Rate Variation of DNA Sequence Evolution in the Drosophila Lineages

Toshiyuki S. Takano

Department of Population Genetics, National Institute of Genetics, Mishima, Shizuoka-ken 411-8540, Japan

Manuscript received September 19, 1997
Accepted for publication February 4, 1998

ABSTRACT

Rate constancy of DNA sequence evolution was examined for three species of Drosophila, using two samples: the published sequences of eight genes from regions of the normal recombination rates and new data of the four *AS-C* (*ac*, *sc*, *l'sc* and *ase*) and *ci* genes. The *AS-C* and *ci* genes were chosen because these genes are located in the regions of very reduced recombination in *Drosophila melanogaster* and their locations remain unchanged throughout the entire lineages involved, yielding less effect of ancestral polymorphism in the study of rate constancy. The synonymous substitution pattern of the three lineages was found to be erratic in both samples. The dispersion index for replacement substitution was relatively high for the *per*, *G6pd* and *ac* genes. A significant heterogeneity was found in the number of synonymous substitutions in the three lineages between the two samples of genes with different recombination rates. This is partly due to a lack of the lineage effect in the *D. melanogaster* and *Drosophila simulans* lineages in the *AS-C* and *ci* genes in contrast to Akashi's observation of genes in regions of normal recombination. The higher codon bias in *Drosophila yakuba* as compared with *D. melanogaster* and *D. simulans* was observed in the four *AS-C* genes, which suggests change(s) in action of natural selection involved in codon usage on these genes. Fluctuating selection intensity may also be responsible for the observed locus-lineage interaction effects in synonymous substitution.

THE constancy of the evolutionary rate across multiple lineages was first examined by Ohta and Kimura (1971). Subsequently Langley and Fitch (1974) and Gillespie (1989) have distinguished locus-lineage interaction effects from lineage effects. Episodic protein sequence evolution has been taken as evidence of positive natural selection in action (Gillespie 1989), while lineage dependent evolution may arise due to generation-time and/or population-size effects (Ohta 1995; Akashi 1996). Formally, other biological causes, however, can be responsible for irregular molecular evolution, and this issue still remains to be solved.

The constancy of the rate has been examined using the variance-to-mean ratio of the number of substitutions among lineages (dispersion index) as a measure (Ohta and Kimura 1971; Gillespie 1989). This ratio should be one under the assumption of simple Poisson process and no estimation error. When the variance in the number of substitutions among lineages is greater than the expectation, several mechanisms can be put forward to explain the presence of locus-lineage interaction: changes in mutation rate at the individual gene basis, time-heterogeneous directional selection (Gillespie 1991), and fluctuating neutral space (Takahata 1987). Region-dependent changes in recombination rate can also result in changes in substitution rate of nearly neutral mutations in the regions involved through the background selection effects (Charlesworth *et al.* 1993), and thus may lead to the inflation of the dispersion index. This effect should be distinguished from generation-time and population-size effects, which affect all the genes in the genome in the same direction. The final goal of this study is to know how often the degree and pattern of natural selection varies in time and to evaluate the action of natural selection in molecular evolution on this basis. The test of rate constancy among closely related species is the first step to estimate the frequency of changes in selective forces acting on individual genes and the relative contribution of the factors as mentioned above. In this context, knowledge of evolutionary-rate variation is important to deepen our understanding of the mechanisms of molecular evolution.

Among several methods used to examine the action of natural selection, the rate constancy test also has a couple of advantages. First, an excess of replacement substitutions compared with the number of synonymous substitutions is not necessarily required to get a significant result, meaning a higher detection power in certain situations. Second, independent analysis for replacement and synonymous substitutions can be made. The drawback of the necessity of assumptions of absolute divergence time is overcome in this study by the use of multiple-gene comparison.

Many of the previous analyses on rate-constancy have paid less attention to ancestral polymorphism, although Gillespie and Langley (1979) and Hudson (1983) have done analytical and simulation studies of its effect on the rate-constancy test. On the other hand, Taka-

Author's e-mail: totakano@lab.nig.ac.jp

hata (1986) has developed means of estimating the effective population size in an ancestral population using the mean and variance of the number of substitutions between two extant species among genes under constant-rate assumption. However, there seems to be considerable rate variation among loci at the silent site (Sharp and Li 1989; Moriyama and Gojobori 1992; Wolfe and Sharp 1993; Ohta and Ina 1995). It is highly desirable to discriminate one effect from the other.

Recently, Moriyama and Powell (1996) summarized intraspecific variation in Drosophila. The amount of variation can be as high as 0.03 in terms of nucleotide difference per silent site in *Drosophila melanogaster* (*Amy*; Inomata *et al.* 1995) and 0.08 in *Drosophila simulans* (*Est-6*; Karotam *et al.* 1995), which is expected to have a big sampling variance in a sample of two sequences (Nei and Tajima 1981). On the other hand, the smallest estimate obtained in both *D. melanogaster* and *D. simulans* is 0. It is well known that in Drosophila the degree of intraspecies variation is highly dependent on the recombination rate in the region involved (Begun and Aquadro 1992). Thus, genes with low silent variation exclusively come from regions of very reduced recombination rates such as the tip of the *X* chromosome and the fourth chromosome. The changes of the banding pattern of the salivary gland polytene chromosome have been studied in some detail, and the structural changes were inferred for the *D. melanogaster* species subgroup (Lemeunier and Ashburner 1976). A comparison of *D. melanogaster* and *Drosophila yakuba* chromosomes, for instance, revealed the occurrences of at least four inversions on the *X* chromosome. As a result, the 2B region of *D. melanogaster* is located far from the tip of the *X* chromosome in *D. yakuba* due to a long inversion covering 2B through 6D. Frequent chromosomal structural changes mean that the degree of intraspecies variation may vary greatly, even in relatively closely related species pairs. In the presence of high variation and frequent chromosomal structural changes, the ancestral polymorphism could seriously inflate the rate variation in the case of a low degree of between-species divergence, 0.5, per site, whereas a comparison of more distantly related species pairs decreases the accuracy of estimates of the number of substitutions. The impact of ancestral polymorphism is presented graphically in Figure 1.

This article presents an analysis of the evolutionary rate variation among three species of Drosophila, *D. melanogaster*, *D. simulans* and *D. yakuba*, using the published sequences of eight genes from regions of normal recombination rates. At the same time, to exclude possible effects of ancestral polymorphism, we have taken advantage of the fact that the genes on the tip of the *X* chromosome and the fourth chromosome have very reduced levels of intraspecies variation and no structural changes were inferred in these regions among these three species (Lemeunier and Ashburner 1976). This provides a unique opportunity to study the fluctuation of the evolutionary rate, under the condition of the least effect of ancestral polymorphism. As a check on ancestral polymorphism, new data for four genes of the *achaete-scute* complex (*AS-C*) on the tip of the *X* chromosome and the *ci* gene on the fourth chromosome were also studied in this article. The results obtained suggest irregular synonymous substitution patterns in the three lineages both in the eight genes of normal recombination rate and in the four genes from the regions of reduced recombination. This study also provides evidence for the change in codon bias of the four *AS-C* genes between *D. melanogaster* and *D. simulans*, on the one hand, and *D. yakuba*, on the other hand.

MATERIALS AND METHODS

Sequence sources: The number of sequences studied, their sources and GenBank accession numbers (release 98.0) are as follows: eleven alleles of *D. melanogaster* (excluding Adh^{FChD}), six alleles of *D. simulans* and twelve strains of *D. yakuba* for the *Adh* from McDonald and Kreitman (1991); six alleles of *D. melanogaster* and six of *D. simulans* for the *per* from Kliman and Hey (1993a), and a *D. yakuba per* sequence (X61127) from Thackeray and Kyriacou (1990); the sequence of TN329 line of *D. melanogaster* (L22721) from Inomata *et al.* (1995) and a *D. simulans* (D17734) and a *D. yakuba* sequence (D17738) from Shibata and Yamazaki (1995) for the *Amy*; a *sry* sequence of *D. melanogaster* (X03121) from Vincent *et al.* (1985), and a *sry* sequence of *D. simulans* (U64718) and *D. yakuba* (U64719) from Caccone *et al.* (1996); a *D. melanogaster nullo* sequence (X65444) from Rose and Wieschaus (1992), and a *nullo* sequence of *D. simulans* (U44733) and *D. yakuba* (U44732) from Caccone *et al.* (1996); *G6pd* sequences of *D. melanogaster* (L13900), *D. simulans* (L13878) from Eanes *et al.* (1993) and a *D. yakuba* sequence (U42750) from Eanes *et al.* (1996); five strains of *D. melanogaster*, five of *D. simulans* and four of *D. yakuba* for the *boss* (Ayala and Hartl 1993); five strains of *D. melanogaster*, five of *D. simulans* and five of *D. yakuba* for the *Rh3* (Ayala *et al.* 1993); a *D. melanogaster* (M17120) from Villares and Cabrera (1987) and a *D. simulans* (X62400) from Martin-Campos *et al.* (1992) for the *ac*; a *sc* sequence of *D. melanogaster* (M17119) from Villares and Cabrera (1987); a *l'sc* sequence of *D. melanogaster* (X12549) from Alonso and Cabrera (1988); six alleles of *D. melanogaster* and six of *D. simulans* from Hilton *et al.* (1994) for the *ase*; ten strains of *D. melanogaster* and nine of *D. simulans* for the *ci* (Orenic *et al.* 1990; Berry *et al.* 1991).

Fly stock, cloning, sequencing and cytological examination: A strain of *D. yakuba* (stock no. 14021-0261.0) was provided by the National Drosophila Species Resource Center at Bowling Green, Ohio. A P1 phage clone, DS06327, covering a large portion of *AS-C* was gifted from the laboratory of T. Yamazaki (Kyushu University, Fukuoka, Japan). The genomic library of the *D. yakuba* strain was constructed in the λDASHII vector. The preparation of DNA probe consisted of the following: The fragments of the *ac*, *sc*, and *l'sc* genes were amplified by PCR from DS06327 and cloned into the pCRII vector by using the TA-cloning kit (Invitrogen, NV Leek, The Netherlands). The primers used for the amplification were derived from the published sequences as follows: 5′-TGTTTTTACTTGGCTCT GATGT-3′, and 5′-GTGTTATGGTTGGGTGCGACTA-3′ for the *ac* gene from the sequence in Villares and Cabrera (1987); 5′-AGGGTTTAGGACGAAGGGACT-3′, and 5′-AGA

AAATAGGGCGTGGTGGTAA-3′ for the *sc* (Villares and Cabrera 1987); and 5′-ATTCCAGCAGCAGCATTACCAG-3′, and 5′-TTTTCTATCATTGTCTTCCATT-3′ for the *l'sc* (Alonso and Cabrera 1988). The primers used for the construction of the *ase* and *ci* gene probes were the same as in Hilton *et al.* (1994) and Berry *et al.* (1991), respectively. The fragments were amplified from the *D. melanogaster* strains. The labeling of DNA probes with digoxigenin-dUTP and hybridization was carried out following the protocol from Boehringer Mannheim (Mannheim, Germany). A single phage clone of *D. yakuba* hybridized to the probe of *D. melanogaster* was selected for each gene, and the product of subcloning was inserted into the pBluescriptII vector. The DNA sequences for both strands were determined on an ABI PRISM 310 Genetic Analyzer (Perkin-Elmer Applied Biosystems, Foster City, CA).

The DNA sequences of the *sc* and *l'sc* genes were also determined for a strain of *D. simulans* collected in the Congo, Sim-5(G20), which was established by half sib-mating for 20 generations. A 1.2-kb *sc* and a 0.9-kb *l'sc* fragment (the same primers as those in probe construction), and a 0.6-kb *sc* (primers: 5′-CTCAGGGCATAATACCTACTA-3′ and 5′-AAACTGG TCTCGTGGTTCTTA-3′) and a 0.7-kb *l'sc* (5′-CTGCCGAA GTCGCTGCCTCTG-3′, and 5′-GCTCCCGTCAAGAACTGT TGC-3′) fragment were amplified, purified through columns, and used as templates for direct sequencing. The nucleotide sequence data reported in this article will appear in the DNA Data Bank of Japan, European Molecular Biology Laboratory and GenBank nucleotide sequence databases with the following accession numbers: AB005751 and AB005797–AB005802.

In situ hybridization with digoxigenin-labeled probes was performed on the salivary gland chromosomes of *D. yakuba*. The *D. yakuba sc*, *l'sc* and *ci* clones were labeled via PCR using the primers (5′-CTCATTCTTGTGTAAGGTGTC-3′, and 5′-AGGCACTTGGTTCAACTCAAC-3′) for the *sc*, and the same primers as in the screening probe construction for the *l'sc* (a 0.7kb fragment) and *ci* (Berry *et al.* 1991).

Sequence analysis: The numbers of synonymous and nonsynonymous sites were calculated according to Ina (1995). The arithmetic mean of the three species was employed in the following analysis. The ratio of transitional to transversional substitution rates was obtained as that at the third iteration cycle described in Ina's method 2, using all the synonymous changes among the three species involved. For 11 of the 13 genes, one sequence, randomly chosen from each species, was used to count the transitional and transversional synonymous changes at the third codon position and to estimate the number of synonymous and replacement sites when two or more sequences were available. These are listed below: *a* sequences of *D. melanogaster* and *D. yakuba* and *c* sequence of *D. simulans* (X57365; X57361) in McDonald and Kreitman (1991) for the *Adh*; ME-NJ1 of *D. melanogaster* (L07825) and SI-CA1 of *D. simulans* (L07826) for the *per* (Kliman and Hey 1993a); ME-NJ1 of *D. melanogaster* and SI-CA1 of *D. simulans* for the *ase* (Hilton *et al.* 1994); a *D. melanogaster* (X54360) and a *D. simulans* sequence with T at position 2338 of the *ci* (Orenic *et al.* 1990; Berry *et al.* 1991). For the remaining two genes, the transitional and transversional synonymous substitutions at the third codon position were counted for the entire group of samples, 14 strains from three species for the *boss* (Ayala and Hartl 1993) and 15 for the *Rh3* (Ayala *et al.* 1993), but the subsequent calculation was made in the same manner as for the other genes by using *a* sequence of *D. melanogaster*, *f* of *D. simulans* and *k* of *D. yakuba* in Ayala and Hartl (1993) for the *boss*, and *a* sequence of *D. melanogaster*, *f* of *D. simulans* and *p* of *D. yakuba* in Ayala *et al.* (1993) for the *Rh3*.

The number of substitutions was estimated by Kimura's (1980) formulae, and the means for all the combinations were given when two or more sequences were available. In addition to this estimate, the numbers of substitutions were counted for each lineage on the basis of the following parsimonious assumptions without correction. The nucleotide at the node was estimated as one(s) to require the smallest number of total substitutions. When the nucleotide at the node could not determine uniquely, the same probability was given to each evolutionary pathway except for the following cases: There were 11 codons for 13 genes, in which two or more substitutions occurred and two different pathways to require different numbers of replacement substitutions were possible. The weight factors for each pathway were calculated according to Miyata and Yasunaga (1980). Based on the factors obtained, these 11 codons were classified into two groups: one with roughly equal weight for two pathways (weight factors for favored pathway range from 0.50 to 0.65) and one with biased weight (weight factors for favored pathway range from 0.85 to 0.99). Thus, equal weight (0.5) was given to two pathways for the former (six codons) and the pathways to require the smaller number of replacements were taken uniquely for the latter (5 codons).

For segregating sites within species, first the probability that the ancestral nucleotide of that species was one of the segregating nucleotides was assumed to be given by their frequencies (Watterson and Guess 1977), and the same counting method was applied to each pathway. For example, let us assume that species A and B have nucleotides G, and that G and T are segregating in species C with frequencies of (5/6) and (1/6), respectively. In this case, the nucleotide at the node is inferred to be G irrespective of the ancestral nucleotide of species C. The probability that the ancestral nucleotide of species C is G is (5/6), in which case the number of substitutions are calculated to be (1/6) in the species C lineage. The probability that the ancestral nucleotide of species C is T is (1/6), where the number of substitutions are calculated to be (1 + 5/6). In sum, the number of substitutions in the species C lineage is estimated to be

$$\left(\frac{5}{6}\times\frac{1}{6}\right)+\left\{\frac{1}{6}\times\left(1+\frac{5}{6}\right)\right\}=\frac{4}{9}.$$

These numbers of substitutions obtained without correction were employed for all the statistical tests except for the relative rate test (Wu and Li 1985).

RESULTS

Evolutionary rate variation in eight genes from regions of normal recombination: DNA sequences from *D. melanogaster*, *D. simulans* and *D. yakuba* were compared for eight genes: the *Adh* (cytological map position, 35B), *per* (3B), *Amy* (54A), *sry* (99D), *nullo* (6F), *G6pd* (18D), *boss* (96F) and *Rh3* (92D). None of these genes are located in the regions of very reduced recombination rate such as the tip of the *X* and the proximal regions of the chromosomes. *D. simulans* is one of the most closely related species to *D. melanogaster*, and *D. yakuba* is distantly related to these two species (Species C in Figure 1). Thus, the *D. yakuba* lineage in this study represents the entire lineage from the common ancestor of *D. melanogaster* and *D. simulans* to *D. yakuba*. The main purpose of this study is to compare the substitution pattern in the three lineages among different genes and to know the degree of variation. With this aim, no information about direction of mutations in the

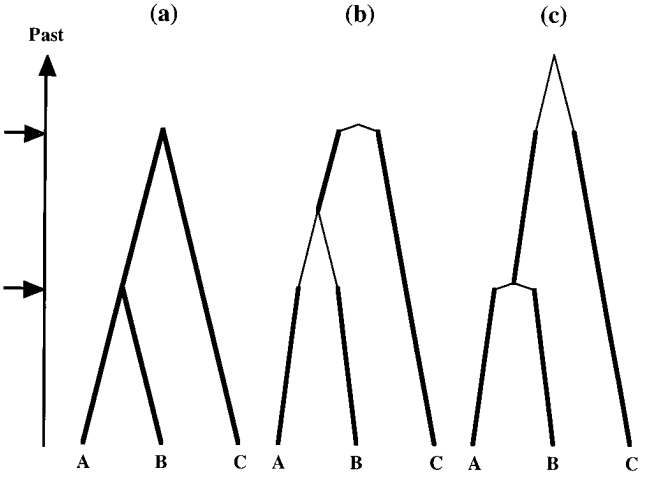

Figure 1.—Possible gene trees of a sample of three genes, one from each of three species. Thin lines in (b) and (c) stand for ancestral polymorphism at the time of species separation shown by arrows. (a) shows a history without ancestral-population variation. (b) and (c) depict the impact of ancestral polymorphism: (b) high variation in ancestral population of species A and B but low variation in common ancestor of three species, and the opposite pattern in (c). The mixture of these histories among genes produces variation in the number of substitutions inferred for three lineages even under the condition of constant evolutionary rate.

D. yakuba lineage (either of occurrence of mutation in the lineage leading to *D. yakuba* or in the internal branch from the common ancestor of the three species through the common ancestor of *D. melanogaster* and *D. simulans*) does not affect the results. Tables 1 and 2 show the estimates of the number of synonymous and replacement substitutions per site by Kimura's two parameter method (1980) and the estimated numbers of substitutions in the three lineages on a basis of parsimony assumptions, respectively.

Nucleotide substitution pattern in the three lineages was examined for lineage effects and locus-lineage interaction effects. The lineage effect was clearly seen, both in the synonymous and replacement substitutions, as the higher evolutionary rate in the *D. melanogaster* in comparison with the *D. simulans* lineage as described in Akashi (1995, 1996). For synonymous substitution, the relative rate test (Wu and Li 1985) showed the ratio of *melanogaster-simulans* difference to its standard deviation as 2.0 or more in two genes, the proximal gene of *Amy* and *sry*. And similar results were found for replacement substitution in three genes, the proximal gene of *Amy*, *sry* and *G6pd*. Only the replacement substitution of the *G6pd* showed significantly faster evolution in *D. simulans* than in *D. melanogaster*. Since the amount of within-species variation is greater in *D. simulans* than in *D. melanogaster* (Moriyama and Powell 1996), a difference in mutation rate between the two species is an unlikely cause for the higher evolutionary rate in *D. melanogaster*. The most likely explanation is a lower efficiency of selection in the *D. melanogaster* lineage due to a smaller population size (Akashi 1995) and a lower recombination rate in *D. melanogaster* (Ohnishi and Voelker 1979; True *et al.* 1996; Hamblin and Aquadro 1996) as compared to *D. simulans*.

Besides the lineage-dependent pattern of evolution, the synonymous changes in the three lineages occurred irregularly among the eight genes (3×8 test of independence in Sokal and Rohlf 1995; G with Williams's correction $= 31.8$ with 14 df, $P < 0.005$). Although the same kind of statistical test was not applied due to the small number of replacement substitutions in each lineage, the replacement substitution pattern for the *G6pd* and *per* genes showed a higher dispersion index. The faster protein sequence evolution of the *G6pd* gene in *D. simulans* has been described well in Eanes *et al.* (1993, 1996).

One possible explanation for the observed rate-irregularity is gene-dependent fluctuation of mutation rate. If the rate of mutation varies among lineages, then there will be a correlation between synonymous and replacement substitution rates. However, statistically significant heterogeneities were found in the numbers of synonymous and replacement substitutions for the *per* (a larger ratio of replacement to synonymous substitutions in the *D. yakuba* than the other two lineages: Fisher's exact test, $P = 0.0006$), and *Rh3* (a larger ratio of replacement to synonymous substitutions in the *D. melanogaster* lineage as compared to the other lineages: Fisher's exact test, $P = 0.017$). Since any one of the three lineages could show a significant deviation from the other two lineages, we may take the significance level of 0.017 ($= 1 - 095^{1/3}$) instead of 0.05. There is still a significant heterogeneity between the two types of substitutions in these two genes. Thus, it is suggested that fluctuation of mutation rate alone cannot explain the whole picture of the observed substitution patterns.

Cytological examination of *AS-C* genes and *ci* gene in *D. yakuba*: Lemeunier and Ashburner (1976) compared the banding pattern of the salivary chromosomes among six sibling species of *D. melanogaster*. The findings showed that the fourth chromosomes of *D. melanogaster* and *D. yakuba* were generally similar to each other, but not enough so to be conclusive. The identity of the tip of the *X* chromosomes of the three species is also ambiguous (Horton 1939). Then, *in situ* hybridization was performed on the polytene chromosomes of *D. yakuba* with the *D. yakuba* clones. Figure 2 shows that the *l'sc* and *ci* genes are located on the tip of the *X* and the fourth chromosome in *D. yakuba*, respectively, as in *D. melanogaster*.

Evolutionary rate variation in four genes of *achaete-scute* complex and *ci* gene sampled from regions of reduced recombination rates: The four *AS-C* genes, *ac*, *sc*, *l'sc* and *ase*, and the *ci* gene were cloned from a genomic library of a *D. yakuba* strain, and their nucleotide sequences were determined. The length of se-

TABLE 1

Coding sequence divergence among *D. melanogaster*, *D. simulans* and *D. yakuba*

Gene	Synonymous		Replacement	
	Length	Estimate of number of substitutions per site	Length	Estimate of number of substitutions per site
Adh[a]				
mel-sim	199.4	0.0559 ± 0.0174	565.6	0.0043 ± 0.0028
mel-yak		0.1577 ± 0.0313		0.0115 ± 0.0045
sim-yak		0.1338 ± 0.0281		0.0107 ± 0.0044
per[a]				
mel-sim	410.8	0.1216 ± 0.0188	1268.2	0.0020 ± 0.0012
mel-yak		0.3112 ± 0.0332		0.0242 ± 0.0044
sim-yak		0.2654 ± 0.0295		0.0242 ± 0.0044
Amy-proximal				
mel-sim	356.6	0.0928 ± 0.0170	1122.5	0.0099 ± 0.0030
mel-yak		0.1224 ± 0.0200		0.0180 ± 0.0040
sim-yak		0.0862 ± 0.0163		0.0099 ± 0.0030
sry				
mel-sim	412.1	0.1229 ± 0.0189	1043.9	0.0116 ± 0.0034
mel-yak		0.2728 ± 0.0313		0.0283 ± 0.0053
sim-yak		0.1905 ± 0.0246		0.0204 ± 0.0045
nullo				
mel-sim	136.8	0.0941 ± 0.0279	384.2	0.0212 ± 0.0075
mel-yak		0.5274 ± 0.0878		0.0568 ± 0.0125
sim-yak		0.4684 ± 0.0790		0.0624 ± 0.0132
G6pd				
mel-sim	377.7	0.1113 ± 0.0182	1177.3	0.0189 ± 0.0040
mel-yak		0.1485 ± 0.0221		0.0163 ± 0.0038
sim-yak		0.1536 ± 0.0222		0.0268 ± 0.0048
boss[a]				
mel-sim	416.3	0.0902 ± 0.0156	1149.7	0.0042 ± 0.0019
mel-yak		0.1874 ± 0.0243		0.0053 ± 0.0021
sim-yak		0.1670 ± 0.0225		0.0018 ± 0.0012
Rh3[a]				
mel-sim	307.0	0.0833 ± 0.0174	839.0	0.0036 ± 0.0021
mel-yak		0.1644 ± 0.0256		0.0036 ± 0.0021
sim-yak		0.1527 ± 0.0247		0.0
ac				
mel-sim	168.2	0.0971 ± 0.0261	431.8	0.0093 ± 0.0047
mel-yak		0.1622 ± 0.0351		0.0429 ± 0.0102
sim-yak		0.1949 ± 0.0399		0.0356 ± 0.0092
sc				
mel-sim	253.4	0.0929 ± 0.0203	742.6	0.0150 ± 0.0045
mel-yak		0.2396 ± 0.0363		0.0233 ± 0.0057
sim-yak		0.2009 ± 0.0322		0.0191 ± 0.0051
l'sc				
mel-sim	198.4	0.0364 ± 0.0140	566.6	0.0107 ± 0.0044
mel-yak		0.2738 ± 0.0460		0.0197 ± 0.0060
sim-yak		0.2669 ± 0.0454		0.0161 ± 0.0054
ase[a]				
mel-sim	290.7	0.0506 ± 0.0139	771.4	0.0168 ± 0.0047
mel-yak		0.2484 ± 0.0353		0.0339 ± 0.0068
sim-yak		0.2390 ± 0.0342		0.0244 ± 0.0057
ci[a]				
mel-sim	246.2	0.1011 ± 0.0218	711.8	0.0303 ± 0.0066
mel-yak		0.2214 ± 0.0354		0.0464 ± 0.0083
sim-yak		0.2497 ± 0.0382		0.0406 ± 0.0077

The numbers of synonymous and replacement sites were calculated by Ina's (1995) method. The number of substitutions per site was estimated by Kimura's (1980) formulae.

[a] The numbers of differences are the means for multiple alleles described in materials and methods.

TABLE 2

Number of substitutions in three lineages and dispersion index in eight genes from regions of normal recombination rates

	Synonymous					Replacement				
		Number of substitutions			Dispersion		Number of substitutions			Dispersion
Gene	Length	*D. melanogaster*	*D. simulans*	*D. yakuba*	index	Length	*D. melanogaster*	*D. simulans*	*D. yakuba*	index
Adh	199.4	7.9 (8.4)	3.2 (4.7)	23.5 (21.0)	0.4	565.6	1.5 (1.5)	1.0 (1.0)	5.0 (5.0)	0.1
per	410.8	34.4 (30.8)	15.6 (21.8)	93.4 (75.3)	0.6	1268.2	1.3 (1.3)	1.2 (1.6)	29.4 (28.9)	7.5
Amy-p	356.6	23.0 (21.0)	10.1 (10.0)	20.6 (18.5)	2.7	1122.5	10.1 (10.0)	1.0 (1.0)	10.1 (10.5)	4.3
sry	412.1	42.3 (35.3)	8.4 (11.3)	70.1 (57.3)	4.7	1043.9	10.2 (10.0)	1.9 (2.0)	19.4 (19.0)	2.3
nullo	136.8	10.5 (8.3)	2.4 (4.3)	61.7 (42.8)	8.4	384.2	3.0 (3.0)	5.1 (5.0)	18.8 (18.5)	1.1
G6pd	377.7	20.1 (18.2)	22.0 (20.2)	36.0 (31.7)	7.6	1177.3	5.0 (5.8)	17.3 (17.8)	14.2 (14.3)	10.9
boss	416.3	23.0 (22.2)	14.5 (16.6)	55.0 (47.4)	0.4	1149.7	4.4 (4.4)	0.4 (0.4)	1.6 (2.0)	2.8
Rh3	307.0	14.6 (13.9)	11.0 (14.3)	35.9 (31.8)	0.8	839.0	3.0 (3.0)	0.0 (0.0)	0.0 (0.0)	3.3
Total		175.8 (158.2)	87.1 (103.4)	396.3 (325.9)			38.4 (39.0)	28.0 (28.8)	98.7 (98.2)	

The number of substitutions in the three lineages was calculated from the number of substitutions in the three species pairs, which was derived from the product of the number of synonymous or replacement sites and Kimura's (1980) estimate of the number of substitutions per site. Numbers in parentheses are the number of substitutions inferred on parsimony assumptions (see materials and methods). Following Gillespie's (1989) method, dispersion indices for synonymous and replacement substitutions were weighted by synonymous and replacement weight factors, respectively. These factors were calculated from the pooled number of substitutions in the three lineages: 0.800 for the *D. melanogaster*, 0.396 for the *D. simulans* and 1.804 for the *D. yakuba* lineage in synonymous substitution; and 0.698 for *D. melanogaster*, 0.509 for *D. simulans* and 1.794 for *D. yakuba* in replacement substitution. The totals do not, in many cases, match the sum in the columns due to the accumulation of rounding errors.

quence determined was 878 bp for the *ac*, 1314 bp for the *sc*, 924 bp for the *l'sc*, 1564 bp for the *ase* gene, all of which include the entire putative coding sequences. Because only partial sequences of the *D. simulans ase* were available, the corresponding coding regions of the *ase* genes were employed in the analysis. A 1293-bp fragment of the *ci* gene was sequenced, and this region contained the whole of the coding regions surveyed in Berry *et al.* (1991). In addition, the entire coding regions of the *sc* and *l'sc* of *D. simulans* were amplified and their sequences were determined for the analysis.

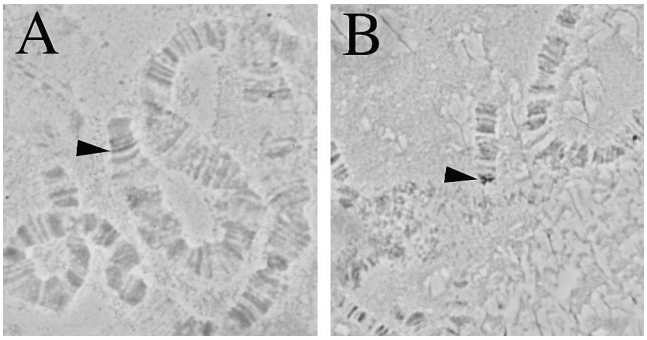

Figure 2.—Localization of the *l'sc* and *ci* genes on the polytene chromosome of *D. yakuba*. *In situ* hybridization using the *D. yakuba l'sc* and *ci* clones as probes to salivary gland chromosomes indicated that the *l'sc* (A) and *ci* (B) genes are located on the tip of the *X* and the fourth chromosome, respectively, just as in *D. melanogaster*. The *sc* gene is also localized on the tip of the *X* chromosome (data not shown).

A few of the insertions/deletions in the coding regions among the three species were observed as follows: a two-amino-acid insertion in the *D. yakuba sc* sequence and a one-amino-acid deletion in the *D. melanogaster sc*; a one-amino-acid deletion in the *D. yakuba l'sc*; a two-amino-acid insertion in the *D. melanogaster ase*; one- and nine-amino-acid insertions and a two-amino-acid deletion in the *D. yakuba ase*; two two-amino-acid insertions in the *D. yakuba ci*; no length variation in the *ac*. A Gln-rich sequence was found in the *ase* gene of *D. melanogaster* [$(Gln)_5$-Val-Gln], which is the same as in the *ac* and *sc* genes (Villares and Cabrera 1987). There was, however, a nine-amino-acid insertion (duplication) of the Gln-rich sequence in *D. yakuba*, resulting in the sequence $(Gln)_{12}$-Pro-Gln-Val-Gln. Besides these gaps, there may be a difference in the translation start site of

```
523 ATAACGCGAGGGTTTAGGACGAAGGGACTCATTCTTGTGTAAGGTGTCAAACGATCAAGT
    ************************************************************
    ATAACGCGAGGGTTTAGGACGAAGGGACTCATTCTTGTGTAAGGTGTCAAACGATCAAGT

583 TCAAGTATTGTACTCTGTTCATTTATTTTTTTCTGTTGATCGTTATCCGGAAAGTGAAAG
    *****  ** *** ** * ************************* ******** * ****
    TCAAGCCTTCTACCCTCTGCATTTATTTTTTTCTGTTGATCGTTCTCCGGAAAATAAAAG

                          ***
643 AAAGCTCCGA------GTGTGTTAATGAAAAACAATAATAAT---ACAACGAAAAGCACT
    **********      * ***** * *****  ********    ************** *
    AAAGCTCCGAATCTGTGCGTGTTTAAGAAAACAAATAATAAGAGTACAACGAAAAGCATT

       ###
694 ACCATGTCATCGAGTGTGCTGTCCACCAACGAAACGTTTCCAACGACCATCAATTCGGCA
    * ************************ ***** ***************** **********
    ATCATGTCATCGAGTGTGCTGTCCACAAACGATACGTTTCCAACGACCATTAATTCGGCA
```

Figure 3.—Substitution in the first initiation codon of the *sc* transcript in *D. yakuba*. Upper and lower sequences show the *D. melanogaster* and *D. yakuba* sequences, respectively. * indicates the ATG codon in *D. melanogaster* as well as in *D. simulans*, but AAG in the *D. yakuba* sequence. # is the conserved ATG codon for all the three species. Gaps are shown by hyphens. The sequence and position of *D. melanogaster* refer to the sequence in Villares and Cabrera (1987).

the *sc* transcripts as shown in Figure 3. These gaps, including the sequence anterior to the conserved ATG codon in the *sc* sequence, were excluded in the following analyses.

The summary of DNA variation among the three species is given in Tables 1 and 3. The ratio of replacement to synonymous substitutions for the pooled data varied significantly among the three lineages (G with Williams's correction = 11.6, $P < 0.005$). This is partly due to the accelerated protein sequence evolution in the *D. melanogaster* lineage (Akashi 1996) but no acceleration for synonymous substitution. No lineage effect in synonymous substitution provides an interesting contrast to the pattern observed for the eight genes from the regions of normal recombination, in which the greater number of substitutions in the *D. melanogaster* lineage as compared to those of the *D. simulans* lineage was observed in synonymous substitution as well as in replacement substitution (Table 2). The reduced effectiveness of selection in regions of reduced recombination is a plausible cause for no effect of the population sizes in the two species.

When we examined the pattern of substitution in the three lineages, there was a significant heterogeneity among the five genes for synonymous substitution as in the genes with normal recombination (G with Williams's correction = 19.6, $P < 0.025$). This result suggests that the presence of locus-lineage interaction for synonymous substitution rate does not simply come from effects of ancestral polymorphism. For replacement substitution, only the *ac* gene showed a significantly different substitution pattern from the other four genes (G with Williams's correction = 4.2, $P < 0.05$ for the numbers of substitutions between *D. melanogaster* + *D. simulans* and *D. yakuba*). Lower frequency of statistical significance for replacement substitution patterns is at least partly due to the small number of substitutions in each lineage, and it should not be simply compared to synonymous substitution.

The ratio of replacement to synonymous substitutions was significantly higher in the *D. melanogaster* for the *ase* (corrected $G = 6.1$, $P < 0.025$, for *D. melanogaster* vs. *D. simulans* + *D. yakuba*), and higher in the *D. melanogaster* and *simulans* lineages for the *l'sc* (corrected $G = 4.9$, $P < 0.05$, for *D. melanogaster* + *D. simulans* vs. *D. yakuba*). These results again suggest that the observed erratic substitution patterns are not entirely because of gene-dependent fluctuation of mutation rate.

TABLE 3

Number of substitutions in three lineages and dispersion index in five genes from regions of reduced recombination rates

	Synonymous					Replacement				
		Number of substitutions			Dispersion		Number of substitutions			Dispersion
Gene	Length	*D. melanogaster*	*D. simulans*	*D. yakuba*	index	Length	*D. melanogaster*	*D. simulans*	*D. yakuba*	index
ac	168.2	5.4	10.9	21.9	2.7	431.8	3.6	0.4	14.9	2.7
		(5.5)	(9.5)	(19.0)			(3.7)	(0.7)	(13.7)	
sc	253.4	16.7	6.9	44.0	3.3	742.6	7.1	4.0	10.2	0.2
		(14.7)	(7.7)	(36.7)			(7.0)	(4.0)	(10.0)	
l'sc	198.4	4.3	2.9	50.0	3.1	566.6	4.1	2.0	7.1	0.0
		(4.0)	(3.0)	(40.0)			(4.0)	(2.0)	(7.0)	
ase	290.7	8.7	6.0	63.5	1.5	771.4	10.2	2.8	16.0	0.5
		(9.0)	(6.0)	(52.0)			(10.8)	(3.2)	(15.7)	
ci	246.2	9.0	15.9	45.5	2.2	711.8	12.8	8.7	20.2	0.8
		(9.8)	(14.8)	(36.8)			(12.7)	(9.0)	(20.2)	
Total		44.1	42.6	224.9			37.8	18.0	68.4	
		(43.0)	(41.0)	(184.5)			(38.2)	(18.8)	(66.5)	

The weight factors for the *D. melanogaster*, *D. simulans* and *D. yakuba* lineages are 0.424, 0.411, and 2.165 for synonymous substitution and 0.913, 0.434 and 1.653 for replacement substitution, respectively. The mismatches in the total numbers are due to rounding.

TABLE 4

Comparison of numbers of substitutions in three lineages between five genes in regions with reduced recombination rates and eight genes with normal recombination rates

	Synonymous			Replacement		
	D. melanogaster	*D. simulans*	*D. yakuba*	*D. melanogaster*	*D. simulans*	*D. yakuba*
Four *ASC* and *ci* genes	43.0	41.0	184.5	38.2	18.8	66.5
Eight genes[a]	158.2	103.4	325.9	39.0	28.8	98.2
	$G = 15.8$[b]			$G = 2.0$		

Assignment of substitutions in the three lineages was inferred on a basis of parsimony assumptions (see materials and methods).

[a] The *Adh*, *per*, *Amy*, *sry*, *nullo*, *G6pd*, *boss* and *Rh3* genes are involved.

[b] The *G*-test of independence with Williams's correction was highly significant in synonymous substitution ($P < 0.001$), but not in replacement substitution.

Comparison of the four *AS-C* and *ci* genes with the remaining eight genes: A comparison was made of the pooled number of the substitutions in the three lineages (Table 4). There was a significant heterogeneity between the *AS-C* and *ci* genes, and the remaining eight genes for synonymous substitution, but not for replacement substitution. One reason for the heterogeneity in synonymous substitution can be attributed to a lack of difference between the *D. melanogaster* and *D. simulans* lineages of the four *AS-C* and *ci* genes. This is most likely due to the low efficiency of natural selection in regions of reduced recombination, irrespective of the population size difference between the two species. Indeed, the codon bias for these genes is generally very low for the *D. melanogaster* genes as described in Kliman and Hey (1993b). A significant heterogeneity between the two samples of genes still existed for a *D. melanogaster* + *D. simulans* vs. *D. yakuba* comparison (corrected $G = 13.7$, $P < 0.001$). As mentioned in the Introduction, ancestral polymorphism is one of the possible causes. The current findings may be explained by the higher variation in an ancestral population of *D. melanogaster* and *D. simulans* than that in an ancestral population of the three species (Tree b in Figure 1). This, however, seems unlikely because replacement substitution showed the opposite pattern, that is, a higher ratio of the number of substitutions in the *D. yakuba* lineage as compared to those of the *D. melanogaster* and *D. simulans* lineages for genes with higher recombination rates than the four *AS-C* and *ci* genes. The present finding may also be explained if the following conditions exist: Natural selection acts on synonymous changes and the degree of preference for codons varies among species, depending on genes.

Akashi (1995) classified codons into two types, "major (preferred)" and "nonmajor (unpreferred)." Major codons are defined as codons which appear in higher frequency in higher-biased genes. Synonymous substitutions are subsequently classified into preferred substitutions which mean substitutions from a nonmajor codon to a major codon, unpreferred substitutions from a major codon to a nonmajor codon, and the others as explained in Akashi (1995). The numbers of these substitutions in this study are given in Table 5. Because we cannot determine the direction of the substitutions in the *D. yakuba* lineage without any outgroup species, the numbers of homologous codons at which *D. yakuba* encodes a major codon and the common ancestor of *D. melanogaster* and *D. simulans* encodes a nonmajor codon (y_{maj}/ms_{non}) and the codons in the opposite configuration (y_{non}/ms_{maj}) for the substitutions in the *D. yakuba* lineage are given in Table 5. An excess of preferred substitutions over unpreferred substitutions (or the number of y_{maj}/ms_{non} over that of y_{non}/ms_{maj}) was observed in the *D. yakuba* lineage for the genes from the regions of reduced recombination, although in the *ci* gene the two types of substitutions occurred almost the same number of times (corrected $G = 17.4$, $P < 0.001$, for 19.9 preferred and 36.3 unpreferred substitutions in the *D. melanogaster* and *D. simulans* lineages, and 89.2 y_{maj}/ms_{non} and 41.2 y_{non}/ms_{maj} in *D. yakuba*). The y_{maj}/ms_{non} substitution in Table 5 refers to the occurrence of preferred codon substitution in the lineage leading to *D. yakuba* or unpreferred codon substitution in the internal branch from the common ancestor of the three species through the common ancestor of *D. melanogaster* and *D. simulans*. In any case, the important finding is that the change of codon bias depends on genes. Indeed, a significant heterogeneity in the numbers of y_{maj}/ms_{non} and y_{non}/ms_{maj} in the *D. yakuba* lineage existed between the genes from different recombination rates (corrected $G = 19.0$, $P < 0.001$), but neither in the *D. melanogaster* nor in the *D. simulans* lineage. The higher codon bias for the four *AS-C* genes in *D. yakuba* as compared to *D. melanogaster*/*D. simulans* suggests change(s) in selection intensity on codon usage for the four *AS-C* genes during the evolutionary course of these three species. The higher number of y_{maj}/ms_{non} substitutions is one of the causes of a significant heterogeneity in the substitution pattern between the genes

TABLE 5

"Preferred" and "unpreferred" synonymous substitutions

	D. melanogaster			*D. simulans*			*D. yakuba*		
Gene	Preferred	Unpreferred	Neutral	Preferred	Unpreferred	Neutral	y_{maj}/ms_{non}	y_{non}/ms_{maj}	Neutral
ac	1.0	4.5	0	5.5	3.0	1.0	10.0	4.0	5.0
sc	2.0	7.3	5.3	3.0	3.3	1.3	16.7	9.0	11.0
l'sc	1.0	2.0	1.0	2.0	1.0	0	27.0	6.0	7.0
ase	1.8	3.2	4.0	1.0	3.0	2.0	22.5	10.5	19.0
ci	0.3	3.0	6.5	2.3	6.0	6.4	13.0	11.7	12.2
Total	6.1	20.0	16.8	13.8	16.3	10.7	89.2	41.2	54.2
Adh	1.8	6.7	0	1.1	2.4	1.2	11.2	4.4	5.4
per	1.7	18.0	11.2	4.9	11.6	5.2	24.3	30.0	21.1
Amy-p	1.0	17.0	3.0	0	7.0	3.0	9.5	6.0	3.0
sry	16.7	13.3	5.3	4.7	6.3	0.3	9.0	31.7	16.7
nullo	0.7	7.3	0.3	2.0	0.7	1.7	13.7	15.8	13.3
G6pd	2.0	14.3	1.8	7.5	6.8	5.8	6.8	19.0	5.8
boss	1.8	14.0	6.4	4.0	7.8	4.8	24.6	12.4	10.5
Rh3	1.9	7.3	4.7	2.4	9.3	2.6	9.4	13.6	8.8
Total	27.6	97.9	32.7	26.6	51.9	24.6	108.5	132.9	84.6

Preferred and unpreferred substitutions are based on Akashi (1995). y_{maj}/ms_{non} stands for codons that encode a major codon in *D. yakuba* and a nonmajor codon in the common ancestor of *D. melanogaster* and *D. simulans*, and y_{non}/ms_{maj} represents codons that encode a nonmajor codon in *D. yakuba* and a major codon in the common ancestor of *D. melanogaster* and *D. simulans*.

with reduced and normal recombination rates. This may also have made a contribution to the erratic synonymous substitution pattern among the three lineages. In fact, there was a significant heterogeneity in the numbers of y_{maj}/ms_{non} and y_{non}/ms_{maj} for the eight genes in regions of normal recombination (corrected $G = 25.8$, $P < 0.001$).

DISCUSSION

While the dispersion index has been used as a parameter to test rate constancy in the previous studies (Gillespie 1989; Ohta 1995), there are a few criticisms of this index. Bulmer (1989) pointed out that the corrections for multiple hits significantly inflate the index even in the distance of 0.25. He developed new test statistics, assuming a star phylogeny. This assumption does not hold for the three species studied here. For this reason, the number of substitutions inferred under the parsimony assumptions were employed for the statistical tests in this study, although the mean number of substitutions per site in each lineage was small even for the synonymous ones: 0.05 in the *D. melanogaster* lineage, 0.03 in *D. simulans* and 0.18 in *D. yakuba*. Goldman (1994) criticized the statistical significance of the dispersion index in Gillespie's study (1989). Gene-dependent construction of phylogenies with estimated branch length in Goldman's study is problematic as mentioned in Nielsen (1997). In fact, Goldman assumed the branching order of {[(goat – rabbit) – human] – mouse} for the α-*hemoglobin* gene, and [(cow – mouse) – human] for the *cytochrome oxidase 2* gene. The difference in phylogenies assumed among genes should be tested as well. High degrees of ancestral polymorphism would be required to explain the difference in the two genes. Although more careful examination of effects of sampling designs and assumptions is needed, the index can be a useful measure of rate variation in certain conditions.

The principal findings of this study were (1) erratic synonymous substitution patterns among the three lineages and (2) lineage effect in protein sequence evolution. The first finding was obtained independently from the published sequences of the eight genes from the regions of normal recombination and the five genes located in regions of greatly reduced recombination rates in *D. melanogaster*. Consequently, the erratic synonymous substitution patterns cannot be attributed to the effects of ancestral polymorphism. One possible cause is fluctuation of the mutation rate, but the high occurrence of significant heterogeneity of synonymous and replacement substitution patterns does not support this hypothesis. An alternative, and more plausible, explanation is the changes in degree of selective constraint involved in codon usage. Indeed, higher codon bias was observed in the four *AS-C* genes of *D. yakuba* as compared to *D. melanogaster* and *D. simulans*.

The estimated dispersion index for the replacement substitutions was generally low, and only the *per*, *G6pd* and *ac* genes showed an indication of erratic evolution among the three lineages. In contrast to this finding, previous reports on three orders of mammals show generally higher dispersion indices for replacement substitutions than for synonymous ones (Gillespie 1989; Ohta 1995). It is of great interest to analyze possible

causes for the difference in the rate variation between the two studies, one of Drosophila and the other of mammals, in detail. Effects of the differences in the analytical designs between the two studies, however, should be taken into consideration at first. The smaller number of genes examined in this study, as compared with 49 genes in Ohta (1995), could produce a larger correlation between the weight factors and the number of substitutions at individual genes. Thus, the relatively low dispersion indices in the Drosophila genes could be contributed by the sample size. The level of divergence in this study is also lower than in the mammalian studies: the mean numbers of substitutions per site in all the lineages in the present study and in Ohta (1995) were 0.27 and 0.67 for the synonymous substitutions and 0.03 and 0.16 for the replacement substitutions, respectively. Ohta (1995) found a significant correlation between the dispersion index and the number of substitutions, which could be due to the correction bias (Bulmer 1989) and/or other reasons (Ohta 1995). Thus, the difference in the levels of divergence is a possible explanation for the difference in dispersion index between the two studies. In addition, branching patterns may also affect the levels of the dispersion index. The difference among the weight factors for the three lineages is smaller in Ohta (1995) than in this study. The three weight factors for the replacement substitutions are 1.3, 1.0 and 0.8 in Ohta (1995), while those in this study were 1.7, 0.9 and 0.4 for the five genes from the regions of reduced recombination (Table 3). Bias in the weight factors might reduce the power of detection of rate variation as compared with a star phylogeny with the same levels of total divergence.

Besides these technical biases due to the different research designs employed in each study, the higher levels of overdispersion for the replacement substitutions in the mammals as compared to Drosophila may reflect differences in the degree of functional redundancy due to gene duplications. Accelerated protein sequence evolution has been observed during the limited periods following gene duplications in a number of genes (Li 1985; Ohta 1994; Iwabe *et al.* 1996), although whether the accelerated rates are due to relaxation of selective constraints or accumulation of advantageous mutations is a controversial issue. Furthermore, a significant fraction of erratic evolution observed in the mammals comes from multiplicated genes such as the growth hormone and prolactin genes (Gillespie 1989; Ohta 1995). It is certainly true that there is a big difference in the number of genes between mammals and Drosophila (Lewin 1997). Advances in genome projects for many different taxa will help measure and compare the degree of redundancy or gene duplication in genomes. Keeping in mind the possible biases in comparisons, further study is needed to understand the relationship between the degree of redundancy and frequency of episodes in molecular evolution.

Another factor which may affect relative degrees of rate variation for the synonymous and replacement substitutions is the efficiency of natural selection. As mentioned above, the higher rate variation in the Drosophila genes could be due to a combination of weak selection with fluctuating selection intensities and fluctuating population sizes. Lower selection efficiency on the synonymous substitutions due to smaller population sizes of the mammals and thus lower rate variation could partly explain the relatively higher dispersion index for the replacement substitutions.

Akashi (1995, 1996) discovered the lineage effect, that is, the accelerated protein sequence evolution and the accumulation of unpreferred synonymous changes in the *D. melanogaster* lineage as compared to the *D. simulans* lineage. The present study showed no acceleration in the synonymous substitution rate in *D. melanogaster* for the five genes from the regions with reduced recombination, while a significantly higher accumulation of synonymous substitutions was observed for the eight genes from the regions with variable recombination rates in the *D. melanogaster* lineage as compared to the *D. simulans* lineage. The absence of lineage effect in the four *AS-C* and *ci* genes can be explained by less efficient natural selection in the regions of very reduced recombination rates. On the other hand, there still exists a significantly greater accumulation of replacement substitutions in the five genes in the regions of low recombination in the *D. melanogaster* lineage as compared to the *D. simulans* lineage ($G = 6.7$ with Williams's correction, $P < 0.01$). This suggests that a significant fraction of replacement changes has greater selection coefficients than synonymous ones, but that it may be fixed in smaller populations such as *D. melanogaster*.

The four *AS-C* genes, the proneural *ac*, *sc* and *l'sc* genes, and the neural precursor gene *ase*, encode the basic helix-loop-helix (bHLH) class of transcription regulators. Although these proteins are not completely interchangeable in several aspects such as the numerator function of the *sc* (*sisterless b*) in sex determination, a great deal of functional redundancy has been reported among these genes in the proneural function promoting neural development (*e.g.*, Parkhurst *et al.* 1993; Balcells *et al.* 1988; Jiménez and Campos-Ortega 1990; Parras *et al.* 1996). Based on their functions in neural development and sex determination, the level of expression of these genes is considered to be very important in the course of development. Comparison of the numbers of "preferred" and "unpreferred" codon changes revealed a significant difference between the *D. melanogaster-simulans* lineage and the *D. yakuba* lineage, suggesting change(s) in selection intensity. Natural selection intensity varies due to changes in the required level of expression. Efficiency of selection can also vary according to the recombination rate of the region involved, although the *AS-C* genes are cytologically located in the tip region of the *X* chromosome in

all three species. Investigation of intraspecific variation of these genes and of other genes in this region in *D. yakuba* will help determine which is a better explanation for the variation in natural selection intensity.

Nonrandomness of distribution of intra- or interspecific variation along DNA sequences is often studied by window analysis. This analysis, coupled with knowledge of protein structure and its function, can be a powerful tool to detect the action of natural selection. Nonrandomness, however, does not necessarily come from variation in mutation rate or degree of selective constraint along the sequence alone. If we can infer which lineages have mutations (based on the parsimonious assumption or some other methods), nonrandom distribution of substitutions may be further divided into lineage-dependent/lineage-specific clustering and lineage-independent clustering. The former refers to a clustering of substitutions along sequences in individual lineages, and the occurrence of lineage-dependent clustering may be due to epistatic interaction among sites including compensatory mutations. The pattern observed in the *G6pd* gene is a good example of the lineage-specific clustering of substitutions. As mentioned above, Eanes *et al.* (1993, 1996) showed the accelerated protein sequence evolution in the *D. simulans* lineage (15 replacement substitutions in *D. simulans* and 6 in *D. melanogaster*). Many of the replacement changes in *D. simulans* are found in the very narrow regions in exons 2 and 3, but the substitutions in the *D. melanogaster* lineage are evenly distributed over the entire region examined. The *ac* and *per* genes showed relatively higher rates of protein sequence evolution in the *D. yakuba* lineage. In this context, it would be intriguing to test for presence of lineage-specific clustering of the substitutions for these genes using outgroup sequences. When we have a large enough number of substitutions in each of the lineages compared, computer simulation analysis will provide a mean of statistical analysis of lineage-dependent clustering effects.

We now come to understand that, besides gene duplication and protein sequence variation, changes in tissue or developmental specificity and in expression levels have played crucial roles in the morphological and developmental evolution of organisms. It is certainly a necessary task to survey the kind of changes that have occurred during the course of evolution. Further DNA sequence analysis may detect even small changes in the action of natural selection, for instance, suggested by the codon usage change in the four *AS-C* genes with the redundant function. This information, in turn, may provide insights into evolutionary changes in expression levels and changes responsible for morphological evolution. Interspecific variation may also be useful as a source to study genetic interaction among sites in one molecule.

I am grateful to Tomoko Ohta, Masaru Iizuka and Yasuo Ina for their suggestion and encouragement. I thank Shigeo Hayashi for permission to use sequencing facilities in his laboratory. I also thank Yumiko Yamashita for her technical assistance with this project and Leah Gilner and two anonymous reviewers for improving the manuscript. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan.

LITERATURE CITED

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

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

Alonso, M. C., and C. V. Cabrera, 1988 The achaete-scute gene complex of *Drosophila melanogaster* comprises four homologous genes. EMBO J. **7:** 2585–2591.

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.

Balcells, L., J. Modolell and M. Ruiz-Gómez, 1988 A unitary basis for different *Hairy-wing* mutations of *Drosophila melanogaster*. EMBO J. **7:** 3899–3906.

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.

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., 1989 Estimating the variability of substitution rates. Genetics **123:** 615–619.

Caccone, A., E. N. Moriyama, J. M. Gleason, L. Nigro and J. R. Powell, 1996 A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism data. Mol. Biol. Evol. **13:** 1224–1232.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

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.

Eanes, W. F., M. Kirchner, J. Yoon, C. H. Biermann, I. Wang *et al.*, 1996 Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus in *Drosophila melanogaster* and *Drosophila simulans*. Genetics **144:** 1027–1041.

Gillespie, J. H., 1989 Lineage effects and the index of dispersion of molecular evolution. Mol. Biol. Evol. **6:** 636–647.

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

Gillespie, J. H., and C. H. Langley, 1979 Are evolutionary rates really variable? J. Mol. Evol. **13:** 27–34.

Goldman, N., 1994 Variance to mean ratio, $R(t)$, for Poisson processes on phylogenetic trees. Mol. Phylogenet. Evol. **3:** 230–239.

Hamblin, M. T., and C. F. Aquadro, 1996 High nucleotide sequence variation in a region of low recombination in *Drosophila simulans* is consistent with the background selection model. Mol. Biol. Evol. **13:** 1133–1140.

Hilton, H., R. M. Kliman and J. Hey, 1994 Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48:** 1900–1913.

Horton, I. H., 1939 A comparison of the salivary gland chromosomes of *Drosophila melanogaster* and *D. simulans*. *Genetics* **24:** 234–243.

Hudson, R. R., 1983 Testing the constant-rate neutral allele model with protein sequence data. Evolution **37:** 203–217.

Ina, Y., 1995 New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J. Mol. Evol. **40:** 190–226.

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.

Iwabe, N., K. Kuma and T. Miyata, 1996 Evolution of gene families and relationship with organismal evolution: rapid divergence of tissue-specific genes in the early evolution of chordates. Mol. Biol. Evol. **13:** 483–493.

Jiménez, F., and J. A. Campos-Ortega, 1990 Defective neuroblast commitment in mutants of the *achaete-scute* complex and adjacent genes of D. melanogaster. Neuron **5:** 81–89.

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

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.

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

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

Langley, C. H., and W. M. Fitch, 1974 An examination of the constancy of the rate of molecular evolution. J. Mol. Evol. **3:** 161–177.

Lemeunier, F., and M. Ashburner, 1976 Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. Proc. R. Soc. Lond. **193B:** 275–294.

Lewin, B., 1997 *Genes, VI*. Oxford University Press, Oxford.

Li, W.-H., 1985 Accelerated evolution following gene duplication and its implication for the neutralist-selectionist controversy, pp. 333–352 in *Population Genetics and Molecular Evolution*, edited by T. Ohta and K. Aoki. Japan Scientific Societies Press, Tokyo and Springer-Verlag, Berlin.

Martín-Campos, J. M., J. M. Comerón, N. Miyashita and M. Aguadé, 1992 Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics **130:** 805–816.

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

Miyata, T., and T. Yasunaga, 1980 Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J. Mol. Evol. **16:** 23–36.

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 J. R. Powell, 1996 Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13:** 261–277.

Nei, M., and F. Tajima, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97:** 145–163.

Nielsen, R., 1997 Robustness of the estimator of the index of dispersion for DNA sequences. Mol. Phylogenet. Evol. **7:** 346–351.

Ohnishi, S., and R. A. Voelker, 1979 Comparative studies of allozyme loci in *Drosophila simulans* and *D. melanogaster*. II. Gene arrangement on the third chromosome. Jpn. J. Genet. **54:** 203–209.

Ohta, T., 1994 Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics **138:** 1331–1337.

Ohta, T., 1995 Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40:** 56–63.

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.

Ohta, T., and M. Kimura, 1971 On the constancy of the evolutionary rate of cistrons. J. Mol. Evol. **1:** 18–25.

Orenic, T. V., D. C. Slusarski, K. L. Kroll and R. A. Holmgren, 1990 Cloning and characterization of the segment polarity gene *cubitus interruptus Dominant* of *Drosophila*. Genes Dev. **4:** 1053–1067.

Parkhurst, S. M., H. D. Lipshitz and D. Ish-Horowicz, 1993 *achaete-scute* feminizing activities and *Drosophila* sex determination. Development **117:** 737–749.

Parras, C., L. A. Garcia-Alonso, I. Rodríguez and F. Jiménez, 1996 Control of neural precursor specification by proneural proteins in the CNS of *Drosophila*. EMBO J. **15:** 6394–6399.

Rose, L. S., and E. Wieschaus, 1992 The Drosophila cellularization gene nullo produces a blastoderm-specific transcript whose levels respond to the nucleocytoplasmic ratio. Genes Dev. **6:** 1255–1268.

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

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.

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

Takahata, N., 1986 An attempt to estimate the effective size of the ancestral species common to two extant species from which homologous genes are sequenced. Genet. Res. **48:** 187–190.

Takahata, N., 1987 On the overdispersed molecular clock. Genetics **116:** 169–179.

Thackeray, J. R., and C. P. Kyriacou, 1990 Molecular evolution in the *Drosophila yakuba period* locus. J. Mol. Evol. **31:** 389–401.

True, J. R., J. M. Mercer and C. C. Laurie, 1996 Differences in crossover frequency and distribution among three sibling species of Drosophila. Genetics **142:** 507–523.

Villares, R., and C. V. Cabrera, 1987 The *achaete-scute* gene complex of *D. melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to *myc*. Cell **50:** 415–424.

Vincent, A., H. V. Colot and M. Rosbash, 1985 Sequence and structure of the *serendipity* locus of *Drosophila melanogaster*: a densely transcribed region including a blastoderm-specific gene. J. Mol. Biol. **186:** 149–166.

Watterson, G. A., and H. A. Guess, 1977 Is the most frequent allele the oldest? Theor. Popul. Biol. **11:** 141–160.

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

Wu, C.-I., and W.-H. Li, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741–1745.

Communicating editor: A. G. Clark