Copyright © 1996 by the Genetics Society of America

Molecular Evolution Between *Drosophila melanogaster* and *D. simulans*: Reduced Codon Bias, Faster Rates of Amino Acid Substitution, and Larger Proteins in *D. melanogaster*

Hiroshi Akashi

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

Manuscript received January 2, 1996
Accepted for publication August 9, 1996

ABSTRACT

Both natural selection and mutational biases contribute to variation in codon usage bias within Drosophila species. This study addresses the cause of codon bias differences between the sibling species, *Drosophila melanogaster* and *D. simulans*. Under a model of mutation-selection-drift, variation in mutational processes between species predicts greater base composition differences in neutrally evolving regions than in highly biased genes. Variation in selection intensity, however, predicts larger base composition differences in highly biased loci. Greater differences in the G+C content of 34 coding regions than 46 intron sequences between *D. melanogaster* and *D. simulans* suggest that *D. melanogaster* has undergone a reduction in selection intensity for codon bias. Computer simulations suggest at least a fivefold reduction in $N_e s$ at silent sites in this lineage. Other classes of molecular change show lineage effects between these species. Rates of amino acid substitution are higher in the *D. melanogaster* lineage than in *D. simulans* in 14 genes for which outgroup sequences are available. Surprisingly, protein *sizes* are larger in *D. melanogaster* than in *D. simulans* in the 34 genes compared between the two species. A substantial fraction of silent, replacement, and insertion/deletion mutations in coding regions may be weakly selected in Drosophila.

PATTERNS of codon usage and synonymous DNA divergence suggest a form of selection-mutation-drift at silent sites in *Escherichia coli* and *Saccharomyces cerevisiae*. In these species, codon usage is biased toward "major" codons, which are generally recognized by abundant tRNAs for each amino acid (IKEMURA 1981, 1982; BENNETZEN and HALL 1982; GROSJEAN and FIERS 1982). Among genes, the degree to which codon usage is biased correlates positively with expression levels (GRANTHAM *et al.* 1981; BENNETZEN and HALL 1982; GOUY and GAUTIER 1982; IKEMURA 1985). An inverse relationship between silent DNA divergence and codon usage bias is consistent with stronger selection for translationally superior codons in highly expressed genes (SHARP and LI 1987; but see EYRE-WALKER and BULMER 1995 for contrary evidence). These patterns, termed "major codon preference", suggest a balance between natural selection favoring major codons for each amino acid and mutation pressure and genetic drift allowing nonmajor codons to persist (SHARP and LI 1986; BULMER 1988; ANDERSSON and KURLAND 1990).

Patterns of codon usage and silent DNA evolution in *Drosophila melanogaster* appear to be similar to those found in *E. coli* and yeast (SHIELDS *et al.* 1988). Codon usage is biased toward a subset of (G- or C-ending) synonymous codons for each amino acid. However, tRNA abundances and gene expression levels are difficult to quantify in multicellular organisms. Major codons correspond to abundant tRNAs for the three amino acids for which data are available (SHIELDS *et al.* 1988). Anecdotal evidence suggests a relationship between expression levels and codon bias; highly expressed genes such as ribosomal proteins and *Adh* show greater codon bias than genes with limited or low expression such as *Adhr*. Similarly to prokaryotes, silent divergence between Drosophila species is inversely related to codon usage bias (SHARP and LI 1989; CARULLI *et al.* 1993).

Population genetic studies have confirmed the action of selection at silent sites in Drosophila. Under major codon preference, synonymous DNA mutations fall into two fitness classes: preferred mutations from nonmajor to major codons and unpreferred changes in the opposite direction (AKASHI 1995). Comparisons between the evolutionary dynamics of these two classes of synonymous mutations have established differences in their effects on fitness. As predicted by major codon preference, preferred mutations segregate at higher frequencies than unpreferred changes in the *Adh* region of *D. pseudoobscura* (H. AKASHI and S. SCHAEFFER, unpublished data). Differences in the ratios of polymorphism to divergence of preferred and unpreferred changes in *D. simulans* confirms the major codon preference model (AKASHI 1995). Maximum likelihood estimates of selection intensity from both frequency and polymor-

Address for correspondence: Hiroshi Akashi, Section of Evolution and Ecology, University of California at Davis, Davis, CA 95616.
E-mail: hakashi@ucdavis.edu

phism/divergence data suggest that, on average, selection at silent sites is very weak in Drosophila ($|N_e s| \approx 1$) (AKASHI 1995; H. AKASHI and S. SCHAEFFER, unpublished observations).

Population genetics theory also predicts a reduction in the effectiveness of natural selection at sites linked to positively and negatively selected mutations (HILL and ROBERTSON 1966; FELSENSTEIN 1974; LI 1987; CHARLESWORTH 1994; BARTON 1995). KLIMAN and HEY (1993) found lower codon usage bias in regions of reduced recombination in *D. melanogaster*, consistent with a reduction in the efficacy of selection for major codons.

Regional mutational biases also contribute to variation in synonymous codon usage in Drosophila. *D. melanogaster* genes vary considerably in G+C content at silent sites. If mutational biases account for this variation, then the G+C content of neutrally evolving sequences will show an association with that of closely linked silent sites. Correlations between intron and silent base composition suggest (assuming intron base composition reflects mutational equilibrium) that regional mutational biases explain ~10% of codon bias variation in *D. melanogaster* (KLIMAN and HEY 1994).

Although the forces governing synonymous codon usage within species have been investigated, the cause(s) of codon bias differences between species have not been explored. If codon bias is maintained under mutation-selection balance, then codon bias evolution should be sensitive to small changes in either selection intensity or mutational biases. Major codon usage is significantly lower in *D. melanogaster* than in its sibling species, *D. simulans* (AKASHI 1995). In this study, comparisons of the G+C content of coding and intron sequences are shown to support a reduction in selection intensity for codon bias in the *D. melanogaster* lineage. Similar lineage effects in protein evolution suggest that both amino acid changes and insertion/deletion mutations in coding regions may be weakly selected.

METHODS AND RESULTS

Quantitative model of major codon preference: The simplest model of major codon preference considers twofold redundant codons in a haploid organism (LI 1987; BULMER 1991; EYRE-WALKER and BULMER 1995). Mutations occur at rates v from nonmajor codons to major codons and u in the opposite direction. Selection intensity is assumed to be constant within a given gene and evolution at all sites is independent (no linkage). Setting the forward and backward substitution rates equal, such that $N_e u \ll 1$ and $N_e v \ll 1$, gives an expression for the steady-state proportion of major codons in a given gene (LI 1987; BULMER 1991)

$$q = \frac{e^{2N_e s}}{e^{2N_e s} + u/v} \tag{1}$$

The frequency of major codons, q, is a function of two parameters: the product of effective population size and selection intensity, $N_e s$, and the ratio of the forward and backward mutation rates, u/v. Note that the *ratio* of the mutation rates, rather than their absolute values, determine levels of codon bias. Equation (1) also applies to diploid organisms under semidominant fitness effects by replacing $2N_e s$ with $4N_e s$. Major codon usage, MCU, is the frequency of major codons within a given gene and corresponds to q in Equation (1). Relatively small changes in either $N_e s$ or u/v will cause major codon usage to reach a new equilibrium value.

Codon bias reduction in *D. melanogaster*: If selection intensity, mutation rates, and population sizes remain relatively constant, major codons will reach an equilibrium frequency and the numbers of unpreferred and preferred fixations will be equal in a given lineage. Table 1 shows the number of synonymous fixations in the *D. melanogaster* and *D. simulans* lineages for eight genes. Data for three genes are added to a previous analysis (AKASHI 1995). Two or more alleles of these genes have been sequenced from both *D. melanogaster* and *D. simulans* and from at least one outgroup species within the *D. melanogaster* subgroup. Parsimony assumptions and outgroup sequences were used to infer the lineage in which fixations have occurred. For a given codon encoding different nucleotides in *D. melanogaster* and *D. simulans*, substitutions were assigned to minimize the number of changes in the phylogenetic tree. Codons for which multiple trees give the fewest number of changes were not included in the analysis. The

TABLE 1

Synonymous fixations in *D. melanogaster* and *D. simulans*

	mel		*sim*	
Gene	Unpreferred	Preferred	Unpreferred	Preferred
Adh	1	0	0	0
Adhr	5	1	1	1
Amy	5	1	0	0
boss	7	0	1	1
Mlc1	3	0	0	1
per	10	0	3	2
Pgi	12	3	1	4
Rh3	6	0	3	0
Total	49	5	9	9

Sequences were aligned using the CLUSTAL (HIGGINS and SHARP 1988) method with the default parameter settings in the DNAStar computer application. Major codons are defined in AKASHI (1995). The number of unpreferred mutations for the *per* locus differs from that in AKASHI (1995) because the CLUSTAL method gives a slightly different alignment than the method used in the previous analysis. GenBank/EMBL DNA sequence library (release 81.0) accession numbers or references are: *Adh* (*mel*-11: M17827-28, M19547, M17830-37, M22210, *sim*-6: M36581, X00607, X57361-64, *ere:* X54116, *ore:* M37837, *yak*-13: X54120, X57365-76, *tei:* X54118). *Adhr* [*mel*-11: (KREITMAN and HUDSON 1991), *sim*-5: (SUMNER 1991), *ere:* X54116, *yak:* X54120, *tei:* X54118]. *Amy* (*mel*-18: L22716-35, *sim*-2: D17733-34, *ere*-2: D17727-28, *ore*-2: D21128-29, *yak*-2: D17737-38, *tei*-2: D17735-36). *boss* [*mel*-6, *sim*-5, *tei*-3, *yak*-4 (AYALA and HARTL 1993)]. *Mlc1* (*mel*-16: L37312-27, *sim*-9: L08051, L49010-17, *tei:* L49008, *yak:* L49007). *per* (*mel*-6: L07817-19, L07821, L07823, L07825, *sim*-6: L07826-32), yak: X61127). *Pgi* (*mel*-11: L27539-46, L27553-55, *sim*-14 L27547-52, U20556-59, U20564-65, *yak*-13: L27673-85, *tei*-1: J. H. MCDONALD, personal communication). *Rh3* [*mel*-6, *sim*-5, *tei*-5, *yak*-5: (AYALA *et al.* 1993)]. *mel, sim, yak, tei, ere,* and *ore* refer to *D. melanogaster, D. simulans, D. yakuba, D. teisseri, D. erecta,* and *D. orena,* respectively. The number of alleles examined for each gene in each species is shown.

two classes of synonymous changes have substituted at equal rates in *D. simulans*; the null hypothesis of equilibrium codon bias is not rejected. In *D. melanogaster*, all eight genes show an excess of unpreferred fixations. Summing over all loci, unpreferred fixations outnumber preferred fixations by almost tenfold. A *G*-test for goodness of fit with WILLIAMS' correction (SOKAL and ROHLF 1981, p. 705) shows a strong departure from the expectation of equal rates of fixation ($G = 41.4$, $P < 10^{-5}$); major codon usage has undergone a dramatic reduction in *D. melanogaster*.

Major codon usage can also be compared between pairs of homologous genes without reference to an outgroup. Such analyses allow comparison of a larger number of genes between *D. melanogaster* and *D. simulans*. If codon bias has remained at equilibrium between two species, then the number of homologous codons at which one species encodes a major codon and the other encodes a nonmajor codon will equal the number of codons in the opposite configuration (AKASHI 1995). Major codon usage is compared between *D. melanogaster* and *D. simulans* for 34 genes in Table 2. The number of codons at which *D. melanogaster* encodes a major codon and *D. simulans* encodes a nonmajor codon is lower than the number of codons in the opposite configuration. WILCOXON's signed-ranks test (1945) reveals a highly significant difference ($z = 3.66$, $P = 0.003$). Codon bias differences appear to be genome-wide; of 34 genes examined, 25 show higher major codon usage in *D. simulans*, seven show higher bias in *D. melanogaster*, and two show no difference between the species. Pairwise analyses establish consistent differences in major codon usage between these species but do not address whether codon bias has increased in one lineage or decreased in the other. The excess of unpreferred fixations in all eight genes examined in Table 1 suggests a genome-wide reduction of major codon usage in the *D. melanogaster* lineage since its split from the common ancestor to *D. simulans*.

Since pairwise analyses examine a single allele at each locus, the observed differences at silent sites include mutations that have gone to fixation in each lineage and mutations that have accumulated in the sampled sequence since the most recent common ancestor (MRCA) of the alleles segregating within the species. Selection is less effective at preventing weakly selected mutations from reaching appreciable frequencies within populations than at preventing fixation of such changes (KIMURA 1983); the rate at which unpreferred mutations accumulate in a given allele since the most recent common ancestor will be greater than the fixation rate. Under the null hypothesis of no change in major codon usage between two species, the lineage showing a deeper coalescent time (greater time to MRCA) will show a greater proportion of unpreferred changes. In the *D. simulans* and *D. melanogaster* lineages, the time to coalescence for alleles within the species are within an order of magnitude of the time over which fixed differences have accumulated (SAWYER and HARTL 1992; HUDSON *et al.* 1994; AKASHI 1995). Estimates of DNA heterozygosities from worldwide samples suggested coalescent times three to five times longer in *D. simulans* than in *D. melanogaster* (AQUADRO 1992) but recent evidence for higher levels of nucleotide variation in east African populations of *D. melanogaster* suggest similar times to the MRCA (BEGUN and AQUADRO 1993, 1995). Thus, *D. melanogaster* and *D. simulans* alleles should show a similar proportion of segregating nonmajor codons. In comparisons between more distantly related species, the contribution of segregating polymorphism to the number of codon bias differences will be small.

Test for changes in mutational biases in *D. melanogaster*: Under the major codon preference model, codon bias reduction in *D. melanogaster* could result from a change in mutational biases or from a reduction in selection intensity at silent sites. Under major codon preference, changes in these parameters make distinguishable predictions for changes in the base composition of sites under different selection pressures. Figure 1a shows the expected difference in codon bias between species under different mutational biases. The change in codon bias is largest for low codon bias genes, remains relatively flat as MCU increases, and decreases in more highly biased genes. Figure 1b shows the expected difference in codon bias between species differing in selection intensity at silent sites. Changes in N_es predict small differences in codon bias in low codon bias genes and greater differences as selection intensity for codon bias increases. The change in codon bias decreases in very highly biased genes, but few loci show such high major codon usage (Table 2).

Figure 1a illustrates that regions under the weakest selection pressure for base composition show the greatest sensitivity to changes in mutational biases. Since all major codons end in either G or C, the mutational hypothesis predicts a greater decrease in G+C content at neutrally evolving sites than at silent sites in coding regions in *D. melanogaster*. Changes in N_es predict the opposite pattern, a larger excess of G/C→A/T changes in coding regions. G/C↔A/T changes were compared between coding sequences and putatively neutral intron sequences between *D. melanogaster* and *D. simulans*. Because introns were difficult to align with outgroup sequences, pairwise comparisons were made at sites fixed in each species but differing between species. Under the null hypothesis of no difference in base composition, the number of sites at which *D. melanogaster* encodes G or C and *D. simulans* encodes A or T ($mel_{GC}sim_{AT}$) should equal the number of sites in the opposite configuration ($mel_{AT}sim_{GC}$). This test is analogous to the pairwise comparisons of MCU above. Among the eight genes for which multiple alleles have been sequenced in *D. melanogaster* and *D. simulans*, introns show little evidence for a change in base composition: 33 intron sites are $mel_{AT}sim_{GC}$ and 28 are $mel_{GC}sim_{AT}$. Silent sites within exons, however, show significant differences in base compositions between these species: 72 sites are $mel_{AT}sim_{GC}$ whereas 18 are *melGCsimAT* (*G*-test for independence with WILLIAM's correction, $G = 11.82$, $P < 0.01$).

Pairwise comparisons allow a much larger number of regions to be examined. The base compositions of 46 intron sequences (Table 3) and 34 coding regions were compared between *D. melanogaster* and *D. simulans*. For these genes, a single allele was examined in each species. Although this comparison does not distinguish between fixed and segregating mutations, similar times to the MRCA in *D. melanogaster* and *D. simulans* suggest that such a procedure will not bias the results (see above). Overall, the G+C content of introns is similar in *D. melanogaster* (%GC = 38.6) and *D. simulans* (%GC = 39.1). Similar to the analyses of fixed differences, the ratio of $mel_{AT}sim_{GC}$ to $mel_{GC}sim_{AT}$ sites is significantly higher at silent sites (377/231) than in introns (252/214), ($G = 6.82$, $P < 0.01$). Mutational differences do not explain codon bias reduction in *D. melanogaster*.

These analyses assume that some intron sites evolve neutrally but do not assume that intron sequences are completely free of functional constraint. A number of findings suggest that intron sequences affect mRNA splicing and gene expression levels (MOUNT *et al.* 1992; SCHAEFFER and MILLER 1993; STEPHAN and KIRBY 1993; LAURIE and STAM 1994; LEICHT *et al.* 1995). However, relatively high rates of nucleotide substitution (HUDSON *et al.* 1994) and insertion/deletion evolution (Table 3), suggest that introns contain a large fraction of unconstrained sites.

Reduced selection intensity at silent sites in *D. melanogaster*: The excess of G/C→A/T mutations in coding regions of *D. melanogaster* suggests a reduction in selection intensity

TABLE 2

Codon bias comparisons between *D. melanogaster* and *D. simulans*

Gene	Map	Codons	MCU *mel*	MCU *sim*	$m_{maj}s_{non}$	$m_{non}s_{maj}$	$m_{maj}s_{non} - m_{non}s_{maj}$
ac	1B1-2	198	44.4	47.0	4	9	−5
Acp26Aa	26A	181	37.0	37.6	4	5	−1
Acp26Ab	26A	87	42.5	44.8	0	2	−2
Act88-F	88F	371	73.3	75.7	4	13	−9
Adh	35B3	255	74.5	76.1	3	7	−4
Adhr	35B3	270	41.1	43.7	4	11	−7
Amy-d	54A1-B1	480	82.5	82.9	13	15	−2
ase	1B3-4	342	41.8	42.4	4	6	−2
bcd	84A1	29	37.9	37.9	0	0	0
boss	96F8-11	518	54.4	55.8	9	16	−7
ci	102A3	294	21.8	21.4	6	5	1
cta	40-F	315	26.7	26.0	6	4	2
Est-6	69A1-5	520	41.5	44.8	10	27	−17
eve	46C3-11	58	65.5	69.0	0	2	−2
GstD1	87B8-9	193	82.4	86.0	1	8	−7
Hsc70-1	70C	103	63.1	64.1	0	1	−1
Hsp83	63B9	375	73.6	74.9	4	9	−5
Mlc1	98B	167	65.9	68.9	0	5	−5
MtnA	85E10-15	40	72.5	72.5	0	0	0
per	3B1-2	556	72.5	73.0	16	19	−3
Pgd	2D3-4	474	71.1	70.7	21	19	2
Pgi		555	71.9	74.1	7	19	−12
pn	2E2-3	357	52.9	54.1	8	12	−4
ref(2)p	37E2-F4	563	48.7	48.3	12	10	2
Rh3	92D1	381	58.0	56.7	12	7	5
sala	33A	131	23.7	24.4	5	6	−1
Sgs3	68C3-5	25	44.0	40.0	1	0	1
Sod	68A8-9	154	68.8	70.8	3	6	−3
su(f)	20E-F	730	49.2	50.8	13	25	−12
tra	73A8-9	171	38.0	37.4	7	6	1
v	10A	379	58.8	61.2	9	18	−9
Yp2	9A-B	342	76.0	76.9	5	8	−3
z	3A3	262	65.6	67.6	4	9	−5
Zw	18D12-13	499	77.6	80.2	6	19	−13
Totals		10,375			201	328	−127

Gene symbols and cytological locations are from FlyBase (1995). The CLUSTAL method was used to align coding sequences. Codons refers to the number of homologous codons encoding the same amino acid in each alignment. MCU is calculated for homologous codons which encode the same amino acid in the two species. $m_{maj}s_{non}$ refers to the number of codons that encode a major codon in *D. melanogaster* and a nonmajor codon in *simulans*. $m_{non}s_{maj}$ refers to codons that encode a nonmajor codon in *D. melanogaster* and a major codon in *simulans*. For loci with multiple alleles sequenced within a species, a single sequence was chosen at random for this analysis. Six genes are added to a previous comparison of major codon usage between these species (Akashi 1995). Some numbers listed in this table differ from those in Table 4 of Akashi (1995) because different alleles were chosen (at random) for this analysis. GenBank/EMBL accession numbers (release 91.0) or references for these genes are: *ac:* M17120, X62400, *Acp26Aa:* Y00219, X70899, *Acp26Ab:* Y00219, X70899, *Act88-F:* M18830, M87274, *Adh:* M17827, X57361, *Adhr:* (Kreitman and Hudson 1991; Sumner 1991), *Amy-d:* L22728, D17733, *ase:* X52892, J. Hey (personal communication), *bcd:* X07870, M32123, *boss:* (Ayala and Hartl 1993), *cid:* X54360, (Berry *et al.* 1991), *cta:* M. Wayne (personal communication), *Est-6:* J04167, L34263, eve: M14767, U32092, *GstD1:* X14233, M84577, *Hsc70-1:* L01501, J01087-9, *Hsp83:* X03810, X03811, *Mlc1:* M10125, L08051, *MtnA:* M12964, M55407, *per:* L07817, L07829, *pgd:* M80598, U02288, *Pgi:* L27539, L27550, *Pn:* Z12141, (Simmons *et al.* 1994), *ref(2)p:* X16993, U23930, *Rh3:* (Ayala *et al.* 1993), *sala:* X57474, M21227, *Sgs3:* (Martin *et al.* 1988), *sod:* X13780, X15685, *su(f):* X62679, L09193, *tra:* M17478, X66930, *v:* M34147, U27204, *Yp2:* L14421, L14426, *z:* L13043, L13049, *Zw:* L13887, L13876. Accession numbers or references are listed for *D. melanogaster* and *simulans*, respectively, for each gene.

at silent sites. The magnitude of changes in $N_e s$ necessary to explain the observed ratio of preferred to unpreferred substitutions was determined by computer simulation.

Simulations of silent fixations under major codon preference require estimates of three parameters: the relative mutation rates between major and nonmajor codons, the number of mutable sites in a given gene, and the intensity of selection for major codons. The relative mutation rates, u/v, for the transitions, A↔G and T↔C, were estimated from the base composition of low codon bias genes. In genes showing the lowest frequency of major codons at fourfold redundant sites, the base composition at twofold redundant sites should reflect

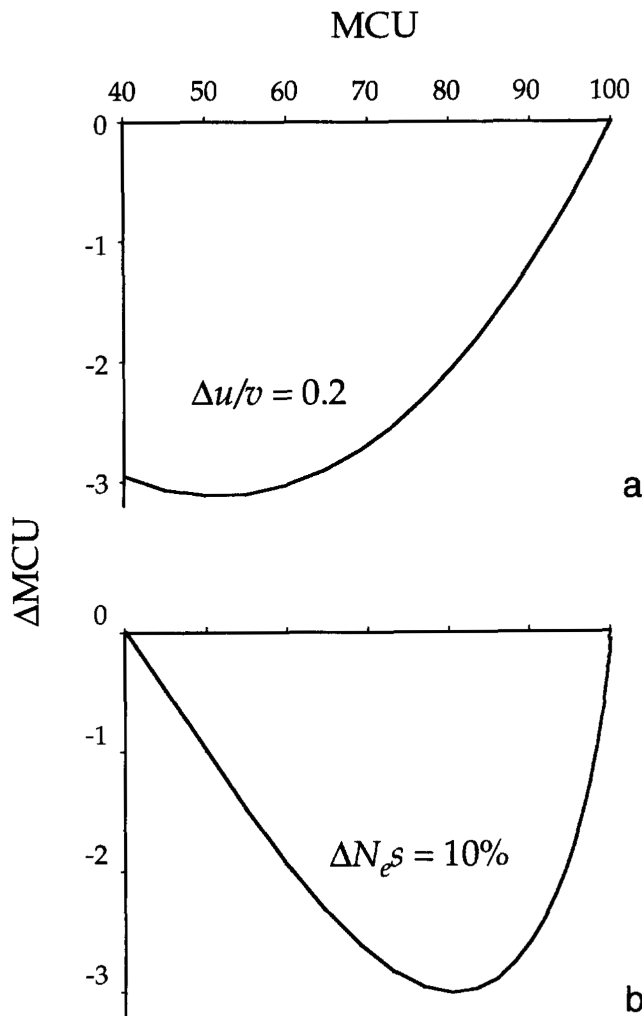

FIGURE 1.—Sensitivity of major codon usage to changes in mutational biases and selection intensity. The x-axis plots an initial state of major codon usage and the y-axis shows the expected percent decline in codon bias for a given parameter change. (a) ΔMCU is shown for an increase of u/v from 1.5 to 1.7. u/v of 1.5 gives an equilibrium mutational base composition of 60% AT, the average base composition of *D. melanogaster* introns (SHIELDS *et al.* 1988; MORIYAMA and HARTL 1993). (b) ΔMCU is plotted for a 10% decrease in N_es. The relationships were determined from Equation (1) and assume that codon bias evolution has reached a steady-state and that changes in mutational bias or selection intensity are uniform across loci.

neutral equilibrium frequencies. Among the 575 *D. melanogaster* loci analyzed in AKASHI (1995), the 20 genes showing the lowest MCU at fourfold redundant sites show base compositions of 58.1% A at A/G twofold redundant sites and 57.9% T at T/C sites. This suggests $u/v \approx 1.5$ for both transition types. Because u/v could only be estimated for transitions, the analyses were limited to the 23 transitions between major and nonmajor codons.

Ancestral codons were inferred at the node connecting *D. melanogaster* and *D. simulans* for the eight genes examined in Table 1. Table 4 shows the numbers of major codons in the ancestral sequence and the observed ratio of preferred and unpreferred substitutions in the *D. melanogaster* lineage. N_es values for the ancestral sequences were calculated from the MCU values using Equation (1).

In the simulations, silent mutations were fixed in each gene until the number of substitutions equaled that observed in the *D. melanogaster* lineage. The probability of fixation for each class of silent mutations equaled the product of the number of mutable sites, the relative probability of mutation at the sites, and the probability of fixation given N_es for a particular simulation. N_es values shown in Table 4 were reduced by a given fraction, and the reduction was assumed constant for the period that mutations accumulated. Fixation probabilities were calculated according to KIMURA (1962). Ratios of preferred to unpreferred substitutions were summed over the eight genes for each iteration, and the distribution of this ratio was determined from 10,000 iterations.

A reduction of N_es is sufficient to explain patterns of silent evolution in the *D. melanogaster* lineage. The ratio of unpreferred to preferred substitutions (42/4) in *D. melanogaster* fell within the 95% range of the simulated values when N_es was reduced to 20% of the ancestral N_es values. This suggests an upper limit of N_es in *D. melanogaster* relative to that in *D. simulans*. However, the data are also consistent with strictly neutral evolution in the *D. melanogaster* lineage following the split with *D. simulans* ($P = 0.20$); the lower bound on N_es in *D. melanogaster* is 0. Other selective forces are not required to explain the change in codon bias in *D. melanogaster* (*i.e.*, positive selection coefficients for mutations from major to nonmajor codons).

To confirm a reduction in selection intensity in *D. melanogaster*, the distribution of unpreferred fixations among different codons was also examined. Under neutral synonymous DNA evolution, the distribution of substitutions should fit a mutational expectation. If mutation rates are assumed to be equal at all sites, then the expected numbers of substitutions are the products of the relative frequencies of each major codon in the ancestral sequences and the total number of substitutions (Table 5). The observed and expected frequencies give a G statistic for goodness of fit of 19.16. Because the expected values for these data are small, the distribution of the test statistic was determined from simulations of neutral evolution. In each iteration, 42 fixations were allowed to occur, each with the probability of a particular unpreferred change equal to the frequency of the ancestral major codon. 73% of the 10,000 simulations showed equal or greater G values than that observed in the data; rates of substitution for unpreferred changes are similar across synonymous families in *D. melanogaster*. Both the ratio of unpreferred to preferred fixations and the distribution of fixations among codons suggest essentially neutral synonymous DNA evolution in *D. melanogaster*.

Accelerated protein evolution in *D. melanogaster*: Table 6 shows the numbers of replacement and silent fixations in the *D. melanogaster* and *D. simulans* lineages since their split from a common ancestor. Because a single allele was examined in each of the species, these numbers combine mutations that have fixed in each lineage and those segregating within species. As noted above, this should not bias the comparison of rates of fixation. Of the 14 loci examined, eight show a greater number of replacement changes in *D. melanogaster*, one shows faster evolution in *D. simulans*, and five show no difference between these species (Wilcoxon's signed-ranks test, $z = 2.14$, $P = 0.032$). Although rates of silent evolution are faster in the *D. melanogaster* lineage (Table 1), the ratio of replacement to silent substitution rates is higher in *D. melanogaster* than in *simulans* (G test for independence with WILLIAMS' correction for the pooled data, $G = 7.44$, $P = 0.007$). The pattern is similar in the eight loci for which segregating mutations can be separated from fixed differences; 18 amino acids have fixed in the *D. melanogaster* lineage and only a single replacement substitution has occurred in *D. simulans*. Proteins are not evolving neutrally in these species. This pattern in isolation,

TABLE 3

Intron comparisons between *D. melanogaster* and *D. simulans*

Intron	Length			Percentage GC				
	mel	*sim*	*mel* − *sim*	*mel*	*sim*	$m_{GC}s_{AT}$	$m_{AT}s_{GC}$	$m_{GC}s_{AT} - m_{AT}s_{GC}$
Acp26Aa int 1	56	56	0	32.1	33.9	0	1	−1
Acp26Ab int 1	61	61	0	36.1	32.8	4	2	2
Act88-F int 1	60	63	−3	48.3	43.3	7	4	3
Adh int 1	65	61	4	47.5	45.8	1	0	1
Adh int 2	70	68	2	25.0	27.9	2	4	−2
Adhr int 1	426	427	−1	34.1	34.1	7	7	0
Adhr int 2	52	51	1	29.4	25.5	5	3	2
ci int 1	64	64	0	15.6	15.6	1	1	0
ci int 2	53	54	−1	18.9	18.9	2	2	0
Est-6 int 1	51	51	0	26.0	22.0	2	0	2
eve int 1	71	70	1	41.4	37.1	3	0	3
Mlc1 int 1	153	154	−1	41.8	43.1	2	4	−2
Mlc1 int 2	985	984	1	39.4	40.9	7	21	−14
Mlc1 int 3	60	60	0	41.7	45.0	1	3	−2
Mlc1 int 4	350	345	5	43.3	43.3	3	3	0
Mlc1 int 5	113	114	−1	47.8	48.7	3	4	−1
Mtn int 1	265	228	37	33.8	36.4	4	10	−6
per int 1	59	69	−10	50.8	55.9	1	4	−3
per int 2	62	60	2	53.3	51.7	2	1	1
per int 3	62	62	0	51.6	53.2	1	2	−1
Pgd int 1	75	75	0	50.7	56.0	1	5	−4
Pgd int 2	1419	1423	−4	46.9	46.0	31	19	−12
Pgi int 1	245	245	0	48.6	49.0	1	2	−1
Pgi int 2	348	358	−10	31.3	31.3	15	15	0
Pgi int 3	62	62	0	30.6	38.7	0	5	−5
Pgi int 4	55	55	0	41.8	41.8	2	2	0
ref(2)p int 1	631	640	−9	34.1	33.7	24	21	3
ref(2)p int 2	55	55	0	29.1	32.7	0	2	−2
sala int 1	57	58	−1	23.2	17.9	3	0	3
Sgs3 int 1	73	74	−1	43.8	45.2	1	2	−1
Sod int 1	725	730	−5	35.5	34.3	23	14	9
su(f) int 1	61	61	0	29.5	31.1	1	2	−1
su(f) int 2	164	186	−22	28.6	30.4	7	10	−3
su(f) int 3	58	58	0	34.5	41.4	0	4	−4
su(f) int 4	300	299	1	44.6	46.0	1	5	−4
su(f) int 5	65	64	1	29.7	31.2	1	2	−1
su(f) int 6	56	56	0	20.4	20.4	2	2	0
su(f) int 7	449	449	0	29.4	34.0	24	44	−20
su(f) int 8	57	57	0	33.3	33.3	2	2	0
tra int 1	248	248	0	42.7	43.1	4	5	−1
tra int 2	57	54	3	24.1	22.2	1	0	1
Yp2 int 1	67	63	4	39.7	39.7	3	3	0
z int 1	121	120	1	39.2	38.3	5	4	1
z int 2	62	62	0	37.1	37.1	1	1	0
Zw int 2	62	62	0	51.6	53.2	1	2	−1
Zw int 3	85	85	0	38.8	40.0	2	3	−1
				38.6	39.1	214	252	−38

Introns were aligned using the CLUSTAL method. The length of introns in the two species (and differences in length) are shown. The GC content is shown for nucleotides that do not align with gaps. $m_{GC}s_{AT}$ refers to the number of sites that show a G or C in *D. melanogaster* and an A or T in *simulans*. $m_{AT}s_{GC}$ refers to the number of sites that show an A or T in *D. melanogaster* and a G or C in *simulans*. The results are not sensitive to the method of alignment; WILBUR-LIPMAN (WILBUR and LIPMAN 1982) and MARTINEZ-NEEDLEMAN-WUNSCH (NEEDLEMAN and WUNSCH 1970; MARTINEZ 1988) methods using the default parameter settings in the DNAStar computer application give similar results (data not shown). GenBank/EMBL accession numbers are listed in Table 1. *ci* intron sequences are from A. BERRY (personal communication).

however, does not elucidate the mechanism of protein evolution. Accelerated amino acid substitutions in *D. melanogaster* could reflect either the fixation of deleterious amino acid mutations ($N_e s \approx -1$) or a faster rate of adaptive evolution (discussed below).

An interesting exception to this pattern has been found at the *Zw* gene. EANES (1994) noted a roughly twofold faster rate of amino acid fixations at this locus in the *D. simulans* lineage than in *D. melanogaster*. *Zw* also shows lower ratios of polymorphism to divergence for replacement than for silent

TABLE 4

Silent-transition fixations in the *D. melanogaster* lineage

Gene	Codons	MCU	$N_e s$	Preferred	Unpreferred
Adh	253	0.850	1.07	1	0
Adhr	211	0.583	0.37	4	1
Amy	460	0.959	1.78	4	0
boss	463	0.706	0.64	6	0
Mlc1	90	0.811	0.93	3	0
per	472	0.928	1.48	9	0
Pgi	531	0.838	1.03	11	3
Rh3	312	0.788	0.86	4	0

The numbers of slient transition sites and the number of inferred fixations are shown. GenBank/EMBL accession numbers and the method of inferring ancestral states are given in Table 1.

mutations (EANES *et al.* 1993), suggesting adaptive protein evolution (MCDONALD and KREITMAN 1991). Sequence data for the outgroup, however, were not available and this gene was not included in the analysis.

Frequency distributions of amino acid polymorphisms in *D. melanogaster* and *D. simulans* are skewed toward rares: Negative and positive directional selection can be distinguished by examining the frequency distributions of mutations segregating within species (OHTA 1975; SAWYER *et al.* 1987). Deleterious amino acid changes should show an excess of rare variants segregating within species, whereas positive selection predicts more common variants. This prediction was tested by comparing the frequency distributions of silent and replacement changes in the loci listed in Table 1. Table 7 shows the numbers of newly arising silent and replacement polymorphisms observed in a single allele in the sample and those found in multiple alleles. All classes of silent mutations were pooled. The newly arising mutations were identified using parsimony analyses and outgroup sequences as described above. Amino acid changes show an excess of singletons relative to silent changes in both *D. melanogaster* and *D. simulans.* A MANTEL-HAENSZEL test with a correction for continuity (MANTEL and HAENSZEL 1959; MANTEL 1963) shows a highly significant departure from equal frequency distributions for the combined data ($z = 2.82$, $P < 0.003$). Amino acid polymorphisms in these species are either deleterious or under a form of balancing selection that skews the frequency distribution toward rare variants (see GILLESPIE 1994a).

TABLE 5

Distribution of unpreferred fixations in *D. melanogaster*

Change	Sites	Expected	Observed
TTC → TTT	74	2.0	1
TAC → TAT	57	1.5	2
TGC → TGT	39	1.0	1
TCG → TCA	35	0.9	0
TCC → TCT	66	1.8	0
ATC → ATT	70	1.9	4
AAG → AAA	96	2.5	3
AAC → AAT	71	1.9	1
AGC → AGT	56	1.5	0
ACC → ACT	74	2.0	1
GTC → GTT	62	1.6	2
GTG → GTA	82	2.2	1
GAC → GAT	94	2.4	4
GAG → GAA	96	2.5	2
GGC → GGT	89	2.4	5
GCC → GCT	92	2.4	1
CTG → TTG	114	3.0	2
CTG → CTA	114	3.0	4
CTC → CTT	29	0.8	2
CAG → CAA	76	2.0	1
CAC → CAT	37	1.0	2
CGC → CGT	48	1.3	2
CCC → CCT	46	1.2	1

Unpreferred substitutions among the 23 transitions between major and nonmajor codons are shown for the eight genes analyzed in Table 1. The expected values are the products of the relative frequencies of the ancestral codons and the total number of unpreferred fixations in the *D. melanogaster* lineage. GenBank/EMBL accession numbers are listed in Table 1.

TABLE 6

Silent and replacement fixations in the *D. melanogaster* and *D. simulans* lineages

	mel		*sim*	
Gene	Replacement	Silent	Replacement	Silent
Adh	1	7	0	3
Adhr	2	10	1	8
Amy-d	12	17	2	14
bcd	0	0	2	0
boss	3	19	0	16
GstD1	4	4	4	5
Mlc1	0	3	0	2
per	2	27	0	20
Pgi	2	17	2	10
ref(2)P	28	18	7	12
Rh3	3	15	0	9
sala	5	5	4	8
Sgs3	1	0	1	2
tra	5	10	5	8
Totals	68	143	28	117

Parsimony assumptions and outgroup sequences from within the *D. melanogaster* subgroup were used to infer the lineage in which replacement fixations have occurred for 14 genes. The numbers of silent and replacement fixations are shown. The numbers of silent substitutions are pooled for preferred, unpreferred, and changes within classes. GenBank/EMBL accession numbers for *D. melanogaster* and *D. simulans* sequences are given in Table 2. GenBank accession numbers for outgroup sequences not listed in Table 1 are as follows: *bcd: t*-M32121, *GstD1: t*-M84579, *y*-M84580, *e*-M84576, *ref(2)p: e*-X69831, *sala: o*-M21579, *Sgs3: y, e* (MARTIN *et al.* 1988), *tra: e*-X66527. *t, y, e,* and *o* refer to *D. teissieri, D. yakuba, erecta,* and *D. orena,* respectively.

Larger proteins in *D. melanogaster* than *D. simulans*: Of the 34 genes examined in Table 2, six show differences in length, and all six are larger in *D. melanogaster* than in *D. simulans* (Table 8; signs test, $P = 0.032$). Overall, *D. melanogaster* encodes 29 more codons than *D. simulans.* Comparisons with outgroup sequences for *per, sala,* and *tra* suggest that this difference is due to length increases in the *D. melanogaster* lineage.

Differences in protein length could result from differences in insertion/deletion mutation rates or from differences in the fitness effects of protein size changes. Intron lengths were

TABLE 7

Frequency distributions of silent and replacement polymorphisms in *D. melanogaster* and *D. simulans*

	mel[a]		*sim*[b]	
Frequency	Replacement	Silent	Replacement	Silent
1	7	19	9	77
>1	0	21	1	52

To maximize both the number of genes and the number of alleles which could be examined, six alleles were examined in each of six genes in *D. melanogaster* and five alleles were examined in each of seven genes in *D. simulans*. *Adh* (KREITMAN 1983) and *Amy* (INOMATA *et al.* 1995) were not examined in *D. melanogaster* because the alleles were not sequenced randomly with respect to knowledge of electrophoretic variation. All GenBank accession numbers and references are as listed in Table 1 except the following genes for which a subset of the available alleles were examined: *Adh* (*sim*-5: M36581, X00607, X57362-64). *Mlc1* (*mel*-6: L37312-17, *sim*-5: L08051, L49010-13). *per* (*sim*-5: L07828-32). *Pgi* (*mel*-6: L27544-46, L27553-55, *sim*-5: L27547-51).

[a] $n = 6$.

[b] $n = 5$.

TABLE 8

Protein length differences between *D. melanogaster* and *D. simulans*

Gene	*mel*	*sim*	*mel* − *sim*
Acp26Aa	265	256	9
Est-6	545	543	2
Hsc-70-1	215	214	1
per	560	559	1
sala	143	140	3
tra	198	185	13
Total	1926	1897	+29

The number of codons aligned between *D. melanogaster* and *simulans* protein sequences are shown for the six loci in Table 2 which show differences in length. All length changes occurred within exons and do not appear to be the result of errors in identifying intron splice sites. GenBank/EMBL accession numbers or references are listed in Table 2.

examined to determine whether mutation pressure has increased genome size in the *D. melanogaster* lineage. *D. melanogaster* introns fall into two size classes: small introns <90 bp and larger introns which cover a large range of sizes from 90 to several thousand bp (MOUNT *et al.* 1992). STEPHAN *et al.* (1994) found that the length of smaller introns tend to be conserved within the *D. melanogaster* subgroup, suggesting stabilizing selection for intron size. Larger introns, however, have undergone numerous size changes and may be good candidates for neutral length evolution (STEPHAN *et al.* 1994). Lengths of 46 introns were compared between *D. melanogaster* and *D. simulans* in Table 3. 13 introns are larger in *D. melanogaster*, 13 are larger in *D. simulans*, and 20 are the same length in the two species (Wilcoxon's signed-ranks test, $z = 0.29$, $P = 0.77$). The similarity of introns length is not confined to small introns; of the 16 introns larger than 90 bp, five are larger in *D. melanogaster*, eight are larger in *D. simulans* and three are the same length in the two species (Wilcoxon's signed-ranks test, $z = 0.94$, $P = 0.34$). Introns show no evidence for differences in mutational processes affecting genome size between these species; natural selection appears to have played a role in protein size evolution.

Although consistent differences in protein lengths are difficult to explain under adaptive or deleterious evolution with respect to protein function, translational selection offers a simple explanation for such patterns. Longer proteins will entail a greater metabolic cost, and will be synthesized more slowly, than shorter proteins. Given insertion/deletion mutations with little effect on protein function, translational selection will favor smaller peptides. Energetically costly insertion mutations may be accumulating in *D. melanogaster* proteins.

DISCUSSION

Lineage effects for codon bias and evidence for nonneutral protein evolution: OHTA has suggested comparisons of rates of protein evolution to test whether a substantial fraction of amino acid changes are "slightly deleterious" (OHTA 1972, 1973). Fixation probabilities for mutations are a monotonically increasing function of the parameter $N_e s$ (KIMURA 1962). Given similar distributions of the fitness effects of newly arising mutations, substitution rates will vary in lineages experiencing different effective population sizes. Adaptive mutations will accumulate more quickly in *larger* populations, whereas deleterious changes will substitute at a greater rate in *smaller* populations. Because fixation probabilities decrease rapidly for negative $N_e s$, rate differences can only be detected for deleterious mutations with small fitness effects ($-3 < N_e s < 0$).

A number of taxa show patterns consistent with the accumulation of weakly deleterious mutations in smaller populations. At the Drosophila *Adh* locus, the replacement substitution rate (relative to the synonymous rate) is 40–50% higher in the presumably smaller populations of Hawaiian Drosophila than in the *D. melanogaster* and *D. obscura* groups (OHTA 1993). Similar rate accelerations in protein evolution have been noted by MORAN (1996) in the presumably smaller populations of endosymbiotic than free-living bacteria and by EASTEAL and COLLET (1994) in the presumably smaller populations of rodents than primates (but see LI *et al.* 1987; OHTA 1992 for an alternative interpretation).

Although OHTA's method is valid, knowledge of the historical effective population sizes of natural populations is often tenuous. Ecological considerations allow reliable estimates of N_e for only a few species and neutral DNA polymorphism can shed light on effective population sizes for the past $4N_e$ generations, far shorter than the amount of time investigated on most lineages. The major codon preference model, however, allows detection of relatively small changes in $N_e s$ at silent sites and knowledge of this parameter can be employed to identify mechanisms of protein evolution.

Consistent accelerations of protein evolution in lineages of reduced selection intensity at silent sites would suggest three possible explanations. Selection coefficients for deleterious protein changes may be of a simi-

lar magnitude as that of unpreferred silent changes and both could be accumulating in a population of smaller effective size. Alternatively, selection coefficients associated with both silent and amino acid changes may be smaller in a given lineage. This would be expected if amino acid mutations affect translational efficiency (*i.e.*, some amino acids are translated more quickly or more accurately than others) and the fitness effects of such changes differ between species. Finally, an excess of amino acid fixations could reflect greater pressure for adaptive protein evolution. Directional selection for replacement changes can accelerate genetic drift at linked silent sites (HILL and ROBERTSON 1966; FELSENSTEIN 1974; BARTON 1995); greater selection intensity for amino acid changes can *result* in a reduction in the effectiveness of selection for codon bias.

Distinguishing between adaptive and deleterious protein evolution requires within- as well as between-species sequence comparisons. Under KIMURA and OHTA's (1971) view of molecular evolution, the critical parameter governing the dynamics of mutations within species and substitution processes between species, $N_e s$, remains constant over evolutionary time. Three types of data can shed light on the sign and magnitude of $N_e s$: frequency distributions of polymorphic mutations, ratios of the numbers of segregating and fixed differences, and substitution rates between species. Negative selection coefficients for amino acid mutations will result in frequency distributions skewed toward rare polymorphisms (OHTA 1975; SAWYER *et al.* 1987), elevated ratios of polymorphism to divergence (BALLARD and KREITMAN 1994; NACHMAN *et al.* 1994; AKASHI 1995), and faster rates of substitution in low codon bias lineages. However, predictions for directional and balancing selection can overlap those of the slightly deleterious model. Balancing selection can give rise to an excess of rare variants (GILLESPIE 1994a) or higher ratios of polymorphism to divergence (WAYNE *et al.* 1996) compared with that of neutrally evolving sites. Adaptive protein evolution can cause a reduction in codon bias (see above). Particular models of directional or balancing selection, however, do not make the same predictions as slightly deleterious evolution for all three aspects of within- and between-species evolution. In the *D. melanogaster* and *D. simulans* lineages, amino acid mutations show frequency distributions skewed toward rares. In addition, both species show higher ratios of polymorphism to divergence for replacement than for silent changes; the sample sizes, however, are small and the latter trends are not statistically significant (data not shown). Confirmation of all three trends in a larger number of genes would strongly support that a substantial fraction of molecular variants are slightly deleterious.

Concave fitness functions and the evolution of near neutrality: Weak selection models of molecular evolution have been criticized for the *ad hoc* assumption of a large fraction of deleterious mutations with a narrow range of fitness effects (GILLESPIE 1994b). The suggestion above, that a detectable fraction of silent, amino acid, and protein length mutations incur fitness costs of $s \approx -1/N_e$ in *D. melanogaster*, may also seem implausible. Given estimates of effective population size on the order of 10^6 in this species (KREITMAN 1983), mutations with appreciable effects on fitness can fall in a range of selection coefficients of six orders of magnitude.

Concave fitness functions are biologically plausible and can lead to weak selection. Such functions arise from models of diminishing returns in fitness as a character evolves toward an "optimal" state. Theoretical studies of enzyme kinetics have shown that changes in enzyme activity can have a diminishing effect on reaction velocity or flux through a biochemical pathway (reviewed in HARTL *et al.*, 1985). HARTL *et al.* (1985) argued that neutral or nearly neutral molecular evolution could result from concave fitness functions for enzyme activity. Natural selection will affect the kinetic properties of an enzyme to the point where detectable changes in activity will have negligible effects on fitness (assuming that flux is proportional to fitness).

The HARTL-DYKHUIZEN-DEAN model can lead to either neutral evolution or mutation-selection-drift balance. GILLESPIE (1994b) showed that, under concave fitness functions, normal distributions of the fitness effects of new mutations about the "parental" value will lead to essentially neutral evolution. If, however, the distribution of fitness effects becomes skewed toward deleterious variants as a character approaches the optimum, then population fitness will reach an equilibrium at which the evolutionary dynamics of newly arising mutations will remain in balance between weak selection, mutation, and genetic drift. At this state, the number of newly arising deleterious mutations will outnumber the number of advantageous changes, but the fixation rates of the two classes will be equal and a relatively constant level of fitness will be maintained. Such a scenario may explain similar patterns of codon bias across species thought to differ substantially in effective population size (AKASHI 1995).

Under a model of concave fitness functions, changes in effective population size will result in evolution to a new steady-state level of fitness. If N_e decreases and the number of deleterious mutations is larger than that of advantageous changes, then rates of substitution will increase in smaller populations and decrease in larger ones. The accelerated rate of silent and replacement substitutions and the increase in protein sizes in *D. melanogaster* could reflect an approach to a lower point on a fitness surface. Interestingly, this model predicts that the accumulation of deleterious mutations will eventually cease and fitness will be maintained at a lower equilibrium level (given a constant N_e).

Although theoretically plausible, empirical evidence for concave fitness functions has been equivocal. Mutation accumulation studies suggested synergistic (greater

than multiplicative) fitness interactions among deleterious mutations in Drosophila (MUKAI 1969). However, MUKAI's results are complicated by the finding that mutations due to transposable element insertions may increase in laboratory populations (S. NUZHDIN, personal communication). Inbreeding studies suggest synergistic epistasis among mutations for some fitness components but not for others (see WILLIS 1993). Weak selection models of molecular evolution, however, require concave fitness surfaces as a character approaches an optimal state. Because small fitness differences are undetectable in natural or laboratory populations, it will be of interest to determine whether comparative sequence analyses can identify synergistic interactions among weakly selected mutations.

I am grateful to MARTY KREITMAN and JOHN GILLESPIE for extensive discussion of this work. I also thank PETER ANDOLFATTO, BRIAN CHARLESWORTH, ADAM EYRE-WALKER, CHUCK LANGLEY, ELI STAHL, WOLFGANG STEPHAN, and an anonymous reviewer for valuable suggestions which improved this manuscript. ANDREW BERRY, ANDREW CLARK, JODY HEY, JOHN H. MCDONALD, and MARTA WAYNE kindly provided unpublished DNA sequences. H.A. was a Howard Hughes Medical Institute Predoctoral Fellow and is currently funded by a postdoctoral fellowship from the Center for Population Biology at the University of California at Davis.

LITERATURE CITED

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

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

AQUADRO, C. F. 1992 Why is the genome variable? Insights from *Drosophila.* Trends Genet. **8:** 355–362.

AYALA, F. J., and D. L. HARTL, 1993 Molecular drift of the *bride of sevenless* (*boss*) gene in *Drosophila.* Mol. Biol. Evol. **10:** 1030–1040.

AYALA, F. J., B. S. W. CHANG, and D. L. HARTL, 1993 Molecular evolution of the *Rh3* gene in Drosophila. Genetica **92:** 23–32.

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

BARTON, N. H., 1995 Linkage and the limits to natural selection. Genetics **140:** 821–841.

BEGUN, D. J., and C. F. AQUADRO, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **356:** 519–520.

BEGUN, D. J., and C. F. AQUADRO, 1995 Molecular variation at the *vermilion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans.* Genetics **140:** 1019–1032.

BENNETZEN, J. L., and B. D. HALL, 1982 Codon selection in yeast. J. Biol. Chem. **257:** 3026–3031.

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

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.

CARULLI, J. P., D. E. KRANE, D. L. HARTL and H. OCHMAN, 1993 Compositional heterogeneity and patterns of molecular evolution in the Drosophila genome. Genetics **134:** 837–845.

CHARLESWORTH, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63:** 213–227.

EANES, W. F., 1994 Patterns of polymorphism and between species divergence in the enzymes of central metabolism, pp. 18–28 in *Non-Neutral Evolution: Theories and Molecular Data,* edited by B. GOLDING. Chapman and Hall, New York.

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.

EASTEAL, S., and C. COLLET, 1994 Consistent variation in amino-acid substitution rate, despite uniformity of mutation rate: protein evolution in mammals is not neutral. Mol. Biol. Evol. **11:** 643–647.

EYRE-WALKER, A., and M. BULMER, 1995 Synonymous substitution rates in enterobacteria. Genetics **140:** 1407–1412.

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

FLYBASE, 1995 The Drosophila Genetic Database. Available from the ftp.bio.indiana.edu network server and Gopher site.

GILLESPIE, J. H., 1994a Alternatives to the neutral theory, pp. 1–17 in *Non-Neutral Evolution: Theories and Molecular Data,* edited by B. GOLDING. Chapman and Hall, New York.

GILLESPIE, J. H., 1994b Substitutional processes in molecular evolution. III. Deleterious alleles. Genetics **138:** 943–952.

GOUY, M., and C. GAUTIER, 1982 Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10:** 7055–7064.

GRANTHAM, R., C. GAUTIER, M. GOUY, M. JACOBZONE and R. MERCIER, 1981 Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9:** r43–79.

GROSJEAN, H., and W. FIERS, 1982 Preferential codon usage in prokaryotic genes: the optimal codon-anti-codon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18:** 199–209.

HARTL, D. L., D. E. DYKHUIZEN and A. M. DEAN, 1985 Limits of adaptation: the evolution of selective neutrality. Genetics **111:** 655–674.

HIGGINS, D. G., and P. M. SHARP, 1988 CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene **73:** 237–244.

HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.

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., 1981 Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translation system. J. Mol. Biol. **151:** 389–409.

IKEMURA, T., 1982 Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. **158:** 573–597.

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

INOMATA, N., H. SHIBATA, E. OKUYAMA and T. YAMAZAKI, 1995 Evolutionary relationships and sequence variation of alpha-amylase variants encoded by duplicated genes in the Amy locus of *Drosophila melanogaster.* Genetics **141:** 237–244.

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.

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.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogater* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

LAURIE, C. C., and L. F. STAM, 1994 The effect of an intron polymor-

phism on alcohol dehydrogenase expression in *Drosophila melanogaster.* Genetics **138:** 379–385.

LEICHT, B. G., S. V. MUSE, M. HANCZYC and A. G. CLARK, 1995 Constraints on intron evolution in the gene encoding the myosin alkali light chain in Drosophila. Genetics **139:** 299–308.

LI, W.-H., 1987 Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. **24:** 337–345.

LI, W.-H., M. TANIMURA and P. M. SHARP, 1987 An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. **25:** 330–342.

MANTEL, N., 1963 Chi-squared tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. J. Am. Stat. Assoc. **58:** 690–700.

MANTEL, N., and W. HAENSZEL, 1959 Statistical aspects of the analysis of data from the retrospective analysis of disease. J. Natl. Cancer Inst. **22:** 719–748.

MARTIN, C. H., C. A. MAYEDA and E. M. MEYEROWITZ, 1988 Evolution and expression of the *Sgs-3* glue gene of Drosophila. J. Mol. Biol. **201:** 273–287.

MARTINEZ, H. M., 1988 A flexible multiple sequence alignment program. Nucleic Acids. Res. **16:** 1683–1691.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila.* Nature **351:** 652–654.

MORAN, N. A., 1996 Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93:** 2873–2878.

MORIYAMA, E. N., and D. L. HARTL, 1993 Codon usage bias and base composition of nuclear genes in *Drosophila.* Genetics **134:** 847–858.

MOUNT, S. M., C. BURKS, G. HERTZ, G. D. STORMO, O. WHITE *et al.,* 1992 Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids. Res. **20:** 4255–4262.

MUKAI, T., 1969 The genetic structure of natural populations of *Drosophila melanogaster.* VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. Genetics **61:** 749–761.

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.

NEEDLEMAN, S. B., and C. D. WUNSCH, 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48:** 443–453.

OHTA, T., 1972 Evolutionary rate of cistrons and DNA divergence. J. Mol. Evol. **1:** 150–157.

OHTA, T., 1973 Slightly deleterious mutant substitutions in evolution. Nature **246:** 96–98.

OHTA, T., 1975 Statistical analyses of *Drosophila* and human protein polymorphisms. Proc. Natl. Acad. Sci. USA **72:** 3194–3196.

OHTA, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **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.

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.

SAWYER, S. A., and D. L. HARTL, 1992 Population genetics of polymorphism and divergence. Genetics **132:** 1161–1176.

SCHAEFFER, S. W., and E. L. MILLER, 1993 Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura.* Genetics **135:** 541–552.

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, 1987 The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4:** 222–230.

SHARP, P. M., and W.-H. LI, 1989 On the rate of DNA sequence evolution in *Drosophila.* J. Mol. Biol. **28:** 398–402.

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.

SIMMONS, G. M., W. KWOK, P. MATULONIS and T. VENKATESH, 1994 Polymorphism and divergence at the *prune* locus in *Drosophila melanogaster* and *D. simulans.* Mol. Biol. Evol. **11:** 666–671.

SOKAL, R. R., and F. J. ROHLF, 1981 *Biometry.* W. H. Freeman and Co., New York.

STEPHAN, W., and D. A. KIRBY, 1993 RNA folding in Drosophila shows a distance effect for compensatory fitness interactions. Genetics **135:** 97–103.

STEPHAN, W., V. S. RODRIGUEZ, B. ZHOU, and J. PARSCH, 1994 Molecular evolution of the metallothionein gene *Mtn* in the *melanogaster* species group: results from *Drosophila annanassae.* Genetics **138:** 135–143.

SUMNER, C., 1991 Nucleotide polymorphism in Alcohol dehydrogenase duplicate of *Drosophila simulans*: implications for the neutral theory. Undergraduate thesis, Princeton University.

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.

WILBUR, W. J., and D. J. LIPMAN, 1982 Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80:** 726–730.

WILLIS, J. H., 1993 Effects of different levels of inbreeding on fitness components in *Mimulus guttatus.* Evolution **47:** 864–876.

WILCOXON, F., 1945 Individual comparisons by ranking methods. Biometrics Bull. **1:** 80–83.

Communicating editor: A. G. CLARK