Codon Bias Differentiates Between the Duplicated *Amylase* Loci Following Gene Duplication in Drosophila

Ze Zhang,*[,1] Nobuyuki Inomata,*[,2] Tomohiro Ohba,* Marie-Louise Cariou[†] and Tsuneyuki Yamazaki*

**Laboratory of Molecular Population Genetics, Department of Biology, Graduate School of Sciences, Kyushu University, Fukuoka 812-8581, Japan and [†]Populations, Genetique et Evolution, Centre National de la Recherche Scientifique, 91198 Gif sur Yvette cedex, France*

Manuscript received July 26, 2001
Accepted for publication April 18, 2002

ABSTRACT

We examined the pattern of synonymous substitutions in the duplicated *Amylase* (*Amy*) genes (called the *Amy*1- and *Amy*3-type genes, respectively) in the *Drosophila montium* species subgroup. The GC content at the third synonymous codon sites of the *Amy*1-type genes was higher than that of the *Amy*3-type genes, while the GC content in the 5′-flanking region was the same in both genes. This suggests that the difference in the GC content at third synonymous sites between the duplicated genes is not due to the temporal or regional changes in mutation bias. We inferred the direction of synonymous substitutions along branches of a phylogeny. In most lineages, there were more synonymous substitutions from G/C (G or C) to A/T (A or T) than from A/T to G/C. However, in one lineage leading to the *Amy*1-type genes, which is immediately after gene duplication but before speciation of the *montium* species, synonymous substitutions from A/T to G/C were predominant. According to a simple model of synonymous DNA evolution in which major codons are selectively advantageous within each codon family, we estimated the selection intensity for specific lineages in a phylogeny on the basis of inferred patterns of synonymous substitutions. Our result suggested that the difference in GC content at synonymous sites between the two *Amy*-type genes was due to the change of selection intensity immediately after gene duplication but before speciation of the *montium* species.

ADAPTIVE evolution of amino acid substitutions caused by positive Darwinian selection is one of the most important mechanisms for the functional divergence between members of a multigene family (Hughes 2000). Therefore, previous studies have concentrated on detecting an excess of (amino acid) replacement substitutions by comparing the patterns of replacement and synonymous substitutions. This procedure has been used to infer evolutionary forces and provided evidence for adaptive amino acid evolution after gene duplication (Messier and Stewart 1997; Zhang *et al.* 1998; Hughes *et al.* 2000).

Increasingly more findings suggest that most synonymous changes in unicellular organisms and Drosophila are not neutral. Indeed, synonymous codon usage bias is ubiquitous in *Escherichia coli*, *Saccharomyces cerevisiae*, and Drosophila. In these organisms, codon usage is biased toward a subset of major codons (G- or C-ending codons), which generally code for the most abundant tRNA(s) (Ikemura 1981, 1982; Bennetzen and Hall 1982; Grosjean and Fiers 1982; Shields *et al.* 1988; Moriyama and Powell 1997). In addition, the positive relationship between expression levels and codon bias was observed. That is, highly expressed genes show greater codon bias than genes with limited or low expression (Gouy and Gautier 1982; Shields *et al.* 1988). Furthermore, the efficacy of natural selection on codon usage is a function of recombination rate (Kliman and Hey 1993; Comeron *et al.* 1999; Takano 1999). In Drosophila, on the basis of patterns of polymorphism and divergence at synonymous sites, it has been found that synonymous substitutions were subject to weak selection against major and nonmajor codons (Akashi 1995). All of these studies have suggested the action of natural selection on synonymous (silent) sites in Drosophila. However, the possible role of synonymous substitutions following gene duplication has been scarcely evaluated. In particular, very few cases of weak selection causing divergence at the synonymous sites between members of multigenes have been documented so far.

The *Amy* genes of Drosophila encoding α-amylase proteins, which break starch into maltose and glucose and interact directly with food environments, constitute a relatively small multigene family with two to seven copies (Bahn 1967; Brown *et al.* 1990; Da Lage *et al.* 1992; Shibata and Yamazaki 1995; Popadic *et al.* 1996; Steinemann and Steinemann 1999; Inomata and Yamazaki 2000). The organization and molecular evolu-

[1]*Present address:* Laboratory of Digital Agriculture and Bioinformatics, Southwest Agricultural University, Chongqing 400716, People's Republic of China.

[2]*Corresponding author:* Laboratory of Molecular Population Genetics, Department of Biology, Graduate School of Sciences, Kyushu University, Fukuoka 812-8581, Japan.
E-mail: ninomscb@mbox.nc.kyushu-u.ac.jp

tion of the *Amy* multigene family in the *melanogaster* species subgroup and several other species have been well characterized. Previous studies showed that the members of the *Amy* multigene family have evolved in a concerted manner (Hickey *et al.* 1991; Popadic and Anderson 1995; Shibata and Yamazaki 1995; Inomata and Yamazaki 2000). On the other hand, *Drosophila kikkawai* and its sibling species were found to have two types of very diverged *Amy* genes (the *Amy*1- and *Amy*3-type genes) encoding active amylase isozymes (Inomata and Yamazaki 2000). Their expression patterns in different food environments diverged after gene duplication but before speciation. The *Amy*1-type genes have higher GC content at the third position of codons and more biased codon usage than do the *Amy*3-type genes. These results suggest the presence of some relationship between regulatory and synonymous evolution after gene duplication.

To elucidate what evolutionary forces have acted on the *Amy*1- and *Amy*3-type genes, we sequenced the full length of both genes of the *montium* species. Here, we describe evolutionary patterns of the two *Amy*-type genes and propose that the divergence at the synonymous sites between them is due to the change of selection intensity immediately after gene duplication but before speciation of the *montium* species.

MATERIALS AND METHODS

DNA sequences: Genomic DNA libraries of *D. nagarholensis* (strain name: PGE in Centre National de la Recherche Scientifique), *D. punjabiensis* (strain name: 14028-0531.0 in Bowling Green State University), and *D. watanabei* (strain name: SWB248 in Tokyo Metropolitan University) were constructed. It should be noted that on the basis of a morphological analysis the species of the stock number 14028-0531.0 at Bowling Green State University was regarded as *D. punjabiensis*, although it is described as *D. jambulina*. The *Amy*1, *Amy*2 (*Amy*1-type), and *Amy*3 (*Amy*3-type) genes were isolated from the genomic libraries by plaque hybridization using recombinant plasmids with the PCR product containing the partial *Amy*1- or *Amy*3-type gene from each species as probes. They were sequenced on both strands of DNA using ABI automated sequencer Model 377 and a DNA sequencing kit (BigDye terminator cycle sequencing ready reaction, ABI) with the synthetic oligonucleotide primers. The new sequences obtained in this study were deposited in the DNA Data Bank of Japan (DDBJ) and their accession numbers are AB078765–AB078773. All other *Amy* sequences of *D. kikkawai*, *D. bocki*, *D. leontia*, and *D. lini* (accession nos. AB035055–AB035069), which came from the genomic libraries (Inomata and Yamazaki 2000) were obtained from the DDBJ. The *Amy* sequences of *D. virilis* (accession no. U02029) and *Scaptodrosophila lebanonensis* (accession no. AB078774) were used as outgroups in the phylogenetic tree reconstruction and in the inference of patterns of synonymous substitutions. The *Amy* sequences of *D. pseudoobscura* (accession no. X76240) and *D. melanogaster* (accession no. L22730) were also included in the phylogenetic tree.

AMY protein electrophoresis: The samples for AMY protein electrophoresis were collected as follows. Adult flies of the three *montium* species were transferred to the two test foods, glucose medium [10% glucose (w/v), 5% killed yeast (w/v), 0.6% agar (w/v), and 0.4% propionic acid (v/v) in distilled water] and starch medium [10% soluble starch (w/v), 5% killed yeast (w/v), 0.6% agar (w/v), and 0.4% propionic acid (v/v) in distilled water]. They laid eggs for 3 days at 22°. After laying eggs, 10 adult flies were randomly collected without distinguishing sexes and frozen at −70°. Ten third instar larvae grown on glucose medium and an additional 10 third instar larvae grown on starch medium were also randomly collected without distinguishing sexes. Larvae were washed with distilled water and then stored at −70°.

The samples were homogenized by sonication in a buffer [pH 8.9; 0.1 m Tris-borate, 5 mm $MgCl_2$, and 10% sucrose (w/v)]. Before electrophoresis the protein content of each sample was measured by the BCA protein assay reagent (Pierce, Rockford, IL). Then, the samples with the equal protein content were applied to a polyacrylamide gel [5% acrylamide (w/v), 0.2% bis-acrylamide (w/v), 20 mm $CaCl_2$, and 0.1 m Tris-borate] in a 0.1 m Tris-borate (pH 8.9) buffer. After running for 3 hr at 4° and 300 V, the gel was incubated at 37° in starch solution [1% soluble starch (w/v), 0.1 m Tris-HCl (pH 7.4), and 20 mm $CaCl_2$] for 1 hr. The gels were then washed with water and stained in I_2-KI solution. The band mobility was referred to as AMY1 and AMY3 isozymes in *D. melanogaster* (Inomata *et al.* 1995).

Data analysis: Alignment of DNA sequences was performed using the CLUSTAL W program (Thomson *et al.* 1994). Gap alignment in the 5′-flanking regions was corrected by hand. Codon usage bias (effective number of codons [ENC]; Wright 1990) and GC content at synonymous third codon position was computed using the DnaSP program, version 3.50 (Rozas and Rozas 1999). A neighbor-joining (NJ) tree from the 1000-bootstrap resampling with Jukes and Cantor's (1969) distance was produced by using the CLUSTAL W. The PAML program, version 3.0 (Yang 2000), employing the maximum likelihood (ML) method, and the PAUP program, version 4.0 (Swofford 1998), employing the maximum parsimony (MP) method, were also used to construct phylogenetic trees.

The likelihood ratio test was used to test two evolutionary models. The null hypothesis was that there would be no lineage-specific effects of evolutionary rate with constant *dN*/*dS* ratio throughout lineages, and the alternative hypothesis was that there would be an independent *dN*/*dS* ratio for every lineage. We incorporated transition/transversion bias and biased codon frequencies into the models. Estimation of *dN*/*dS* ratio was performed by the ML method using the PAML program. According to the best model obtained by the likelihood ratio test, the ancestral sequences at the nodes were estimated by the ML method using the PAML program. Then their GC content and the number of nucleotide substitutions at synonymous third position along branches were counted.

RESULTS

Phylogenetic tree of the *Amy* genes: In *D. kikkawai* and its sibling species, there were three or four paralogous genes. On the basis of the restriction maps and subsequent sequencing, the *Amy*1 gene was distinguished from the *Amy*2, although they were similar to each other, while the sequences of the *Amy*3 and *Amy*4 genes were identical. Therefore, the numbering of *Amy*3 and *Amy*4 was arbitrary (Inomata and Yamazaki 2000). The *Amy*1 and *Amy*2 genes and the *Amy*3 and *Amy*4 genes are called the *Amy*1-type gene and the *Amy*3-type gene, respectively (Inomata and Yamazaki 2000). The *Amy* genes cloned

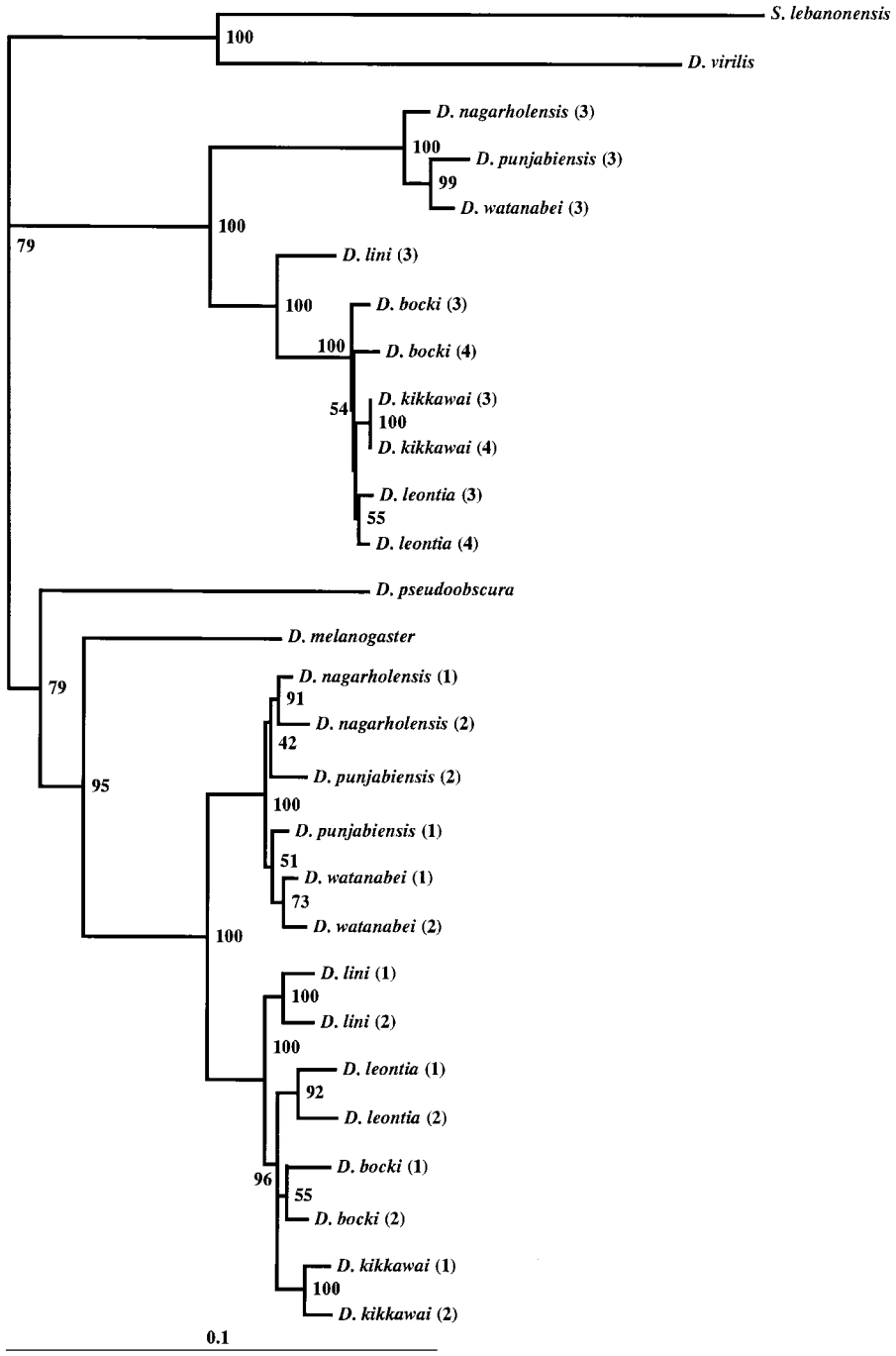

FIGURE 1.—A neighbor-joining tree for the two *Amy*-type genes of the seven *montium* species. Bootstrap value from 1000 replications is shown along each branch. The *Amy* sequences of *D. virilis* and *S. lebanonensis* were used as an outgroup. The *Amylase* gene locus is indicated in parentheses.

from genomic libraries of the three *montium* species, *D. nagarholensis*, *D. punjabiensis*, and *D. watanabei*, could be assigned to each gene type on the basis of their flanking sequences. Figure 1 shows an NJ tree constructed using coding regions of the *Amy*1- and *Amy*3-type genes in *D. bocki*, *D. kikkawai*, *D. leontia*, *D. lini*, *D. nagarholensis*, *D. punjabiensis*, and *D. watanabei*. The *Amy*3-type genes were outside of the *Amy* genes of *D. melanogaster* and *D. pseudoobscura*, but the bootstrap value for the *Amy* gene of *D. pseudoobscura* was not high (79%). The ML method supported the branching pattern of the NJ tree. However, the *Amy* gene of *D. pseudoobscura* was outside of the two *Amy* gene types when the MP method was used (data not shown). Furthermore, although the location of the *Amy* gene of *D. melanogaster* in the MP tree was the same as that in the NJ tree, its bootstrap value was not high (75%). Therefore, the placement of the *Amy* genes of *D. melanogaster* and *D. pseudoobscura* is not clear. In both types of *Amy* genes the branching pattern in the *kikkawai* complex (*D. bocki*, *D. kikkawai*, *D. leontia*, and *D. lini*) was consistent with the previous report (INOMATA and YAMAZAKI 2000) and those four species clustered together with *D. nagarholensis*, *D. punjabiensis*, and *D. watanabei*. For the *Amy*3-type gene *D. punjabiensis* clus-

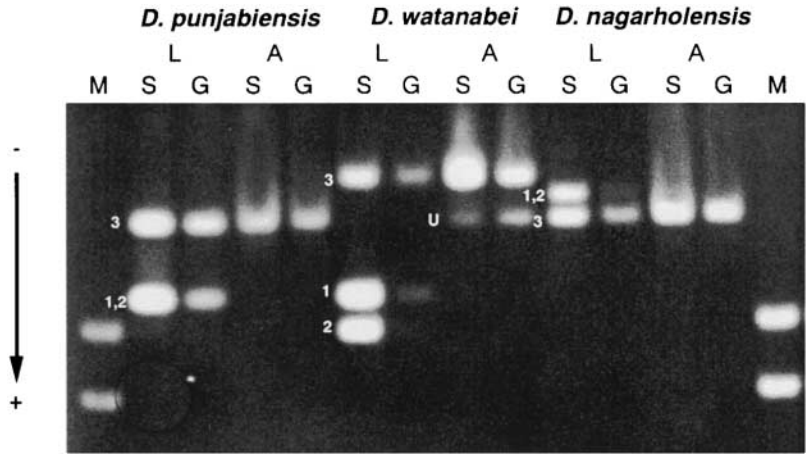

FIGURE 2.—Electrophoretic pattern of AMY isozymes at two stages, larval (L) and adult (A), on two test media in *D. punjabiensis*, *D. watanabei*, and *D. nagarholensis*. G and S indicate glucose and starch media. White numbers indicate AMY isozymes encoded by the *Amy*1, *Amy*2, and *Amy*3 genes. U indicates an unassigned isozyme. The AMY1 and AMY3 isozymes of *D. melanogaster* were used as mobility markers (M).

tered with *D. watanabei* and then with *D. nagarholensis*. On the other hand, for the *Amy*1-type gene the branching pattern of those three species was not clear (see Figure 1). Therefore, for further analyses, we used, for simplicity, the *Amy*1 and *Amy*3 genes as representatives of each gene type, and to exclude the uncertainty of the topology we chose the four *montium* species (*D. kikkawai*, *D. lini*, *D. nagarholensis*, and *D. watanabei*) and two outgroup species (*D. virilis* and *S. lebanonensis*). Their topology is shown in Figure 4. However, including other species (*D. punjabiensis*, *D. pseudoobscura*, and *D. melanogaster*), as in Figure 1, did not change our results fundamentally (data not shown).

Electrophoretic pattern of AMY isozymes: Figure 2 shows electrophoretic pattern of AMY isozymes on two media (glucose and starch) at two stages (larval and adult). In Drosophila the mobility of AMY isozymes is determined mostly by the charge differences of putative mature proteins (INOMATA *et al.* 1995; MATSUO *et al.* 1999). Therefore, we scored −1, 0, or +1 for each amino acid with negative, neutral, or positive charges, respectively. On the basis of the charge differences, we inferred which gene copy encodes an AMY isozyme. The net charges of the *Amy*1, *Amy*2, and *Amy*3 genes were, respectively, −6, −6, and −7 in *D. nagarholensis*; −9, −9, and −7 in *D. punjabiensis*; and −9, −10, and −6 in *D. watanabei*. However, there was an exception for the relationship between the mobility and the charge. Although the *Amy*3 gene in *D. watanabei* has the same charge as the *Amy*1 and *Amy*2 genes in *D. nagarholensis*, we regarded the slowest isozyme in *D. watanabei* as the product of the *Amy*3 gene and the slowest isozyme in *D. nagarholensis* as the product of the *Amy*1 and *Amy*2 genes. This exception might be due to the difference in the three-dimensional structure of the AMY proteins. In *D. watanabei* an additional isozyme was observed. It could be encoded by the fourth *Amy* copy or by the allele of *Amy*1, *Amy*2, or *Amy*3 genes. Hereafter, we call the AMY isozymes encoded by the *Amy*1-type gene and *Amy*3-type gene the AMY1 isozymes and AMY3 isozymes, respectively.

As previously reported in *D. bocki*, *D. kikkawai*, *D. leontia*, and *D. lini* (INOMATA and YAMAZAKI 2000), the AMY3 isozymes of the three species analyzed in this study, *D. nagarholensis*, *D. punjabiensis*, and *D. watanabei*, were observed at both larval and adult stages, and their activities were lower on glucose medium than on starch medium. In contrast, the AMY1 isozymes were observed only in larvae and their activities were higher on starch medium. This indicates that expression of the *Amy*1-type gene is more regulated than that of the *Amy*3-type gene at the transcriptional level, since amylase activity is mostly determined by the amount of mRNA (BENKEL and HICKEY 1986; YAMATE and YAMAZAKI 1999).

Base composition of the two *Amy*-type genes: Codon bias and GC content are summarized in Table 1. Figure 3 shows average GC content in the two *Amy*-type genes. As demonstrated in previous studies (INOMATA and YAMAZAKI 2000), the *Amy*1-type genes have higher GC content at synonymous third codon positions. In addition, GC content at fourfold degenerate sites is also higher in the *Amy*1-type genes than in the *Amy*3-type genes. On the other hand, there were few differences in GC content at the first and second codon positions between the two *Amy*-type genes (Figure 3). Therefore, the differences in base composition between them can be attributed to base composition at synonymous sites.

In the *Amy*1- and *Amy*3-type genes the average GC content at synonymous third codon position was 88.7 and 69.8%, respectively, and the average codon usage bias measured by ENC (WRIGHT 1990) was 29.9 and 41.8, respectively (see Table 1). On the other hand, the average GC content of the intron of the *Amy*1- and *Amy*3-type genes was 42.8 and 35.2%, respectively, and that of the 5′-flanking region was 46.1 and 45.0%, respectively (see Table 1). The difference in GC content between the two *Amy*-type genes in the noncoding regions is relatively small compared with that in the coding regions, suggesting the similar mutational bias between the two *Amy*-type genes in the noncoding regions.

Patterns of synonymous substitutions: For further examination of the difference in base composition, we

TABLE 1

GC content and codon bias in the duplicated *Amylase* genes

	GC3s[a] (%)	Intron[b]	5′-flanking[c] (%)	ENC
*Amy*1 type[d]				
D. bocki (1)	87.1	47.6 (63)	45.0	30.5
D. kikkawai (1)	87.1	46.0 (63)	45.0	31.4
D. leontia (1)	86.5	47.6 (63)	46.0	30.4
D. lini (1)	87.9	44.4 (63)	47.0	30.3
D. nagarholensis (1)	90.1	34.9 (63)	46.0	28.8
D. punjabiensis (1)	90.9	41.8 (55)	45.5	28.7
D. watanabei (1)	91.5	41.8 (55)	46.8	28.2
D. bocki (2)	87.9	49.2 (59)	45.8	30.5
D. kikkawai (2)	87.5	46.0 (63)	44.5	31.8
D. leontia (2)	86.5	46.0 (63)	40.0	31.2
D. lini (2)	87.9	46.0 (63)	45.8	30.2
D. nagarholensis (2)	88.8	35.7 (56)	49.2 (264 bp)	29.6
D. punjabiensis (2)	90.7	33.9 (56)	49.2 (262 bp)	28.7
D. watanabei (2)	91.8	38.1 (63)	49.2 (254 bp)	28.4
Average (*Amy*1 type)	88.7	42.8 (60.6)	46.1	29.9
*Amy*3 type[d]				
D. bocki (3)	71.2	37.3 (67)	43.5	40.6
D. bocki (4)	69.9	38.8 (67)	44.8	42.3
D. kikkawai (3)	71.4	38.1 (63)	44.0	40.9
D. kikkawai (4)	71.4	38.1 (63)	44.0	40.9
D. leontia (3)	70.1	37.3 (67)	43.3	41.4
D. leontia (4)	71.0	37.3 (67)	43.8	41.1
D. lini (3)	71.4	38.8 (67)	46.0	42.2
D. nagarholensis (3)	68.3	42.4 (66)	47.5	42.2
D. punjabiensis (3)	66.4	40.9 (66)	46.8	44.1
D. watanabei (3)	67.2	40.9 (66)	46.0	42.7
Average (*Amy*3 type)	69.8	39.0 (65.9)	45.0	41.8
Non-*montium* species				
S. lebanonensis	66.7	37.5 (56)	40.5	42.8
D. virilis	70.0	49.2 (59)	43.3	37.3
D. melanogaster	89.0	—	40.1 (322 bp)	28.7
D. pseudoobscura	87.9	62.0 (71)	51.4 (177 bp)	28.8

[a] The GC content at synonymous third position.
[b] The length of intron is indicated in parentheses. The *Amylase* gene of *D. melanogaster* has no intron.
[c] The GC content in the 400 bp of 5′-flanking nucleotide sequences is listed. When <400 bp are available, the length of nucleotide is shown in parentheses.
[d] The *Amylase* gene locus is indicated in parentheses.

estimated the ancestral sequences at each node by the maximum likelihood method. Then we computed their GC content at synonymous third codon positions and divided synonymous substitutions into G/C (G or C) → A/T (A or T) and A/T → G/C substitution along each branch. Before estimation, we tested the constancy of the lineage-specific *dN*/*dS* ratio on the topology shown in Figure 4 using the likelihood ratio test. Twice the difference in log-likelihood scores between the null hypothesis (constant ratio) and the alternative (lineage-specific ratio) was 49.30 and then the ratio constancy was rejected (d.f. = 17, $P < 0.005$). Therefore, we employed the model with the lineage-specific ratio for estimation of the ancestral sequences. The estimated number of synonymous and replacement substitutions and direction of synonymous substitutions at third codon positions for each branch are summarized in Table 2. The total number of synonymous substitutions along branches leading to the *Amy*1- and *Amy*3-type genes was 60 and 85, respectively. The estimates of *dS* along branches leading to the *Amy*1- and *Amy*3-type genes were 0.5816 and 0.6348, respectively, and then the total number of synonymous substitutions estimated using these values reached ~140 and 150, respectively. Therefore, the total number of synonymous substitutions was underestimated. This is because no multiple-hit correction was made. However, for our analysis direction of substitutions or substitutional bias, rather than their total number, is important. After *Amy*1/*Amy*3 duplication but before *montium* speciation, there was a highly significant difference in the G/C (G or C) ↔ A/T (A or T) substitution pattern between the *Amy*1- and *Amy*3-type genes

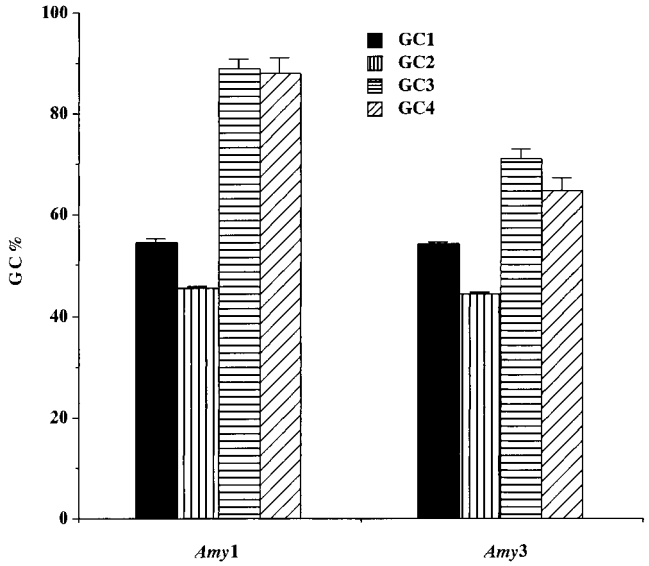

FIGURE 3.—Average GC content at different sites of codons for the two *Amy* (*Amy*1 and *Amy*3)-type genes in the *montium* species. GC1, GC2, GC3, and GC4 refer to GC content at the first, second, and third codon position and fourfold degenerate sites of codons, respectively.

(see Table 3, G with Williams' correction = 59.16, d.f. = 1, $P \gg 0.01$). On the other hand, after *montium* speciation there was no difference in the G/C ↔ A/T substitution pattern between the two *Amy*-type genes (see Table 4, G with Williams' correction = 0.72, d.f. = 1, $P > 0.5$), although substitutions from G/C to A/T were predominant in both *Amy*-type genes (G with Williams' correction = 7.24, d.f. = 1, $P < 0.01$ for *Amy*1-type gene; G with Williams' correction = 5.85, d.f. = 1, $P < 0.05$ for *Amy*3-type genes). These observations indicate that the direction of synonymous substitutions has changed between the two *Amy*-type genes immediately after gene duplication but before speciation of the *montium* species. That is, an excess of synonymous substitutions from A/T to G/C has occurred only in the lineage leading to the *Amy*1-type genes, whereas the G/C to A/T substitutions have been generally predominant throughout all other lineages (see Table 2).

Divergent evolution at synonymous sites: To infer the possible causes for synonymous changes, we consider the simplest model for major codon preference. Assuming the mutation rate is constant for the two *Amy*-type genes, it is very likely that changes in the pattern of synonymous substitutions are due to the fluctuation of selective constraint after gene duplication. To investigate the dynamics of the fluctuation of selective constraint, consider a population of N diploid individuals at mutation-selection-drift equilibrium, assuming that the internal nodes 1, 3, and 6 in Figure 3 are at statistical equilibrium. For simplicity, assume two states, major and nonmajor codon, and that the actual population size is equal to the effective size (N). Here, G- or C-ending codons and A- or T-ending codons are defined as major and nonmajor codons, respectively, and their frequencies are equal to GC content at synonymous third positions. This assumption could be reasonable, since codon preference pattern is very similar among Drosophila species examined (AKASHI 1994, 1995; AKASHI and SCHAEFFER 1997). And let s be the selective advantage of major codons over nonmajor codons under semidominance. The genetic model is

$$1 + s \qquad \overset{u}{\rightarrow} \qquad 1$$

G- or C-ending codons (major) $\overset{\leftarrow}{v}$ A- or T-ending codons (nonmajor),

(LI 1987; BULMER 1991), where u is the mutation rate from a major codon to a nonmajor codon and v is the reverse mutation rate. The number of synonymous substitutions from G/C to A/T, k_{AT}, from the ancestral node (*e.g.*, node 1 in Figure 4) to the second node (*e.g.*, node 3 in Figure 4) is given by

$$k_{AT} = 2Nuq\frac{-S}{2N(1 - e^{S})}t, \tag{1}$$

where q is the frequency of major codons in the ancestral node (*e.g.*, node 1 in Figure 4), $S = 4Ns$, $-S/(2N(1 - e^{S}))$ is the ultimate fixation probability of nonmajor codons whose initial frequency is $\frac{1}{2}N$, and t is the number of generations from the ancestral node to the second node. Similarly, the number of synonymous substitutions from A/T to G/C, k_{GC}, from the ancestral node (*e.g.*, node 1 in Figure 4) to the second node (*e.g.*, node 3 in Figure 4) is given by

$$k_{GC} = 2Nv(1 - q)\frac{S}{2N(1 - e^{-S})}t, \tag{2}$$

where $(1 - q)$ is the frequency of nonmajor codons in the ancestral node, and $S/(2N(1 - e^{-S}))$ is the ultimate fixation probability of major codons whose initial frequency is $\frac{1}{2}N$. On the basis of Equations 1 and 2, N and t should be canceled out by taking the ratio k_{AT}/k_{GC}, in the case of the duplicated genes, and then we have

$$S = \ln\left(\frac{u}{v}\right) + \ln\left(\frac{q}{1 - q}\right) - \ln\left(\frac{k_{AT}}{k_{GC}}\right), \tag{3}$$

where S is a function of mutation bias (u/v). For a given lineage, q and k_{AT}/k_{GC} can be estimated by the ML method. Suppose the estimates of k_{AT}/k_{GC} for the *Amy*1- and *Amy*3-type gene lineages are $(k_{AT}/k_{GC})_1$ and $(k_{AT}/k_{GC})_3$, respectively. Here, estimates of q at node 1, $(k_{AT}/k_{GC})_1$, and $(k_{AT}/k_{GC})_3$, were 0.846, 0.270, and 8.857, respectively. Note that, on the basis of Equation 3, the difference in selection intensities at the same u/v between the two lineages is

$$S_1 - S_3 = \ln\left(\frac{k_{AT}}{k_{GC}}\right)_1 - \ln\left(\frac{k_{AT}}{k_{GC}}\right)_3 = \text{constant}. \tag{4}$$

We estimated the changes of selection intensity ($S =$

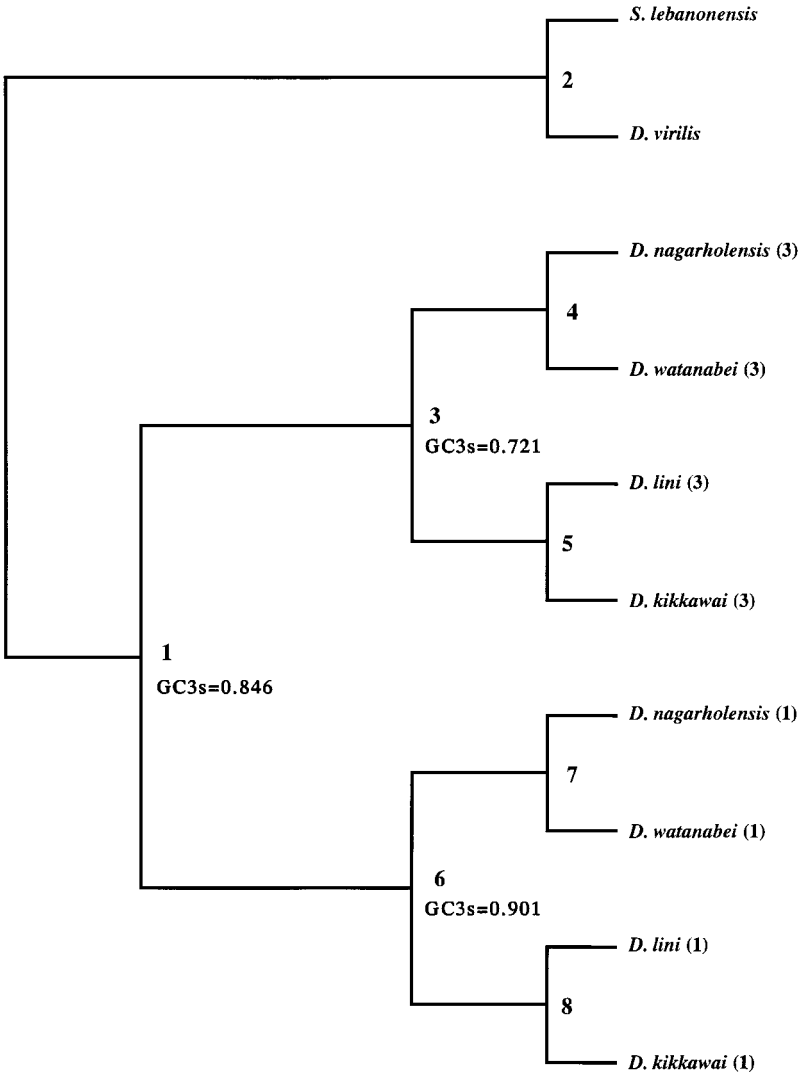

FIGURE 4.—*Amylase* gene tree used for estimation of the pattern of substitutions. The numbers in parentheses represent the *Amy* gene type. GC3s indicates GC content at synonymous third codon position.

$4Ns$) with the increase of mutation bias (u/v) for the two lineages, nodes 1–3 and nodes 1–6, respectively. Figure 5 shows that selection intensity (Ns) of the *Amy*1-type gene lineage is always larger than that of the *Amy*3-type gene lineage under the same u/v and that their difference is ~1. Furthermore, selection intensity of the *Amy*1-type gene lineage was $Ns > \frac{1}{2}$ in any u/v, suggesting that the major codons have been preferred.

DISCUSSION

We could not estimate the direction of substitutions between the two *Amy*-type genes in the noncoding regions by the ML method because the two regions were too diverged and could not be aligned. Therefore, we cannot directly infer what changes have occurred in the noncoding regions after *Amy* gene duplication before speciation of the *montium* species. However, differences in base composition between the coding and noncoding regions of the two *Amy*-type genes are striking, especially those between synonymous sites and the 5′-flanking region. This result indicated that an excess of synonymous substitutions from AT to GC has occurred in the *Amy*1 gene lineage after gene duplication but before speciation of the *montium* species and that this has resulted in higher GC content in the *Amy*1-type genes. One of the plausible explanations for the excess of synonymous substitutions from AT to GC is the temporal or regional changes in mutation bias. If the excess of synonymous substitutions is due to the changes in mutation bias after gene duplication but before speciation of the *montium* species, GC content in the noncoding regions should differ between the two *Amy*-type genes. However, the GC content in the 5′-flanking region was the same in both genes (Table 1). Although GC content of the intron was higher in the *Amy*1-type genes than in the *Amy*3-type genes, the difference was smaller than that in GC content at synonymous sites (27.1% increase in *Amy*1 synonymous sites, while 9.7% increase in the *Amy*1 intron). This observation is not likely to support the temporal or regional changes in mutation bias. Alternatively, it suggests that a small difference in selection

TABLE 2

The patterns of nucleotide substitutions along phylogenetic tree branches in the *Amylase* coding sequence

Branch[a]	A/T → G/C[b]	G/C → A/T[c]	Others	Sum	*dN*	*dS*	*dN*/*dS*
1 ↔ 2	NS	NS	NS	NS	0.0348	0.6336	0.0549
2 → *S. lebanonensis*	21	67	14	102	0.0483	0.9280	0.0521
2 → *D. virilis*	12	44	6	62	0.0459	0.5722	0.0803
To *Amy*3 genes							
1 → 3 (Ancestal to *Amy*3)	7	62	16	85	0.0081	0.6348	0.0128
3 → 4	25	43	10	78	0.0042	0.4001	0.0105
4 → *D. nagarholensis* (3)	1	0	1	2	0.0015	0.0156	0.0935
4 → *D. watanabei* (3)	6	11	1	18	0.0046	0.0653	0.0711
3 → 5	0	0	0	0	0.0037	0.0191	0.1943
5 → *D. lini* (3)	7	10	3	20	0.0016	0.0894	0.0178
5 → *D. kikkawai* (3)	6	7	4	17	0.0069	0.0863	0.0802
*Amy*3 lineage after speciation[d]	45	71	19	135	0.0225	0.6758	0.0333
To *Amy*1 genes							
1 → 6 (Ancestal to *Amy*1)	37	10	13	60	0.0107	0.5816	0.0185
6 → 7	12	6	10	28	0.0029	0.1414	0.0204
7 → *D. nagarholensis* (1)	1	7	1	9	0.0007	0.0440	0.0166
7 → *D. watanabei* (1)	3	1	2	6	0.0030	0.0352	0.0845
6 → 8	0	0	0	0	0.0039	0.0292	0.1324
8 → *D. lini* (1)	1	11	3	15	0.0001	0.0726	0.0010
8 → *D. kikkawai* (1)	1	13	2	16	0.0045	0.0945	0.0481
*Amy*1 lineage after speciation[d]	18	38	18	74	0.0151	0.4169	0.0362

[a] Each branch is represented by an ancestral node to descendant one shown in Figure 4.
[b] The number of synonymous substitutions from A or T to G or C at synonymous third positions.
[c] The number of synonymous substitutions from G or C to A or T at synonymous third positions.
[d] The sum of the substitutions after speciation of the four *montium* species.

intensity has caused the synonymous divergence between the two *Amy*-type genes.

Here, consider the plausible selection intensity in the two *Amy* gene lineages. At equilibrium and $s = 0$ in the model described above, $q = v/(u + v)$ (Sueoka 1962); then mutation bias (the ratio of mutation rates), u/v, is equal to the ratio of AT to GC content. Assuming equilibrium and neutral evolution in the noncoding regions, mutation bias is ~1.2–1.6 in the two *Amy* gene regions. In *D. melanogaster*, mutation bias is suggested to be 1.5 (Akashi 1996); thus those values are consistent with the previous study. Therefore, roughly speaking, selection intensity, *Ns*, could be 1 and 0 in the *Amy*1 and *Amy*3 lineages, respectively (see Figure 5). That is, after gene duplication but before speciation, synonymous sites of the *Amy*1-type gene have been under weak selection where major codons have selective advantage over nonmajor codons (*e.g.*, Akashi 1995), while those of the *Amy*3-type gene have evolved in neutral fashion.

The difference in the selection intensity between the two *Amy*-type genes could result from an increase in the magnitude of selection in the *Amy*1-type gene lineage, a decrease in the *Amy*3-type gene lineage, or both, comparing to the ancestral state. At present, we cannot estimate the ancestral state and our present results give only weak support for the increase in the selection intensity in the *Amy*1-type gene lineage. On the other hand, since synonymous sites in Drosophila genes have been

TABLE 3

Substitution bias after gene duplication but before *montium* speciation

	*Amy*1	*Amy*3
A/T → G/C	37	7
G/C → A/T	10	62

The substitutions along the internal branches between nodes 1 and 3 and nodes 1 and 6, which indicate after gene duplication but before speciation of the four *montium* species. G_{Will} indicates *G* value with Williams' correction. $G_{Will} = 59.16$; $P \ll 0.001$.

TABLE 4

Substitution bias after *montium* speciation

	*Amy*1	*Amy*3
A/T → G/C	18	45
G/C → A/T	38	71

The sum of the substitutions along all branches after divergence at nodes 3 and 6, which indicate after speciation of the four *montium* species. G_{Will} indicates *G* value with Williams' correction. $G_{Will} = 0.72$; $P > 0.05$.

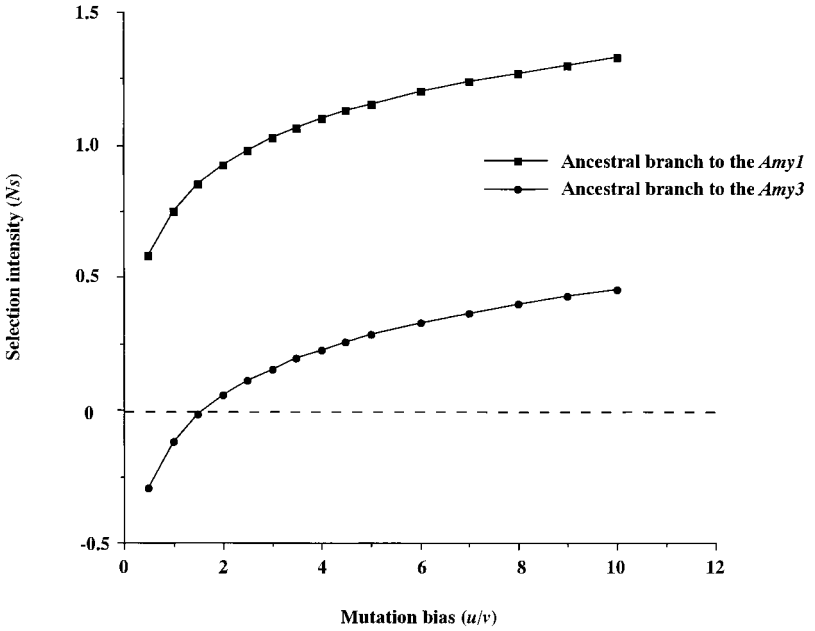

FIGURE 5.—Estimates of selection intensity (Ns) with different mutation biases (u/v) for the *Amy*1 and *Amy*3 lineages.

under weak selection (*e.g.*, AKASHI 1995), the weakened selection in the *Amy*3-type gene lineage is plausible.

The weakened selection intensity in the *Amy*3-type gene lineage is likely to be caused by the relaxation of selective constraint following gene duplication, the difference in recombination environment, or both. In *D. kikkawai* the *Amy*1-type genes reside in a chromosomal arm, suggesting a normal recombination rate, whereas the *Amy*3-type genes are located near the centromere, suggesting a low recombination rate (INOMATA and YAMAZAKI 2000). A lower recombination rate leads to reduction of selective efficacy (HILL and ROBERTSON 1966; BEGUN and AQUADRO 1992; COMERON *et al.* 1999; MCVEAN and CHARLESWORTH 2000). Therefore, assuming that the chromosomal locations of the two *Amy*-type genes in the ancestral *montium* species were the same, the difference in recombination environment might result in the different selection intensity between the *Amy*1- and *Amy*3-type genes. However, in the case of the duplicated genes the weakened selection intensity is likely to be more easily explained by a fluctuation of selection coefficient, s, itself because of functional redundancy following gene duplication.

As shown in the present and previous studies (INOMATA and YAMAZAKI 2000), the *montium* species shows more biased codon usage in the *Amy*1-type genes than in the *Amy*3-type genes. The degree to which codon usage is biased toward major codons is associated with gene expression levels (GOUY and GAUTIER 1982; GROSJEAN and FIERS 1982; SHIELDS *et al.* 1988). Amylase activity encoded by the *Amy*1-type genes strikingly differed in response to food environments and developmental stages, where activity level at the larval stage was highest on a starch (substrate) food environment. On the other hand, activity of amylase isozymes that are encoded by the *Amy*3-type genes was almost the same (see Figure 2; INOMATA and YAMAZAKI 2000). Since amylase activity is determined mostly by the amount of mRNA (BENKEL and HICKEY 1986; YAMATE and YAMAZAKI 1999), patterns of amylase activity reflect the mRNA expression profile. These quantitative differences of mRNAs between the two type genes could be due to the changes of the regulatory elements such as *cis*-sequences and/or *trans*-acting elements. Similar to the four species in the *kikkawai* complex (INOMATA and YAMAZAKI 2000), only some *cis*-regulatory sequences were found in the 5′-flanking region of the *Amy*3-type genes, whereas the *cis*-regulatory sequences of the *Amy*1-type genes were well conserved compared with other Drosophila *Amy* genes (data not shown). Together with those observations, the most plausible scenario is as follows: After gene duplication, the duplicated genes could have functional redundancy, resulting in the weakened selection. Therefore, one of the duplicated genes, the *Amy*3-type gene, lost the ancient function, probably in the *cis*-regulatory regions. This caused a lower level of expression, resulting in neutral evolution at synonymous sites. However, we still do not know why the *Amy*3-type genes have not lost their function completely. More studies are needed to address this interesting question.

We thank Drs. T. Ohta, H. Tachida, T. S. Takano, and A. E. Szmidt for fruitful discussion. This work was supported by research grants to N.I. and T.Y. and by a research fellowship to Z.Z. from the Ministry of Education, Science and Culture of Japan and by a research cooperative program (PICS 607) to M.-L.C. from the CNRS.

LITERATURE CITED

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

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

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

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

Bahn, E., 1967 Crossing over in the chromosomal region determining amylase isozymes in *Drosophila melanogaster*. Hereditas **58:** 1–12.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

Benkel, B. F., and D. A. Hickey, 1986 The interaction of genetic and environmental factors in the control of amylase gene expression in *Drosophila melanogaster*. Genetics **114:** 943–954.

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

Brown, C. J., C. F. Aquadro and W. W. Anderson, 1990 DNA sequence evolution of the amylase multigene family in *Drosophila pseudoobscura*. Genetics **126:** 131–138.

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

Comeron, J. M., M. Kreitman and M. Aguade, 1999 Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila. Genetics **151:** 239–249.

Da Lage, J.-L., F. Lemeunier, M.-L. Cariou and J. R. David, 1992 Multiple amylase genes in *Drosophila ananassae* and related species. Genet. Res. Comb. **59:** 85–92.

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

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

Hickey, D. A., L. Bally-Cuif, S. Abukashawa, V. Payant and B. F. Benkel, 1991 Concerted evolution of duplicated protein-coding genes in Drosophila. Proc. Natl. Acad. Sci. USA **88:** 1611–1615.

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

Hughes, A. L., 2000 *Adaptive Evolution of Genes and Genomes*. Oxford University Press, New York.

Hughes, A. L., J. A. Green, J. M. Garbayo and R. M. Roberts, 2000 Adaptive diversification within a large family of recently duplicated, placentally expressed genes. Proc. Natl. Acad. Sci. USA **97:** 3319–3323.

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.

Inomata, N., and T. Yamazaki, 2000 Evolution of nucleotide substitutions and gene regulation in the amylase multigenes in *Drosophila kikkawai* and its sibling species. Mol. Biol. Evol. **17:** 601–615.

Inomata, N., H. Shibata, E. Okuyama and T. Yamazaki, 1995 Evolutionary relationships and sequence variation of α-amylase variants encoded by duplicated genes in the *Amy* locus of *Drosophila melanogaster*. Genetics **141:** 237–244.

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, pp. 21–131 in *Mammalian Protein Metabolism*, edited by H. N. Munro. Academic Press, New York.

Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.

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

Matsuo, Y., N. Inomata and T. Yamazaki, 1999 Evolution of the amylase isozymes in *Drosophila melanogaster* species subgroup. Biochem. Genet. **37:** 289–300.

McVean, G. A. T., and B. Charlesworth, 2000 The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155:** 929–944.

Messier, W., and C.-B. Stewart, 1997 Episodic adaptive evolution of primate lysozymes. Nature **385:** 151–154.

Moriyama, E. N., and J. R. Powell, 1997 Codon usage bias and tRNA abundance in *Drosophila*. J. Mol. Evol. **45:** 514–523.

Popadic, A., and W. W. Anderson, 1995 Evidence for gene conversion in the amylase multigene family of *Drosophila pseudoobscura*. Mol. Biol. Evol. **12:** 564–572.

Popadic, A., R. A. Norman, W. W. Doane and W. W. Anderson, 1996 The evolutionary history of the amylase multigene family in *Drosophila pseudoobscura*. Mol. Biol. Evol. **13:** 883–888.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Shibata, H., and T. Yamazaki, 1995 Molecular evolution of the duplicated *Amy* locus in the *Drosophila melanogaster* species subgroup: concerted evolution only in the coding region and an excess of nonsynonymous substitutions in speciation. Genetics **141:** 223–236.

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.

Steinemann, S., and M. Steinemann, 1999 The amylase gene cluster on the evolving sex chromosomes of *Drosophila miranda*. Genetics **151:** 151–161.

Sueoka, N., 1962 On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA **48:** 582–592.

Swofford, D. L., 1998 PAUP* (phylogenetic analysis using parsimony *and other methods), version 4. Sinauer Associates, Sunderland, MA.

Takano, T. S., 1999 Local recombination and mutation effects on molecular evolution in Drosophila. Genetics **153:** 1285–1296.

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

Wright, F., 1990 The 'effective number of codons' used in a gene. Gene **87:** 23–29.

Yamate, N., and T. Yamazaki, 1999 Is the difference in α-amylase activity in the strains of *Drosophila melanogaster* with different allozymes due to transcriptional control? Biochem. Genet. **37:** 345–356.

Yang, Z., 2000 Phylogenetic analysis by maximum likelihood (PAML), version 3.0. University College, London.

Zhang, J., H. F. Rosenberg and M. Nei, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95:** 3708–3713.

Communicating editor: N. Takahata