Copyright © 1998 by the Genetics Society of America

The Correlation Between Synonymous and Nonsynonymous Substitutions in Drosophila: Mutation, Selection or Relaxed Constraints?

Josep M. Comeron and Martin Kreitman

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

Manuscript received January 22, 1998
Accepted for publication July 9, 1998

ABSTRACT

Codon usage bias, the preferential use of particular codons within each codon family, is characteristic of synonymous base composition in many species, including Drosophila, yeast, and many bacteria. Preferential usage of particular codons in these species is maintained by natural selection acting largely at the level of translation. In Drosophila, as in bacteria, the rate of synonymous substitution per site is negatively correlated with the degree of codon usage bias, indicating stronger selection on codon usage in genes with high codon bias than in genes with low codon bias. Surprisingly, in these organisms, as well as in mammals, the rate of synonymous substitution is also positively correlated with the rate of nonsynonymous substitution. To investigate this correlation, we carried out a phylogenetic analysis of substitutions in 22 genes between two species of Drosophila, *Drosophila pseudoobscura* and *D. subobscura*, in codons that differ by one replacement and one synonymous change. We provide evidence for a relative excess of double substitutions in the same species lineage that cannot be explained by the simultaneous mutation of two adjacent bases. The synonymous changes in these codons also cannot be explained by a shift to a more preferred codon following a replacement substitution. We, therefore, interpret the excess of double codon substitutions within a lineage as being the result of relaxed constraints on both kinds of substitutions in particular codons.

THE rate of synonymous evolution (K_s) in Drosophila genes is correlated with the base composition at synonymous sites and is negatively correlated with the degree of synonymous codon usage bias (Shields *et al.* 1988; Sharp and Li 1989; Moriyama and Gojobori 1992). In the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*, both of which have a strongly biased codon usage, the degree of this bias is strongly positively correlated with the level of expression of the genes. In addition, the prevalent, or major, codon in each codon family is concordant with the most abundant tRNA (Grosjean and Fiers 1982; Sharp *et al.* 1986; Sharp and Li 1987; Bulmer 1991). Both observations suggest that selection for codon bias acts at the level of translation. In Drosophila, selection pressures have also been mainly restricted to post-transcriptional events, and the nucleotide composition of synonymous sites within a gene cannot be explained by regional mutational biases (Moriyama and Hartl 1993; Kliman and Hey 1994). General translational efficiency has been proposed as the main factor determining both the rate of evolution and base composition at synonymous sites. In multicellular organisms, however, both the level of gene expression and tRNA abundances are difficult to quantify and can vary among tissues (Chevallier and Garel 1979; Garel 1982) and developmental stages (White *et al.* 1973).

In Drosophila, bacteria, and also mammals, the rate of nonsynonymous substitution (K_a) is also positively correlated with the rate of synonymous substitution (K_s) (Graur 1985; Sharp and Li 1987; Wolfe and Sharp 1993; Akashi 1994; Comeron and Aguadé 1996). In mammals, the genome is structured into isochores of different nucleotide composition. The synonymous sites composition, and hence the codon usage, and the introns, reflect the different overall base composition of each isochore (Aota and Ikemura 1986; Bernardi and Bernardi 1986). Nucleotide compositional differences across the genome have been interpreted as being the result of differences in mutational patterns among isochores (Ikemura 1985; Bulmer 1987; Filipsky 1987) and do not exhibit a clear relationship with the rate of synonymous substitution (Wolfe *et al.* 1989; Wolfe and Sharp 1993; Mouchiroud *et al.* 1995). Therefore, explanation of the correlation between K_s and K_a in mammals has focused on the possibility of double mutations at adjacent sites, hereafter called doublets. There is disagreement, however, as to whether doublet mutations are sufficiently common to explain the correlation (Ticher and Graur 1989; Wolfe and Sharp 1993; Mouchiroud *et al.* 1995; Ohta and Ina 1995).

The basis of the correlation between K_s and K_a in Drosophila and prokaryotes is largely unexplored. In Drosophila, Akashi (1994) found evidence for selection on translational accuracy by showing that the more

Corresponding author: Josep M. Comeron, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637. E-mail: jcomeron@midway.uchicago.edu

highly functionally and evolutionarily conserved amino acids within a gene also had greater codon bias than less conserved amino acids. He proposed that selection for translational accuracy would lead to a positive correlation between K_s and K_a. The positive correlation between K_s and K_a in species with codon bias selection can also be explained by a hypothesis proposed by Lipman and Wilbur (1985) (and rejected) for mammals (Wolfe and Sharp 1993; Mouchiroud *et al.* 1995). Under their hypothesis, an amino acid replacement change, possibly driven by positive selection, will then favor a synonymous mutation in that amino acid for a more preferred codon.

No attempt has been made to test hypotheses to explain the correlation between K_s and K_a in Drosophila. We test these hypotheses by investigating the evolution of codons with single synonymous and single nonsynonymous substitutions between *D. pseudoobscura* and *D. subobscura*.

MATERIALS AND METHODS

The genes and accession numbers used in this study are as follows: *Adh* (X78384, X62181, M15545), *Adhr* (X78384, Y00602, M55545), *A/A-T/sesB* (S43651, AF025798, AF025799), *Aprt* (M18432, L06281, AF025800), *Arr2/ArrB* (M32141, X54084), *Bcd* (X07870, X55735, X78058), *Cp15* (X02497, X53423; Benson 1995), *Cp16* (X16715, X53423), *Cp18* (X02497, X53423), *Cp19* (X02497, X53423; Benson 1995), *Cpy1* (M62398, AF025803, AF025804), *Ddc* (X04426; Wang *et al.* 1996), *Eno* (X17034, AF025805, AF025806), *Esterase6/5b* (J04167, M55907), *Gad1* (X76198, AF025807, AF025808), *Gapdh2* (M11255/256/259, AF025809, AF025810), *Gart/ade3* (J02527, X06285), *Gld* (M29298/X07358/X13581/582, M29299, AF025811), *Gpdh* (X67650, U59682; Wells 1996), *Mlc1* (M10125, L08052, AF025812), *Pcp* (J02527, X06285), *Rh1/ninaE* (K02135, X65877, AF025813), *Rh2* (M12896, X65878), *Rh3* (M17718, X65879), *Rp49/RpL32* (X00848, S59382, M21333), *sc* (M17119, X96479), *Sod* (M24421, U47871, U47888), *Sry-α* (X03121, L19536, L19535), *Sxl* (M23636, X98370), *Tpi* (X57576/S70377, AF025814, AF025815), *Ubx* (X05723/24/25/27, X05179), *Uro* (X51940, X57113/S94076, AF025816), *Xdh/ry* (Y00307/308, M33977, Y08237), *y* (X04427, Y13909), and *zen* (X68347, X78058). *Antp* and *hsp82* genes were not used in this study because for these genes the interspecific comparison is only possible for less than half of the entire coding sequence. Also, *ATPsyn-β* and *Vha*14 genes were not used in the analyses because they exhibit significant variation in the synonymous substitution rate ($P < 0.01$) between lineages (Zeng *et al.* 1998). All the other genes (longer than 100 codons) for which a comparison between *D. melanogaster* and both species of the obscura group, *D. subobscura* and *D. pseudoobscura*, is possible have been used in the analyses. The sequences were aligned after translation using CLUSTAL W (Thompson *et al.* 1994) with minor manual adjustments to eliminate unnecessary gaps. The numbers of synonymous (K_s) and nonsynonymous (K_a) substitutions per site were estimated as described in Comeron (1995). K_{s1} and K_{a1} also refer to the number of synonymous and nonsynonymous substitutions per site, respectively, but entail only those codons that differ among the homologous sequences by no or one position. The estimated numbers of substitutions per site were obtained by using the program K-Estimator v3.2 available upon request from J.M.C. or from ftp.bio.indiana.edu/molbio/mswin. Correlation probabilities were calculated by applying the *z*-transformation (z^*) suggested by Hotelling (Sokal and Rohlf 1995; Chap. 15) for dealing with small sample size ($n < 50$).

Computer simulations: We generated pseudorandom coding regions of 250 codons with a G + C content of 0.70 at third positions of codons. We assumed both selection coefficients on synonymous (s_s) and nonsynonymous (s_a) mutations to be constant within a gene and to follow a normal distribution of mean and variance of 1.1 and 0.35, and 4.4 and 1.4, respectively, among genes. We used the means $S_s = -1.1$ and $S_a = -4.4$ because they predict averages for K_s of 0.824 and for K_a of 0.082, very close to the observed averages between *D. melanogaster* and the obscura species, *D. subobscura* and *D. pseudoobscura*, assuming $K_n = 1.5$. The number of synonymous and nonsynonymous substitutions was obtained from a Poisson distribution with mean the product of the value K, predicted by s following Kimura (1983), and the number of sites under analysis. A transition:transversion ratio of 2:1 was applied. Partial correlations between s_s and nonsynonymous s_a were obtained by $s_s = f_c \times w \times N(a) + (1 - f_c) \times N(s)$, where f_c indicates the fraction of variance of s_s explained by s_a, w the average ratio s_s/s_a, and $N(a)$ and $N(s)$ the independent values obtained from the normal distribution of selection coefficients on nonsynonymous and synonymous mutations, respectively.

RESULTS

As expected, we observed a significant positive correlation between K_s and K_a between *D. melanogaster* and the obscura group species (*D. pseudoobscura* and *D. subobscura*) for the 35 genes under analysis ($r = 0.452$, $P = 0.006$; Table 1 and Figure 1). This correlation is also detected when those codons with three substitutions are not taken into account ($r = 0.437$, $P = 0.008$). In Drosophila, one or two codons are preferred in each codon family (Shields *et al.* 1988; Moriyama and Gojobori 1992; Akashi 1995), and they are always C- or G-ending. In genes with biased codon usage, most replacement changes will also be a mutation to another preferred codon for the new amino acid based on the definition of preferred codons proposed by Akashi (1995). Nevertheless, about 18% of replacement changes in a preferred codon generate an unpreferred codon; the exact percentage depends on the amino acid composition. Thus, one might expect natural selection to favor preferred synonymous mutations (from unpreferred to preferred codons or to a more frequent synonymous codon) in those codons where the first nonsynonymous substitution has been to an unpreferred codon (Lipman and Wilbur 1985). On an evolutionary time scale, this codon selection is expected to lead to a relative excess of codons with both synonymous and nonsynonymous substitutions. We will refer to this adaptive explanation as hypothesis i.

Two other hypotheses have been proposed to explain this correlation: (ii) an excess of double mutation at adjacent sites, *i.e.*, doublet mutations (Wolfe and Sharp 1993), and (iii) correlated selective constraints on synonymous and nonsynonymous sites over the entire gene, or (iiia) over particular codon positions, where less conserved proteins or amino acids also have relaxed constraints on synonymous sites (Lipman and Wilbur 1985;

TABLE 1

Divergence estimates for the 35 genes under study

Gene	K_s	K_a	K_{s1}[a]	K_{a1}[a]	No. bp	Sp[b]
A/A-T/sesB	0.2639	0.0312	0.2445	0.0108	891	su-ps
Adh	0.6502	0.0523	0.5734	0.0177	762	su-ps
Adhr	1.1317	0.0505	0.9785	0.0202	840	su-ps
Aprt	0.9421	0.0971	0.7939	0.0324	549	su-ps
Arr2	0.6347	0.0137	0.6102	0.0041	1203	ps
Bcd	0.8971	0.1534	0.7067	0.0292	1482	su-ps[c]
Cp15	0.9143	0.3070	0.6597	0.0443	345	su-ps
Cp16	0.7480	0.1621	0.5917	0.0760	414	su
Cp18	0.8740	0.2399	0.6201	0.0905	516	su
Cp19	0.6902	0.2817	0.5674	0.0761	519	su-ps
Cpy1	0.4633	0.0262	0.4485	0.0085	495	su-ps
Ddc	1.0931	0.0567	0.9581	0.0163	1425	su-ps
Eno	0.3471	0.0458	0.3117	0.0137	1299	su-ps
Est6-5b	1.2010	0.1707	0.9188	0.0617	1632	ps
Gad1	0.6496	0.0222	0.6084	0.0053	1383	su-ps
Gapdh2	1.0484	0.0156	1.0006	0.0051	996	su-ps
Gart/ade3	1.1170	0.0807	0.9579	0.0293	4059	ps
Gld	0.9607	0.0541	0.8395	0.0163	1836	su-ps
Gpdh	0.7899	0.0079	0.7617	0.0000	1200	su-ps
Mlc1	0.1495	0.0173	0.1496	0.0173	465	su-ps
Pcp	0.9545	0.1208	0.7035	0.0496	555	ps
Rh1/ninaE	0.5214	0.0170	0.5009	0.0115	1119	su-ps
Rh2	1.1360	0.0577	1.0420	0.0218	1143	ps
Rh3	0.9040	0.0469	0.8248	0.0230	1149	ps
Rp49/RpL32	0.5115	0.0265	0.4827	0.0164	402	su-ps
sc	1.0245[d]	0.3053	1.4804	0.0440	1035	su
Sod	0.7381	0.0857	0.6608	0.0194	459	su-ps
Sry-α	1.4396	0.2561	0.9508	0.0681	1590	su-ps
Sx1	0.4747	0.0414	0.4152	0.0185	1062	su
Tpi	0.6670	0.0510	0.6028	0.0107	741	su-ps
Ubx	0.5970	0.0518	0.5528	0.0082	1167	ps
Uro	0.8593	0.0840	0.7138	0.0281	1056	su-ps
Xdh/ry	1.0069	0.0739	0.9200	0.0218	4005	su-ps
y	1.2094[d]	0.0687	1.0821[d]	0.0132	1623	su
zen	1.2195	0.2475	0.7717	0.0434	1059	su

[a] The number of synonymous and nonsynonymous substitutions per site, respectively, when only those codons with no or only one substitution are taken into account.

[b] The obscura group species that have been compared to *D. melanogaster*; su, *D. subobscura*; and ps, *D. pseudoobscura*. Where both obscura species are available, divergence estimates are the averages.

[c] Divergence estimates are only from the *D. melanogaster-D. pseudoobscura* comparison as the *D. subobscura* sequence is only 93 codons long.

[d] Divergence estimates obtained by applying the Jukes and Cantor (1996) formula instead of the Kimura (1980) two-parameter method because of the inapplicability of the latter method.

Akashi 1994). Selection for translational accuracy has been suggested as a driving force for correlated selective constraints predicted under hypothesis iii (Akashi 1994).

Doublet mutations do not appear to be the main cause of the correlation between K_s and K_a in these genes. According to the doublet mutation hypothesis ii, almost half (47.8%) of adjacent mutations, involving one synonymous and one nonsynonymous mutation, are expected to be in the third position of a codon and in the first position of the next codon because most, but not all (95.4%), substitutions at the first position of a codon are nonsynonymous, and 59.8% of substitutions at the third position are synonymous (substitutions at the second position are always nonsynonymous). Thus, under this hypothesis, K_{s1} and K_{a1} (see materials and methods) should be positively correlated (Wolfe and Sharp 1993) because adjacent mutations in the third position of a codon and the first position of the next codon, will usually be a synonymous and nonsynonymous change, respectively. Contradicting this prediction, there is no evidence of a positive correlation in the 35 genes between K_{s1} and K_{a1} (Table 1) for those codons with no or only one substitution ($r = 0.159$, $P = 0.34$). This lack of evidence for a positive correlation

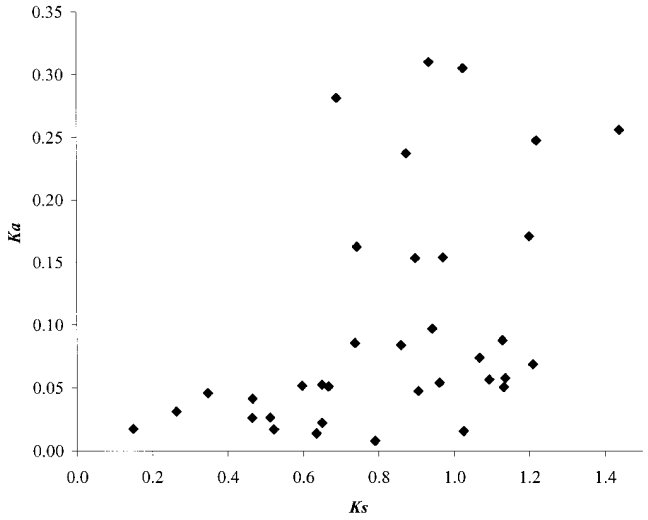

Figure 1.—Relationship between K_s and K_a for the compared 35 genes between *D. melanogaster* and *D. subobscura* or *D. pseudoobscura*.

between K_{s1} and K_{a1} also rules out correlated selection pressures for the whole gene on both kinds of substitutions (hypothesis iii). The observed correlation between K_s and K_a must be due to a relative excess of two substitutions (one synonymous and one nonsynonymous) in the *same* codon. To further test this excess, we have studied the presence of synonymous and nonsynonymous substitutions in codons in the 22 genes where the comparison of the two obscura species (*D. subobscura* and *D. pseudoobscura*) is possible. A *G*-test of independence for the presence of one synonymous and one nonsynonymous substitution in the same codon reveals a strong excess of doubly substituted codons ($G = 22.95$, $P < 0.0001$; Table 2). A further analysis at the codon position of the synonymous and nonsynonymous substitutions in doubly substituted codons allowed us to compare the observed frequencies of substituted positions (first-second, first-third, and second-third) with the expected frequencies on the basis of the frequency of synonymous substitutions at the different codon positions assuming a random coding sequence. A test of goodness of fit (Table 3) fails in detecting an excess of codons substituted at adjacent positions ($G = 0.793$, $P = 0.673$, with the Williams correction for continuity). This result confirms that doublet and adjacent mutations are not the cause of the excess of codons with both one synonymous and one nonsynonymous substitution.

Hypotheses i and iiia, adaptive synonymous substitutions and covarying constraints on particular codons, make distinguishable predictions about the independent occurrence of double mutations in the same codon. Under the covarying constraints hypothesis, independent occurrences of synonymous and replacement substitutions are expected to accumulate in more weakly constrained codons. In contrast, the adaptive substitution hypothesis predicts nonindependence of substitutions. To distinguish between these two predictions, we have carried out a phylogenetic analysis of substitutions in those doubly substituted codons. If synonymous and nonsynonymous substitutions in the same codon are independent events, then they will be expected to have occurred at equal frequency as single mutations in each lineage leading to *D. pseudoobscura* and *D. subobscura* and as double mutations in the same species lineage. The adaptive hypothesis, otherwise, predicts an excess occurrence of the two substitutions in the same species lineage.

We first identified all codons with one synonymous and one nonsynonymous substitution when comparing *D. subobscura* and *D. pseudoobscura* sequences. We then used *D. melanogaster* sequences for the 22 genes whose sequences are available in the three species (Zeng *et al.* 1998; see Table 4) to identify the ancestral sequence for each of these codons. Of the 35 codons with one synonymous and one nonsynonymous substitution between *D. subobscura* and *D. pseudoobscura*, and for which we could unambiguously assign each substitution to one branch, a statistically significant number of them, 27, had both substitutions in one obscura lineage ($\chi^2 = 10.31$, $P = 0.0013$). Also, 48.1% of these codons show nonadjacent substitutions (first and third position), which is very close to the expected 47.8% (see above). Synonymous and nonsynonymous substitutions in the same codon are not independent occurrences, and they

TABLE 2

G*-test of independence for the presence of synonymous and nonsynonymous substitutions in the same codon between *D. subobscura* and *D. pseudoobscura

	Codons with one synonymous substitution	Codons without synonymous substitution
Codons with one nonsynonymous substitution	91	186
Codons without nonsynonymous substitution	1361	5355
	$G = 22.95$, $P < 0.0001$[a]	

[a] $G = 18.637$, $P < 0.0001$ when codons with more than one synonymous or nonsynonymous substitutions are taken into account.

TABLE 3

Substituted positions in codons with one synonymous and one nonsynonymous substitution between *D. subobscura* and *D. pseudoobscura*

Codon positions of the substitutions	Observed	Expected ratio[a]
1,2	5	0.037
1,3	44	0.478
2,3	42	0.485
	$G_{\text{Williams}} = 0.793$, $P = 0.673$	

[a] Expected ratio based on the frequency of synonymous substitutions at the different codon positions in a random coding sequence: 0.046, 0.0, and 0.598, for the first, second, and third codon positions, respectively.

cannot be explained by doublet, and adjacent, mutations or by correlated selective constraints on particular codons. Only the adaptive substitution hypothesis is compatible with the data.

To investigate the possible selective mechanism driving the nonindependent occurrence of synonymous and nonsynonymous substitutions, we considered a further prediction of the adaptive substitution hypothesis. Under any equilibrium model of codon bias maintained by selection, the expected rates of fixation of preferred and unpreferred mutations must be equal, while mutation pressure favors the appearance of unpreferred codons. In contrast, under the adaptive substitution hypothesis for doubly substituted codons, preferred substitutions are expected to exceed unpreferred substitutions. This relative excess of preferred substitutions is expected because the fixation of a nonsynonymous substitution (either by selection or drift) that generates an unpreferred codon will lead to positive selection for a subsequent synonymous substitution to the new preferred codon for that amino acid. The selection mechanism underlying adaptive synonymous substitution is not critical to this argument: selection may be to enhance either translational efficiency or accuracy. The important point is that the consequence of this selection will be to increase the substitution rate of preferred substitutions compared to unpreferred substitutions in the codons that have both synonymous and nonsynonymous substitutions occurring in the same lineage.

The data do not support this prediction. Under the assumption that a nonsynonymous substitution preceded a synonymous substitution, 65% of the inferred ancestral codons are preferred (for the 20 out of the 27 codons that conform to this assumption), while this percentage becomes 40 and 30% for the intermediate and doubly substituted codons, respectively. Thus, there is no evidence of selection for either a preferred synonymous substitution following a nonsynonymous substitution, or equilibrium between preferred and unpreferred codons. A *G*-test of independence among the number of preferred and unpreferred codons of the ancestral and doubly substituted codons reveals a smaller ratio

TABLE 4

Number of codons with one synonymous and one nonsynonymous substitution between *D. subobscura* and *D. pseudoobscura*

	Number of codons with synonymous and nonsynonymous substitution		
Gene[a]	Same obscura lineage	Different obscura lineage	No. codons[b]
A/A-T/sesB	1	1	288
Adh	0	1	254
Adhr	1	0	272
Aprt	1	0	181
Bcd	1	0	93
Cp15	2	0	108
Gad1	1	0	369
Gld	1	2	612
Rh1/ninaE	1	0	370
Sod	1	0	114
Sry-α	5	2	514
Uro	3	1	334
Xdh/ry	9	1	1333
Total	27	8	7005[c]
	$\chi^2 = 10.31$, $P = 0.0013$		

Number of codons where a parsimony approach (see text) allows us to identify the synonymous and nonsynonymous substitutions in the same codon.

[a] Only those genes where one or more codons can be assigned in the previous categories are shown.

[b] The effective number of codons where the comparison among the three species is possible.

[c] The sum of the analyzed codons for the 22 genes where the comparison is possible.

of preferred to unpreferred codons in this latter group ($G = 4.837$, $P = 0.0279$ with the Williams correction). The expected frequency of preferred and unpreferred codons on the basis of random mutation can be estimated from the inferred ancestral codons and a mutational bias toward an A + T to G + C ratio of 60/40. The frequency of preferred codons consequence of both the first nonsynonymous substitution and the subsequent synonymous substitution in the same codon can be explained by random fixation of both the nonsynonymous ($\chi^2 = 0.695$, $P = 0.404$) and the synonymous ($\chi^2 = 1.377$, $P = 0.241$) mutations. By the same argument it is also possible to reject the converse of this adaptive hypothesis for doubly mutated codons, in which the initial occurrence of an unpreferred synonymous substitution established positive selection for an amino acid replacement that would be a more preferred codon (see below). From this analysis, and the preceding ones, we conclude that the data do not support any of the three hypotheses as explanations for the observed correlation between synonymous and replacement rates among genes. An additional alternative hypothesis to explain the data will be considered below.

DISCUSSION

Synonymous and replacement substitutions in the same codon: Our interest in studying codons differing by both a synonymous and an amino acid replacement substitution was motivated by the knowledge that synonymous and amino acid replacement substitution rates are positively correlated in Drosophila genes. The detection of an excess of codons with both a synonymous and a nonsynonymous substitution, not explained by adjacent mutations, confirms that natural selection is acting at the codon, post-transcriptional, level in Drosophila. Thus, codon selection can give rise to coupled synonymous and nonsynonymous substitutions in the same codon. This result does not, however, rule out selection also acting on synonymous sites at the level of mRNA to maintain functional secondary structures (Parsh *et al.* 1997; M. Antezana and M. Kreitman, unpublished data).

The analysis of doubly substituted codons between *D. psedoobscura* and *D. subobscura*, using *D. melanogaster* as an outgroup species, in a relatively large number of genes (22), allowed us to test several hypotheses. The discovery of a greater than expected number of these codons with both substitutions occurring in the same species lineage provides evidence for the nonindependence of replacement and synonymous substitutions in these codons. This nonindependence allowed us to reject the hypothesis of differential constraints among codons, predicted by the translational accuracy model.

The nonindependence of replacement and synonymous substitutions in doubly substituted codons is unlikely to be an artifact of the procedure used to assign changes to each lineage. The possible misassignment of a substitution to the wrong lineage, which might have occurred if the same mutational change had also occurred in the *D. melanogaster* lineage, or if a codon had multiple substitutions at the same nucleotide position, would be expected to obscure rather than enhance the nonrandom pattern of substitution we observed. This finding is also not likely to be due to unequal substitution rates in the *D. pseudoobscura* and the *D. subobscura* lineages, because these rates are very similar (Zeng *et al.* 1998) and because the doubly substituted codons were found to be present in both lineages at nearly equal frequencies (13 and 14). Nor can the nonindependence of doubly substituted codons be due to any systematic shift in codon usage in any of the three lineages. In a separate study, we found the codon usage in these genes to be remarkably similar for almost all of the codon families (Kreitman and Antezana 1998). Finally, the possibility of a single mutational event causing base changes at two adjacent positions can also be eliminated as the cause of this nonindependence of substitutions.

We therefore considered models of selection acting within the context of individual codons, *i.e.*, codon selection. One attractive possibility is the idea that selection drives synonymous substitutions to a new preferred codon following the substitution of an amino acid replacement substitution to an unpreferred codon. Alternatively, one might also consider the reverse process, an adaptive substitution of an amino acid replacement substitution to a new preferred codon following a synonymous substitution to an unpreferred codon. Unfortunately, the evidence does not support either of these adaptive explanations: The majority of synonymous (or replacement) substitutions in these doubly substituted codons were to unpreferred rather than preferred codons. Thus, the second substitution, whichever it is, is not likely to be the result of positive codon selection.

An *ad hoc* explanation for the observed patterns of change in doubly substituted codons can be invoked by assuming a lineage-specific relaxation of selective constraints on particular amino acids within a protein. This relaxed selection at the amino acid level will allow for an increase in the amino acid replacement substitution rate, as well as for an increase in the synonymous substitution rate. The latter is expected to occur because translational accuracy will also be relaxed in these codons. Under this scenario, the first substitution is likely to be a synonymous change to a less preferred codon. The inferred ancestral codons are 68% preferred (for the 25 codons that conform to this assumption) whereas the intermediate codons, after the synonymous substitution, are 24% preferred. The clear increment of unpreferred codons can be explained by mutation on the basis of the expected frequency of preferred and unpreferred codons (see previous section; $\chi^2 = 0.802$, $P = 0.37$). The increased occurrence of the second, nonsynonymous, substitution would not require the action of positive

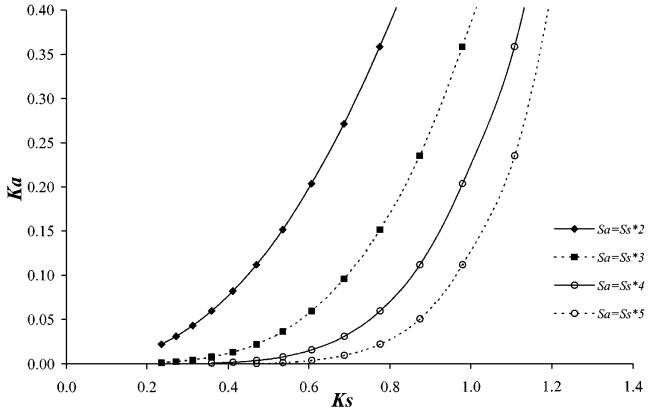

Figure 2.—Expected nonlinear relationship between K_s and K_a based on a linear relationship between selection coefficients on synonymous (s_s) and nonsynonymous (s_a) mutations assuming a number of neutral substitutions per site of 1.5.

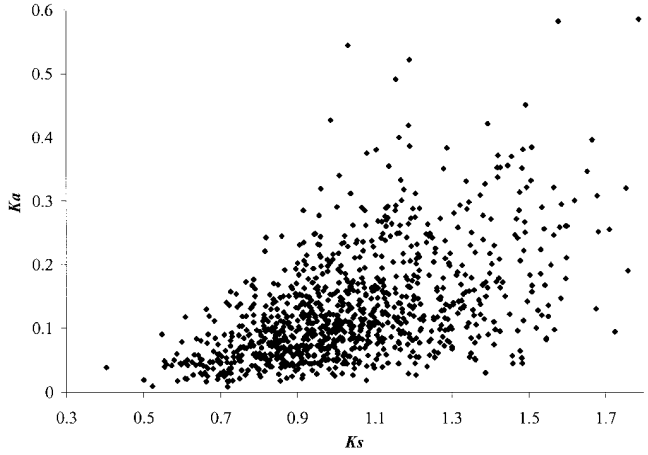

Figure 3.—Relationship between K_s and K_a obtained by computer simulation (see materials and methods) under the proposed covarion-like model assuming that selection coefficients on nonsynonymous sites (s_a) can explain 50% of the variance of the selection coefficients on synonymous sites (s_s); 1000 replicates, $f_c = 0.5$ (see text for details).

selection but only the relaxation of selective constraints on amino acid substitutions in this codon ($\chi^2 = 0.153$, $P = 0.696$, based on the observed number of preferred codons and the expected frequency due to mutation). Such an explanation differs from the already rejected differential constraints hypothesis only in the additional requirement that selective constraints on individual codons change over time within a species lineage. This scenario is reminiscent of one aspect of Fitch's covarion model of protein evolution (Fitch 1971; Miyamoto and Fitch 1995) in its proposal that the substitution of one amino acid in a protein, either by drift or selection, influences the functional constraints on other amino acid positions. (Our *ad hoc* explanation for doubly substituted codons, however, does not claim for a constant fraction of invariable positions.) Under our model, shifting relaxed constraints at the amino acid level is extended to include codon selection by including translational accuracy.

Our analysis indicates, therefore, that only a covarion-like model can explain the observed excess of doubly substituted codons on single-species lineages. Can the covarion-like model also account for the correlation between K_s and K_a observed in Drosophila data (Figure 1)? And if so, does it provide a more plausible explanation for this correlation than does a noncovarion model? To investigate these questions, we first explored the expected relationship between K_s and K_a when selection coefficients for synonymous (s_s) and replacement (s_a) mutations are completely correlated, and selection is constant for all sites (*i.e.*, a noncovarion substitution model). Following Kimura (1983),

$$K = K_n \times S/(1 - e^{-S}),$$

where K_n is the fixation probability of neutral mutations and $S = 4 N_e s$. Figure 2 shows that for linearly related selection coefficients for the two types of mutations, K_s and K_a are positively related, and that the relationship is a strongly nonlinear one. In fact, a similar trend can be observed in the data comparing synonymous and replacement changes between *D. melanogaster* and *D. subobscura* or *D. pseudoobscura* (Figure 1). We have also investigated by simulation the correlation between K_s and K_a for this same model, but when the selection coefficients

TABLE 5

Correlation coefficients between K_s and K_a obtained by computer simulation

Percentage of the variance of s_s explained by s_a[b]	Coefficient of correlation (r)[a]		
	Noncovarion model	Covarion-like model	
		$f_c = 0.25$	$f_c = 0.50$
0	−0.148	−0.046	0.200
25	−0.071	0.063	0.300
50	0.071	0.265	0.477
75	0.308	0.478	0.572

[a] Results obtained after 1000 replicates (see materials and methods for details).
[b] The selection coefficients on synonymous and nonsynonymous mutations, respectively.

on synonymous and replacement mutations are not completely correlated. As indicated in Table 5, a correlation between K_s and K_a in the range observed for Drosophila requires a very strong correlation between s_s and s_a, approaching a value of 1.

Table 5 also gives the results of simulations for a covarion-like model, and Figure 3 shows the data generated for one of the conditions examined. Under this model a fraction, f_c, of amino acid substitutions generates another amino acid substitution as well as a synonymous substitution. As shown in the table, the covarion-like model predicts moderately strong correlations between K_s and K_a for more permissive conditions of the relationship between s_s and s_a than the alternative model. This covarion-like model is therefore at least a viable explanation for the observed correlation between K_s and K_a seen in Drosophila data.

This analysis of the covarion-like model, together with the detected excess of codons with doubly substituted codons in the same species lineage (and with a tendency toward unpreferred codons), support a covarion hypothesis of shifting selective constraints over time on individual amino acids. We propose that selection at the level of translational accuracy couples synonymous and nonsynonymous substitutions in codons with relaxed constraints.

We are grateful to P. Andolfatto, M. Antezana, C. Bergman, A. Llopart, E. Stahl, C. Toomajian, and M. Przeworski for helpful discussions and/or comments on the manuscript. We also thank J. Hey and one anonymous reviewer for valuable comments that substantially improved the manuscript. J.M.C. is funded by a Postdoctoral Fellowship from Ministerio de Educación y Ciencia, Spain. This work was supported by a National Institutes of Health grant GM-39355 to M.K.

LITERATURE CITED

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

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

Aota, S., and T. Ikemura, 1986 Diversity in G+C content at the third positions of codons in vertebrate genes and its cause. Nucleic Acids Res. **14:** 6345–6355.

Benson, A. R., 1995 The molecular evolution of the obscura group *Chorion s15*: a prominent role for codon bias. Ph.D. Thesis, Harvard University.

Bernardi, G., and G. Bernardi, 1986 Compositional constraints and genome evolution. J. Mol. Evol. **24:** 1–11.

Bulmer, M., 1987 Coevolution of codon usage and tRNA abundance. Nature **325:** 728–730.

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

Chevallier, A., and J. P. Garel, 1979 Studies on tRNA adaptation, tRNA turnover, precursor tRNA and tRNA gene distribution in *Bombix mori* by using two-dimensional polyacrylmide gel electrophoresis. Biochemie **61:** 245–262.

Comeron, J. M., 1995 A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J. Mol. Evol. **41:** 1152–1159.

Comeron, J. M., and M. Aguadé, 1996 Synonymous substitutions in the *Xdh* gene of Drosophila: Heterogeneous distribution along the coding region. Genetics **144:** 1053–1062.

Filipski, J., 1987 Correlation between molecular clock ticking, codon usage, fidelity of DNA repair, chromosome banding and chromatin compactness in germline cells. FEBS Lett. **217:** 184–186.

Fitch, W. M., 1971 Rate of change of concomitantly variable codons. J. Mol. Evol. **1:** 84–96.

Garel, J. P., 1982 The silkworm, a model for molecular and cellular biologist. Trends Biochem. Sci. **7:** 105–108.

Graur, D., 1985 Amino acid composition and the evolutionary rates of protein-coding genes. J. Mol. Evol. **22:** 53–63.

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

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

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism* III, edited by H. N. Munro. Academic Press, New York.

Kimura, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

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., and M. Antezana, 1998 The population and evolutionary genetics of codon bias, in *Evolutionary Genetics from Molecules to Morphology*, edited by R. Sing and C. Krimbas. Columbia University Press, New York. (in press).

Lipman, D. J., and W. J. Wilbur, 1985 Interaction of silent and replacement changes in eukaryotic coding sequences. J. Mol. Evol. **21:** 161–167.

Miyamoto, M. M., and W. M. Fitch, 1995 Testing the covarion hypothesis of molecular evolution. Mol. Biol. Evol. **12:** 503–513.

Moriyama, E. N., and T. Gojobori, 1992 Rates of synonymous substitution and base composition of nuclear genes in Drosophila. Genetics **130:** 855–864.

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

Mouchiroud, D., C. Gautier and G. Bernardi, 1995 Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. **40:** 107–113.

Ohta, T., and Y. Ina, 1995 Variation in synonymous substitution rates among mammalian genes and the correlation between synonymous and nonsynonymous divergences. J. Mol. Evol. **41:** 717–720.

Parsh, J., S. Tanda and W. Stephan, 1997 Site-directed mutations reveal long-range compensatory interactions in the *Adh* gene of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **94:** 928–933.

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. Evol. **28:** 398–402.

Sharp, P. M., T. M. F. Tuohy and K. R. Mosurski, 1986 Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. **14:** 5125–5139.

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.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*, Ed. 3. W. H. Freeman and Co., New York.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Ticher, A., and D. Graur, 1989 Nucleic acid composition, codon usage, and rate of synonymous substitution in protein-coding genes. J. Mol. Evol. **28:** 286–298.

Wang, D., J. L. Marsh and F. J. Ayala, 1996 Evolutionary changes in the expression pattern of a developmentally essential gene in

three Drosophila species. Proc. Natl. Acad. Sci. USA **93:** 7103–7107.

Wells, R. S., 1996 Nucleotide variation at the *Gpdh* locus in the genus Drosophila. Genetics **143:** 375–384.

White, B. N., G. M. Tener, J. Holden and D. T. Suzuki, 1973 Analysis of tRNAs during the development of Drosophila. Dev. Biol. **33:** 185–195.

Wolfe, K. H., and P. M. Sharp, 1993 Mammalian gene evolution: Nucleotide sequence divergence between mouse and rat. J. Mol. Evol. **37:** 441–456.

Wolfe, K. H., P. M. Sharp and W.-H. Li, 1989 Mutation rates differ among regions of the mammalian genome. Nature **337:** 283–285.

Zeng, L.-W., J. M. Comeron, B. Chen and M. Kreitman, 1998 The molecular clock revisited: the rate of synonymous vs. replacement change in Drosophila. Genetica **102/103:** 369–382.

Communicating editor: J. Hey