about 90% major codon usage is reached). *Adh* shows high codon bias in the *D. melanogaster* species complex, and the *Zw* and *per* loci show moderately high codon bias. However, *est-6,* for which the *D. simulans* test is significant, is one of the least biased genes (Kliman and Hey 1993a). Of the genes analyzed, it is the only one with more ancestrally unpreferred sites than ancestrally preferred sites. In fact, because of its lower usage of

preferred codons, the power to detect selection in *est-6* might, from Akashi's (1999) analyses, be lower than that for the other genes analyzed.

It is possible, of course, that the test on *est-6* in *D. simulans* is spuriously significant. In general, though, there does seem to be more evidence for selection on codon usage in *D. simulans* than in *D. melanogaster*. One possibility is that selection for preferred codon usage has become stronger in *D. simulans* since its divergence from *D. melanogaster*. The *est-6* gene product is, in fact, found at higher levels in *D. simulans* homogenates at certain stages of the life cycle (Karotam and Oakeshott 1992). However, this may not reflect an increase in expression in *D. simulans;* it could also reflect a decrease in *D. melanogaster*. Alternatively, if selection failed to produce codon bias in *est-6* before the divergence of *D. simulans* and *D. melanogaster*, it might suggest that N_e for this locus was previously low and has since increased, at least in *D. simulans*. It is improbable that ancestral N_e was low for all loci (i.e., for the population in general), since this would be expected to diminish codon bias across the genome. Thus, if selection has been influencing codon usage in *est-6*, another explanation for its low usage of preferred codons is required.

The effectiveness of selection on codon usage at *est-6* in the more distant past may have been reduced by a temporary decrease in recombination in the vicinity of this gene. As predicted by various population genetic models, codon bias is significantly lower in regions of the *D. melanogaster* genome characterized by very low or zero recombination (Kliman and Hey 1993a). In that analysis of 385 *D. melanogaster* loci, the average value of the Codon Adaptation Index [CAI (Sharp and Li 1986)] was 0.439 for genes outside suspected regions of low recombination (i.e., the tip of the X chromosome, centromeres, telomeres, and chromosome 4). All of the genes in the vicinity of *est-6* (polytene map position 69A) have values of CAI below this mean, indicating lower usage of preferred codons. With the exception of *arf* (CAI = 0.470), all genes located between polytene map positions 68C and 73D have values of CAI below that mean. While this is an a posteriori analysis, it might be noteworthy that the average CAI for genes in this interval is 0.335 ($n = 16$), lower than the average value for genes in the presumed hitchhiking regions (0.354), though not as low as the value for genes in hitchhiking regions after excluding genes near autosomal telomeres (0.258) (Kliman and Hey 1993a).

It would be interesting to see if tests reject neutrality at *D. simulans* genes close to *est-6*, and subsequent studies on polymorphism in this species ought to include such loci. With such analyses, it may be possible to assess whether or not the significant test at *est-6* is indicative of a relatively recent change in the effectiveness of weak selection on this region of the *D. simulans* genome.

In general, while rejection of the neutral model is consistent with selection, alternative neutral models might conceivably explain some of the findings presented here. In particular, changing population size will affect the expected frequency distributions of derived states and, by extension, the distribution of the test statistics employed. In a population of gene copies of constant size N, the most recent common ancestor for all extant copies of a given site is expected to be found approximately $2N$ generations into the past. For n sampled copies, this value will generally be less than, though still close to, $2N$ (Kliman and Hey 1993b). If the history of a population is such that it has relatively recently expanded to size N, the common ancestor is expected to have existed more recently. This is because the lengths of the deepest branches in the genealogy of the sampled gene copies will be shortened relative to the lengths expected in a population of constant size, since coalescent times will decrease as N decreases. Genetic hitchhiking will produce a similar effect on the genealogy of sites closely linked to the target of strong, positive selection. For a derived state to be common, it needs to result from a mutation on an internal branch of the genealogy, and any change in genealogical patterns that decreases internal branch length relative to external branch length will decrease the frequency of derived states.

In contrast, derived states will generally be more common if the population size was greater in the relevant past, such that deeper branches of the genealogy are disproportionately long. While there is no reason to assume that such a population model is applicable to *D. simulans*, such a population history has been suggested for *D. melanogaster* (Akashi 1995). Long branches deep in the genealogy are also associated with population subdivision and with balancing selection maintaining a polymorphism.

At this time, there is no reason to suspect a recent increase (gradual or sudden) in the effective population size of either *D. simulans* or *D. melanogaster*. Regardless, if there has increase in N over time, then a test based on the assumption of constant N should be conservative; because deep branches are shortened, the mean and variance of derived states per site should decrease. This was confirmed for the case of *est-6* by modifying the simulation to include a severe bottleneck $0.25N$, $0.5N$, N, and $2N$ generations in the past, where N is the population size of gene copies. When simulating the timing to successive coalescent events, the time to subsequent coalescent events was set to zero generations once the bottleneck period was reached. Rows 2–5 in Table 6 give the expected frequency distributions of derived states for simulated coalescence of four gene copies, incorporating severe bottlenecks at $0.25N$, $0.5N$, $1.0N$, and $2N$ generations in the past. The last two columns in Table 6 give the P values for t tests and U tests, using the simulation

Table 6. Expected frequency distributions of derived states with changing population size[a]

Change in population size	$P(d = 1)$	$P(d = 2)$	$P(d = 3)$	One-tailed P[b]	
				t test	*U* test
No change[c]	0.575657	0.261477	0.162866	0.0224	0.0217
Ancestor small					
0.25*N*	0.828359	0.144713	0.026928	0.0132	0.0036
0.5*N*	0.730025	0.203173	0.066802	0.0179	0.0104
1*N*	0.636102	0.243233	0.120665	0.0204	0.0178
2*N*	0.586349	0.258774	0.154877	0.0224	0.0211
Ancestor large					
0.25*N*	0.483322	0.304136	0.212542	0.0240	0.0260
0.5*N*	0.470077	0.301664	0.228259	0.0240	0.0265
1*N*	0.505480	0.283062	0.211458	0.0236	0.0252
2*N*	0.555728	0.266928	0.177345	0.0226	0.0228

[a] All simulations of coalescence were performed 10^7 times. Simulations did not permit back mutation.
[b] Tests were performed on *D. simulans est-6* data.
[c] *P* values for "no change" are taken from Table 5B.

method, based on the revised frequency distributions. The more recent bottlenecks are associated with considerably lower *P* values for both the *t* test and the *U* test, and all *P* values are below those shown in Table 5b for this data set. Thus, the significance of the *est-6* findings is not an artifact of there being a smaller ancestral population.

If the population size was greater in the past, then the *t* test assuming a constant *N* tends to be somewhat liberal. This is also expected, since the mean and variance of the number of derived states per site will increase as deeper branches in the genealogy are lengthened. The simulation of coalescence was modified such that the population size was increased 10-fold 0.25*N*, 0.5*N*, *N*, and 2*N* generations in the past, and frequency distributions associated with each of these models are given in the last four rows in Table 6. The results of tests based on these revised frequency distributions are shown for *D. simulans est-6*, as before, in the last two columns in Table 6. *P* values were slightly elevated when the change in population size occurs 0.25*N*, 0.5*N*, or *N* generations back. However, the effect is minimal. It might be noted that the effect was greater on the *U* test than on the *t* test. However, there is no evidence that an ancestrally larger population could explain the significant departure from neutral expectations at *est-6* in *D. simulans.*

Still, for cases of borderline statistical significance, one needs to consider the possibility that the current population size is less than the past population size. Akashi (1995) has suggested that apparent loss of selection on codon usage in *D. melanogaster* supports a model of decreased effective population size of this species since its divergence from *D. simulans.* The nearly significant test on *D. melanogaster Adh* needs to be considered in light of evidence for balancing selection acting at this locus (Hudson et al. 1987; Kreitman and Hudson 1991). However, given the low *P* value for the test on *D. melanogaster Zw* and the small effect on statistical significance of a 10-fold larger past population size, the significance of this test should not be simply dismissed as an artifact.

Finally, the analyses described here bring attention to a more general problem associated with tests of neutrality. Rejection of null hypotheses is inferred from the likelihood of obtaining a particular value or greater of a test statistic deemed suitable for the analysis. Akashi and Schaeffer (1997) used the Mann–Whitney *U* test to compare frequency distributions of derived preferred and unpreferred codons in composite data sets for *D. simulans* and *D. pseudoobscura;* in both cases, standard application of the *U* test indicated significant departure from the expectation of the null hypothesis. However, the findings presented here indicate that the null distributions of the *U* and *t* statistics are affected by changing the basic population genetic model employed. For that matter, the use of *any* model has an effect. By employing the commonly used constant-*N* Wright–Fisher model to derive the expected frequency distributions, and sampling from this distribution to construct data sets from which to derive a null *U* distribution, the statistical significance of their earlier results is lost.

As shown earlier for *est-6* in *D. simulans,* the statistical significance of the *U* test (and the *t* test) is also affected by changing population size. The tests were repeated for Akashi and Schaeffer's (1997) *D. simulans* composite data set, after deriving expected frequency distributions of derived states when the population is 10-fold larger in the past. While $P = 0.0520$ with the Monte Carlo approach using a constant-*N* model (see Table 5B), $P > 0.057$ when the expected frequency distribution was derived assuming a population size 10-fold larger 0.25*N*, 0.5*N* and 1*N* generations back, and $P = 0.0536$ when the increase was 2*N* generations back. Statistical significance was actually slightly greater, relative to that from the test based on a constant *N* model, for the *t* tests under the same conditions.

As an alternative, a boostrapping approach was used. In effect, the method assumes an implicit model that perfectly explains the frequency distribution obtained by combining ancestrally preferred and unpreferred sites. This removes certain confounding influences, such as changing population size, that would drive both frequency distributions away from that expected by a given neutral model. For example, if the population is growing, we expect both frequency distributions to shift toward fewer derived states; if simulated data sets used to produce a test statistic distribution are sampled from the constant *N* frequency distribution, *P* values might be misleading. However, even if one finds a neutral model for which the expected mean number of derived states per site equals that of the combined observed data, one can still be criticized for using a model that does not give the same *distribution.* On the other hand, it is hard to

rationalize a model that predicts only polymorphic sites of the 5:1 and 2:4 type and none of the 4:2 and 3:3 type (as seen with *pgi* in *D. simulans;* see Table 2). Thus, while bootstrapping may have certain advantages, it is not the ultimate solution to the problem of choosing an appropriate neutral model.

As a general rule, tests of selective neutrality based on comparison of frequency distributions, while clearly informative, should be used with appropriate caution. Certainly, blind comparison of test statistics to standard probability tables is risky; *P* values from Monte Carlo approaches are probably more accurate. Nevertheless, comparison of frequency distributions of ancestrally preferred and unpreferred synoymous codons, regardless of the choice of specific test statistic or population model, indicates significant departure from neutral expectations for genes of *D. simulans.* The findings are parsimoniously explained by recent, if not ongoing, selection on codon usage.

Acknowledgments. I thank Adam Eyre-Walker, Hiroshi Akashi, and Jody Hey for helpful discussions, as well as Marty Kreitman and two anonymous reviewers for comments that helped me improve the manuscript. I also thank Etsuko Moriyama for providing some of the aligned DNA sequence data sets used to identify polymorphic positions. This research was supported by the Jeffress Memorial Trust and by the Radford University College of Arts and Sciences.

References

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 (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, Schaeffer SW (1997) Natural selection and the frequency distributions of "silent" DNA polymorphism in Drosophila. Genetics 146:295–307

Ayala FJ, Hartl DL (1993) Molecular drift of the *bride-of-sevenless (boss)* gene in *Drosophila.* Mol Biol Evol 10:1030–1040

Bennetzen JL, Hall BD (1982) Codon selection in yeast. J Biol Chem 257:3026–3031

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular evolution. Genetics 134: 1289–1303

Cooke PH, Oakeshort JG (1989) Amino acid polymorphisms for esterase-6 in *Drosophila melanogaster.* Proc Natl Acad Sci USA 86:1426–1430

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

Felsenstein J (1974) The evolutionary advantage of recombination. Genetics 78:737–756

Fu Y-X (1995) Statistical properties of segregating sites. Theor Popul Biol 48:172–197

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

Gouy M, Gauthier C (1982) Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res 10:7055–7074

Hastings C Jr (1955) Approximations for digital computers. Princeton University Press, Princeton, NJ

Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. Genet Res 8:269–294

Hudson RR (1983) Testing the constant-rate neutral allele model with protein sequence data. Evolution 37:203–217

Hudson RR (1990) Gene genealogies and the coalescent process. In: Futuyma D, Antonovics J (eds) Oxford surveys in evolutionary biology, Vol 7. Oxford University Press, Oxford, pp 1–44

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Karotam K, Oakeshott, JG (1992) Regulatory aspects of esterase 6 activity variation in sibling *Drosophila* species. Heredity 71:41–50

Karotam J, Boyce TM, Oakeshott JG (1995) Nucleotide variation at the hypervariable esterase 6 isozyme locus of *Drosophila simulans.* Mol Biol Evol 12:113–122

Kliman RM, Hey J (1993a) Reduced natural selection associated with low recombination in *Drosophila melanogaster.* Mol Biol Evol 10:1239–1258

Kliman RM, Hey J (1993b) DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics 133:375–387

Kreitman M, Hudson RR (1991) Inferring the evolutionary histories of the *Adh* and *Adh–dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics 127:565–582

Laurie CC, Bridgham JT, Choudhary M (1991) Association between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster.* Genetics 129:489–499

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila.* Nature 351:652–654

Moriyama EN, Gojobori T (1992) Rates of synonymous substitution and base composition of nuclear genes in Drosophila. Genetics 130:855–864

Rice WR (1989) Analyzing tables of statistics tests. Evolution 43:223–225

Sawyer SA, Dykhuizen DE, Hartl DL (1987) Confidence intervals for the number of selectively neutral amino acid polymorphisms. Proc Natl Acad Sci USA 84:6225–6228

Sharp PM, Devine KM (1989) Codon usage and gene expression level in *Dictyostelium discoideum:* Highly expressed gene do "prefer" optimal codons. Nucl Acids Res 17:5029–5038

Sharp PM, Li W-H (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 24:28–38

Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila.* J Mol Evol 28:398–402

Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol 5:704–716

Sokal RR, Rohlf FJ (1981) Biometry. Freeman, New York

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276