Rates of Synonymous Substitution and Base Composition of Nuclear Genes in *Drosophila*

Etsuko N. Moriyama and Takashi Gojobori

National Institute of Genetics, Mishima, 411 Japan

Manuscript received October 3, 1991

Accepted for publication January 4, 1992

ABSTRACT

We compared the rates of synonymous (silent) substitution among various genes in a number of species of *Drosophila*. First, we found that even for a particular gene, the rate of synonymous substitution varied considerably with *Drosophila* lineages. Second, we showed a large variation in synonymous substitution rates among nuclear genes in *Drosophila*. These rates of synonymous substitution were correlated negatively with C content and positively with A content at the third codon positions. Nucleotide sequences were also compared between pseudogenes and their functional homologs. The C content of the pseudogenes was lower than that of the functional genes and the A content of the former was higher than that of the latter. Because the synonymous substitution for functional genes and the nucleotide substitution for pseudogenes are exempted from any selective constraint at the protein level, these observations could be explained by a biased pattern of mutation in the *Drosophila* nuclear genome. Such a bias in the mutation pattern may affect the molecular clock (local clock) of each nuclear gene of each species. Finally, we obtained the average rates of synonymous substitution for three gene groups in *Drosophila*; 11.0×10^{-9} , 17.5×10^{-9} and 27.1×10^{-9} /site/year.

As the nucleotide sequence data for the *Drosophila* genome have accumulated, it has become apparent that most nuclear genes in *Drosophila* evolve at rates more than two times higher than those in rodents and other mammals (MORIYAMA 1987; CACONE, AMATO and POWELL 1988; SHARP and LI 1989; ROWAN and HUNT 1991). Moreover, the rate of synonymous substitution of *Drosophila* genes is known to vary between genes (SHIELDS *et al.* 1988; SHARP and LI 1989). The rate of nucleotide substitution can be affected by the mutation rate (KIMURA 1983), the generation time (or the number of replications), and the degree of selective constraints on the nucleotide sequence if the mutations are not selectively neutral (OHTA and TACHIDA 1990). One explanation for the higher rate of synonymous substitution in *Drosophila* is that it has a shorter generation time and, thus, more DNA replications *per year* than mammals (BRITTEN 1986; LI and TANIMURA 1987; KIMURA 1983, 1991). Differences in the accuracy of DNA replication or the efficiency of DNA repair are also possible explanations for the different rates of evolution between *Drosophila* and mammals. It is of interest that the characteristics of DNA polymerases in *Drosophila* may be different from those in mammals; DNA polymerase β in *Drosophila* is about twice as large as its mammalian counterpart and is detected only in the earlier stages of development (SAKAGUCHI and BOYD 1985), and the cryptic proofreading exonuclease activity of DNA polymerase α is unique for the embryo of *Drosophila* (WANG 1991).

SHIELDS *et al.* (1988) have proposed a measure of biased usage of synonymous codons, "scaled χ^2 ." Using this measure, they, as well as SHARP and LI (1989), have shown that the rate of synonymous substitution in *Drosophila* is negatively correlated with the degree of codon usage bias. They have concluded that the variation in substitution rates among nuclear genes in *Drosophila* results from differences in the strengths of selective constraints on synonymous codons. Moreover, SHIELDS *et al.* (1988) have pointed out that codon usage bias in the nuclear genome of *Drosophila* is associated with differences in the G + C content at synonymous sites. STARMER and SULLIVAN (1989) also found shifts in nucleotide frequencies at the third codon positions by comparing the alcohol dehydrogenase (*Adh*) genes among various species of the subgenera *Sophophora* and *Drosophila*. For mammalian genes, TICHER and GRAUR (1989) reported that AT-rich genes accumulate more synonymous substitutions than GC-rich genes, and thus variation in synonymous substitution rates and unequal usage of codons must be due to negative selection against nucleotides A and T in synonymous sites. Further, WOLFE, SHARP and LI (1989) showed a correlation between the rate of synonymous substitution and the G + C content in a comparison of the nuclear genes of rat and mouse. For rodents, the G + C content of the GC-rich genes correlates negatively with substitution rates. On the other hand, that of AT-rich genes correlates positively with substitution rates. They suggested that variation in both synonymous substitution

TABLE 1

A list of *Drosophila* species used in this study

Subgenus	Species group	Complex	Species
<i>Sophophora</i>	<i>melanogaster</i>	<i>melanogaster</i>	<i>D. melanogaster</i> ^a
			<i>D. simulans</i> ^a
			<i>D. mauritiana</i> ^a
			<i>D. sechellia</i> ^a
		<i>yakuba</i>	<i>D. orena</i> ^a
			<i>D. erecta</i>
			<i>D. yakuba</i> ^a
			<i>D. teissieri</i> ^a
			<i>D. pseudoobscura</i> ^a
		<i>obscura</i>	<i>D. subobscura</i>
			<i>D. miranda</i>
			<i>D. ambigua</i> ^a
			<i>D. picticornis</i> ^a
			<i>D. affinisdisjuncta</i> ^a
<i>Drosophila</i>	Hawaiian picture-wing		<i>D. grimshawi</i>
			<i>D. planitibia</i> ^a
			<i>D. heteroneura</i> ^a
			<i>D. differens</i> ^a
			<i>D. silvestris</i> ^a
	<i>repleta</i>	<i>mulleri</i>	<i>D. mulleri</i> ^a
			<i>D. mojavensis</i> ^a
			<i>D. navojoa</i> ^a
			<i>D. hydei</i> ^a
			<i>D. mettleri</i> ^a
	<i>virilis</i>	<i>virilis</i>	<i>D. virilis</i>
			<i>D. funebris</i>
			<i>D. funebris</i>
	<i>Scaptodrosophila</i>	<i>victoria</i>	<i>D. lebanonensis</i> ^a

^a Nucleotide sequences of *Adh* genes are available.

TABLE 2

A list of nuclear genes used in this study

Genes ^a	No. of codons compared	No. of species compared	Abbreviations used in Figure 3
<i>Adh</i>	255	17	A
<i>Amy</i>	493	2	m
<i>ArrB</i>	400	2	r
<i>bcd</i>	478	2	B
<i>Cp15</i>	111	7	5
<i>Cp19</i>	214	6	9
<i>Cp36</i>	232 ^b	2	6
<i>en</i>	530	2	E
<i>Est-6</i>	541	2	s
<i>Gart</i>	1352	2	G
<i>Gld</i>	611	2	l
<i>hsp82</i>	374	4	H
<i>hb</i>	747	2	u
<i>mam</i>	447 ^b	2	M
<i>Pcp</i>	183	2	P
<i>per</i>	870 ^b	4	e
<i>rp49</i>	133	2	R
<i>sev</i>	2484	2	S
<i>Sod</i>	152	3	o
<i>sal</i>	141	3	p
<i>β2 tubulin</i>	445	2	T
<i>Ubx</i>	247 ^b	3	U
<i>Xdh</i>	1297	2	X
3' orf of <i>Adh</i>	270	3	3

^a Gene names and references: *Adh* (see MATERIALS), *Amy* (α -Amylase, BOER and HICKEY 1986; BROWN, AQUADRO and ANDERSON 1990; HICKEY *et al.* 1991), *ArrB* (*Arrestin*, YAMADA *et al.* 1991; KRISHNAN and GANGULY 1990), *bcd* (*bicoid*, BERLETH *et al.* 1988; SEEGER and KAUFMAN 1990), *Cp15* (*Chorion protein 15*, MARTINEZ-CRUZADO *et al.* 1988; MARTINEZ-CRUZADO 1990), *Cp19* (*Chorion protein 19*, MARTINEZ-CRUZADO *et al.* 1988; MARTINEZ-CRUZADO 1990), *Cp36* (*Chorion protein 36*, SPRADLING *et al.* 1987; KONSOLAKI *et al.* 1990), *en* (*engrailed*, POOLE *et al.* 1985; KASSIS *et al.* 1986), *Est-6* (*Esterase-6*, COLLET *et al.* 1990; BRADY, RICHMOND and OAKSHOTT 1990), *Gart* (HENIKOFF and EGHTEDARZADEH 1987), *Gld* (*Glucose dehydrogenase*, KRASNEY, CARR and CAVENER 1990), *hsp82* (*heat shock protein 82*, BLACKMAN and MESELSON 1986), *hb* (*hunchback*, TAUTZ *et al.* 1987; TREIER, PFEIFLE and TAUTZ 1989), *mam* (*mastermind*, SMOLLER *et al.* 1990; NEWFELD, SMOLLER and YEDVOBNICK 1991), *Pcp* (*Pupal cuticle protein*, HENIKOFF *et al.* 1986; HENIKOFF and EGHTEDARZADEH 1987), *per* (*period*, CITRI *et al.* 1987; COLOT, HALL and ROSBASH 1988; THACKERAY and KYRIACOU 1990), *rp49* (*ribosomal protein 49*, O'CONNELL and ROSBASH 1984; AGUADE 1988), *sev* (*sevenless*, BOWTELL, SIMON and RUBIN 1988; MICHAEL, BOWTELL and RUBIN 1990), *Sod* (*Superoxide dismutase*, KWIATOWSKI, PATEL and AYALA 1989; KWIATOWSKI, GONZALEZ and AYALA 1989; KWIATOWSKI and AYALA 1989), *sal* (*spalt*, REUTER, SCHUH and JACKLE 1989), *β2 tubulin* (MICHIELS *et al.* 1987), *Ubx* (*Ultrabithorax*, WEINZIERL *et al.* 1987; WILDE and AKAM 1987), *Xdh* (*Xanthine dehydrogenase*, RILEY 1989).

^b Partial sequences were compared.

rate and base composition can be attributed to systematic differences in the rate and pattern of mutation over regions of the genome.

A consistent explanation for the high and variable synonymous substitution rates of the nuclear genes in *Drosophila* has not been provided until now. To elucidate the evolutionary factors determining substitution rates in *Drosophila*, we compared the rates of synonymous substitution among various genes in a number of *Drosophila* species. We found variation in the rates of synonymous substitution among evolutionary lineages of *Drosophila* as well as a large variation among the nuclear genes in *Drosophila*. The rates of synonymous substitution were shown to be correlated with the base composition of the nuclear genes. Furthermore, we compared the base composition of pseudogenes and their functional homologs. The results obtained imply that synonymous substitutions in the nuclear genome of *Drosophila* are affected by the biased pattern of mutations in this organism.

MATERIALS

***Drosophila* species and sequence data used:** Tables 1 and 2 list the *Drosophila* species and genes examined in this study. All the species belong to the genus *Drosophila*. The nucleotide sequences of 24 *Adh* genes have been reported for 21 species of *Drosophila*, which are composed of four

melanogaster-sibling species (BODMER and ASHBURNER 1984; COYNE and KREITMAN 1986) and five other species of the subgenus *Sophophora* (SCHAEFFER and AQUADRO 1987; JEFFS and ASHBURNER 1991; MARFANY and GONZALEZ-DUARTE 1991), 11 species of the subgenus *Drosophila* (FISHER and MANIATIS 1985; ATKINSON *et al.* 1988; ROWAN and DICKINSON 1988; WEAVER, ANDREWS and SULLIVAN 1989; MENTOTTI-RAYMOND, STARMER and SULLIVAN 1991; ROWAN and HUNT 1991; YUM, STARMER and SULLIVAN 1991), and one species of the subgenus *Scaptodrosophila* (VILLARROYA and JUAN 1991) (see Table 1). The *Adh* locus of each species in

FIGURE 1.—A phylogenetic tree for the *Adh* gene of *Drosophila* species. This tree was constructed by the neighbor-joining method (SAITOU and NEI 1987) using the number of synonymous substitutions per site (NEI and GOJOBORI 1986). The right column represents the G + C content (%) at the third codon positions for the *Adh* gene of each *Drosophila* lineage. Note that this tree is an unrooted tree and the evolutionary position of *D. lebanonensis* (in the subgenus *Scaptodrosophila*) is now controversial (VILLARROYA and JUAN 1991).

the *repleta* species group contains two functional genes, *Adh-1* and -2, and a pseudogene.

Pseudogenes in *Drosophila*: In *Drosophila*, nucleotide sequences for pseudogenes have been reported for the larval cuticle protein (*Lcp*) gene of *D. melanogaster* (SNYDER *et al.* 1982) and for the *Adh* genes of two species in the *melanogaster* species group (JEFFS and ASHBURNER 1991) as well as of four species in the *repleta* species group (FISHER and MANIATIS 1985; ATKINSON *et al.* 1988; MENOTTI-RAYMOND, STARMER and SULLIVAN, 1991; YUM, STARMER and SULLIVAN 1991).

RESULTS

Rates of synonymous substitution and G + C contents of various *Drosophila* species: We compared the nucleotide sequences of the coding regions for the *Adh* genes among 21 *Drosophila* species. Figure 1 shows a phylogenetic tree constructed by the neighbor-joining method (SAITOU and NEI 1987) using the number of synonymous substitutions per site (NEI and GOJOBORI 1986). The neighbor-joining method does not require strict constancy in evolutionary rates. Each branch length represents the number of synonymous substitutions per site. It is clear that in a pair of species each one does not always have a branch of equal length from the divergence point. This implies that the rate of synonymous substitution for the *Adh* gene might vary with the evolutionary lineages in *Drosophila*. In particular, the substitution rates seem to be different among clusters of the subgenus *Sophophora*, Hawaiian *Drosophila* and the *repleta* species group.

TABLE 3

Comparisons of relative rates and G + C content between *Drosophila* lineages

Species ^a							(G + C) % ^d		
A	B	C	D _{AB} ^b	D _{AC} ^b	D _{AB-DAC} ^c		A	B	C
<i>Adh</i>									
me	pi	mu	1.157	0.884	0.273		81.3	61.0	72.0
me	hy	mu	1.225	0.884	0.341*			63.8	
pi	me	ps	1.157	1.063	0.094				76.0
<i>Cp15</i>									
me	gr	vi	1.349	1.022	0.327		64.3	50.0	68.7
gr	me	su	1.349	1.293	0.056				59.5
<i>Cp19</i>									
gr	me	su	1.060	0.975	0.085		52.0	74.0	70.0
<i>hsp82</i>									
vi	me	ps	0.774	0.812	-0.038		71.6	77.8	76.2
<i>per</i>									
vi	me	ps	1.269	1.357	-0.088		69.5	84.2	75.8
<i>Ubx</i>									
fu	me	ps	0.854	0.867	-0.013		68.7	69.8	86.8

^a Species A was used as a reference and relative rates were compared between species B and C. Abbreviations for species are as follows: me for *D. melanogaster*, ps for *D. pseudoobscura*, su for *D. subobscura*, vi for *D. virilis*, fu for *D. funebris*, mu for *D. mulleri*, pi for *D. picticornis*, and gr for *D. grimshawi*. *Adh-1* genes were used for *Adh* of *D. mulleri* and *D. hydei*.

^b Number of synonymous substitutions per site between species A and B, or species A and C.

^c Relative rates were compared between lineages B and C using the method of Wu and Li (1985).

^d G + C % at the third codon positions of each species.

* Significant at the 5% level.

The G + C content at the third codon positions of the *Adh* gene also varies with lineage. The G + C content at the third codon positions of the *Adh* gene ranged from 59 to 87%. The content of G, C, A and T ranged from 25 to 32%, 33 to 55%, 4 to 15% and 9 to 29%, respectively. Since the range of C content is much larger than that of G content, the variation in G + C content may be caused mainly by the variation in C content. The subgenus *Sophophora* has a relatively high G + C content (76–87%). The subgenus *Drosophila* has a relatively low G + C content (59–74%). The Hawaiian species have the lowest value for G + C content (59–61%). The same features are also seen for other genes (the right half of Table 3). The G + C content for the subgenus *Drosophila* (the *repleta* species group and Hawaiian species) is lower than that for the subgenus *Sophophora*. Only for the *Cp15* gene, the G + C content of *D. melanogaster* (in the subgenus *Sophophora*) is lower than that of *D. virilis* (in the subgenus *Drosophila*).

The variation in synonymous substitution rates seems to be related to the differences in G + C content among species. In Table 3, we show the comparisons of synonymous substitution rates and G + C content among species for the *Adh* gene as well as several other genes. The "relative rate test" proposed by Wu

and LI (1985) is useful for comparing the substitution rates between a given pair of species (species B and C in Table 3). In this test, we require no knowledge of the divergence time between species, so that the number of synonymous substitutions per site between these species are compared with each other. The values of $D_{AB}-D_{AC}$ in Table 3 represent the difference between the number of synonymous substitutions per site for lineages B and C after divergence. If the value is positive, lineage C evolves at a relatively lower rate than lineage B. For the *Adh* gene, the substitution rate for *D. mulleri* is significantly lower than that for *D. hydei*. The rates for the *Adh* gene of *D. mulleri* and for the *Cp15* gene of *D. virilis* are also lower than those for *D. picticornis* and *D. grimshawi*, respectively. However, the rates for all the genes of *D. pseudoobscura* and *D. subobscura* are comparable to those for *D. melanogaster*. When we used the nucleotide sequences of *D. pseudoobscura* for these genes as references (species A in Table 3), the same results were obtained (data not shown). On the other hand, the G + C contents of these genes for *D. mulleri* and *D. virilis* are higher than those for *D. hydei* and the Hawaiian species (*D. picticornis* and *D. grimshawi*). Therefore, species with a higher G + C content seem to evolve with lower substitution rates. The same relationship was obtained for the *Adh* genes of other species (Figure 1).

The G + C content of the subgenus *Sophophora* is comparable to that of *D. mulleri*. However, we cannot compare the substitution rates between the subgenera *Sophophora* and *Drosophila*, since there are no out-group species available. If the rate of synonymous substitution is correlated with the G + C content as shown in the subgenus *Drosophila*, the species in the subgenus *Sophophora* would be expected to have substitution rates similar to that of *D. mulleri* and lower than those of other *repleta* species group or the Hawaiian species.

Rates of synonymous substitution and base compositions of the nuclear genes in *Drosophila*: To examine whether the variation in substitution rates among the genes is correlated with base composition, we compared the rates of synonymous substitution and base composition among nuclear genes in *Drosophila*. We estimated the number of synonymous substitutions per site for 16 genes compared between *D. melanogaster* and *D. pseudoobscura*. The number of synonymous substitutions per site ranged from 0.401 for the *Amy* gene to 1.984 for the *Est-6* gene (see Table 5). These values correlated negatively with G + C content at the third codon positions ($r = -0.87$, $P < 0.001$), as well as with the SHIELDS *et al.*'s values (1988) termed "scaled χ^2 " ($r = -0.92$, $P < 0.001$). Correlation coefficients (r) between the contents of C, G, A and T at the third codon positions and synony-

mous substitution rates for the nuclear genes were -0.82 ($P < 0.001$), 0.00 , 0.91 ($P < 0.001$) and 0.60 ($P < 0.05$), respectively. C content was strongly and negatively correlated with substitution rates (Figure 2A). However, G content did not correlate to substitution rates. Moreover, A content was strongly and positively correlated with substitution rates (Figure 2B).

We obtained the same relationships in comparisons between *D. melanogaster* and *D. simulans*, and between *D. melanogaster* and *D. yakuba* ($r = -0.71$, $P < 0.2$, and $r = -0.93$, $P < 0.1$, respectively; Figure 2, C and D), although only a few genes were available for these comparisons. The correlation in the comparison between *D. melanogaster* and *D. virilis* was very weak ($r = -0.20$, $P > 0.5$; Figure 2E). As mentioned before, base composition varies greatly between the species of the subgenera *Sophophora* and *Drosophila*. Therefore, this difference between lineages may have an effect on the correlation between base content and substitution rates. Furthermore, it is possible that the amino acid compositions are different between these two species. Such a difference may affect the base content at the third codon positions. To avoid the effect of differences in amino acid composition on base content, we computed each base content for the codons in the fourfold degeneracy group (Val, Pro, Thr, Ala, Gly, and four codons for Leu, Ser and Arg). For all comparisons including that between *D. melanogaster* and *D. virilis*, we obtained better correlations than those shown in Figure 2. From these results, it seems that the synonymous substitution rates of *Drosophila* nuclear genes are correlated negatively to the G + C or C content at the third codon positions, similarly to the case for variation of the rates among *Drosophila* lineages.

Base compositions of *Drosophila* pseudogenes: Nucleotide changes at the third codon positions of functional genes are subject to weak functional constraints since most of these changes do not affect the amino acid sequence. Nucleotide changes in pseudogenes are expected to be free from functional constraints than those at the third codon position of functional genes since pseudogenes have no function. Therefore, pseudogenes are useful for examining the base compositions of nucleotide sequences without functional constraints. We should note however that the base composition of a pseudogene may retain the features of the original functional gene if the time since becoming a pseudogene is not long.

Table 4 shows the base contents at the third codon positions of pseudogenes and their functional homologues of *Drosophila*. The G + C and C contents of the pseudogene for the *Lcp* gene (*Lcp1ψ*) are considerably lower than those of the functional *Lcp1* gene. The G + C content of *Lcp1* is almost equal to that of

FIGURE 2.—Relationships between base content and number of synonymous substitutions per site for the comparisons between *D. melanogaster* and *D. pseudoobscura* (A and B), *D. melanogaster* and *D. simulans* (C), *D. melanogaster* and species of the *yakuba* complex (D), and *D. melanogaster* and *D. virilis* (E). The nucleotide sequences for the *ArrB* gene of *D. miranda* and for the *Cp15*, *Cp19* and *rp49* genes of *D. subobscura* were used instead of those for *D. pseudoobscura*. The data for species of the *yakuba* complex include those of the *Adh* and *per* genes for *D. yakuba* (●), the *Amy* gene for *D. erecta* (▲), and the *sal* gene for *D. oreana* (■).

the *Ubx* gene (69.8%) and the *Pcp* gene (69.7%), each of which has a relatively low or medium rate of synonymous substitution (see Table 5 and Figure 3). On the other hand, the G + C content of the *Lcp1ψ* is comparable to that of the *Est-6* gene (51.5%), which is a gene with one of the highest synonymous substitution rates. The G + C content and C content for the *Adh* pseudogenes (*Adh-ψ*) are also lower than those of the functional *Adh* genes, although the differences are small. These observations, although there are only a few samples, are consistent with relationships be-

tween the synonymous substitution rates and the base contents.

Evolutionary rates of nuclear genes in *Drosophila*: Using divergence times for *Drosophila* species, we estimated the absolute rates of synonymous substitution for nuclear genes in *Drosophila* (Table 5 and Figure 3). Although the divergence times used here might contain considerable errors, there is a positive correlation between the number of synonymous substitutions per site and divergence time (Figure 3).

The absolute rates of synonymous substitution for

TABLE 4

Comparisons of base content at the third codon positions between pseudogenes and the functional homologs of *Drosophila*

Genes (species)	Base contents (%)				
	A	T	G	C	G + C
<i>Lcp1 (melanogaster)</i>	10.69	17.56	21.37	50.38	71.76
<i>Lcp1ψ (melanogaster)</i>	25.78	21.88	18.75	33.59	52.34
<i>Adh-1 (mulleri)</i>	9.45	18.50	28.74	43.31	72.05
<i>Adh-ψ (mulleri)</i>	16.25	19.17	27.08	37.50	64.58
<i>Adh-1 (mojavensis)</i>	10.63	20.87	27.95	40.55	68.50
<i>Adh-ψ (mojavensis)</i>	15.94	18.73	28.69	36.65	65.34
<i>Adh-1 (hydei)</i>	12.20	24.02	24.80	38.98	63.78
<i>Adh-ψ (hydei)</i>	22.44	26.38	22.05	29.13	51.18
<i>Adh (mettleri)</i>	11.42	23.62	26.77	38.19	64.96
<i>Adh-ψ (mettleri)</i>	15.29	23.53	27.06	34.12	61.18
<i>Adh (teissieri)</i>	3.91	9.38	32.03	54.69	86.72
<i>Adh-ψ (teissieri)</i>	4.69	10.94	34.38	50.00	84.38
<i>Adh (yakuba)</i>	4.30	12.11	32.03	51.56	83.59
<i>Adh-ψ (yakuba)</i>	7.03	12.89	31.25	48.83	80.08

the nuclear genes in *Drosophila* ranged from 6.7×10^{-9} (for the *Amy* gene) to 37.0×10^{-9} (for the *sal* gene)/site/year. The average rate of synonymous substitution for the nuclear genes in *Drosophila* was estimated to be 16.0×10^{-9} /site/year. This value is four times higher than the rate for artiodactyls (4.0×10^{-9} /site/year) and two times higher than that for rodents (7.9×10^{-9} /site/year; LI and TANIMURA 1987). In spite of the variation of the evolutionary rates, all the rates of synonymous substitution for the nuclear genes in *Drosophila* studied here are higher than the rate for artiodactyls. The rates for all genes, excluding the *Amy* and *Cp36* genes, are higher than the rate for rodents.

To obtain average rates of synonymous substitution (local clocks) for *Drosophila* nuclear genes, we classified all the genes used here into three groups based upon the degree of synonymous substitution rates (groups I, II, III in Table 5). Three average rates for *Drosophila* nuclear genes were estimated from regression lines in the three groups. They are 11.0×10^{-9} , 17.5×10^{-9} and 27.1×10^{-9} /site/year for groups I, II and III, respectively.

DISCUSSION

First, we compared the rates of synonymous substitution among *Drosophila* species. We found that synonymous substitution rates varied with *Drosophila* lineage. For the *Adh* and *Cp15* genes, the synonymous substitution rates for the Hawaiian species were much higher than those for *D. mulleri* and the subgenus *Sophophora*. Our results are opposite of those of GODDARD, CACCONE and POWELL (1990) and CACCONE and POWELL (1990). Using DNA-DNA hybridization,

they showed that *D. melanogaster* has a higher evolutionary rate than other *Drosophila* species. However, we estimated directly the number of synonymous substitutions and our comparisons of the substitution rates do not depend on uncertain divergence times between species (Table 3).

Next, we found a correlation between synonymous substitution rates and base content. The G + C and C content at the third codon positions of the nuclear genes are negatively correlated with synonymous substitution rates and the A content is positively correlated. Such correlations were shown both in comparisons of the substitution rates among the *Drosophila* species for a given gene and among various nuclear genes in *Drosophila*. The same relationships have been described in comparisons between human and rat genes (TICHER and GRAUR 1989) and for comparisons between GC-rich genes of rat and mouse (WOLFE, SHARP and LI 1989). Such consistency between the results obtained from *Drosophila* and mammalian genes suggests that the same or similar mechanism may determine the relationships between base composition and substitution rates for these organisms. Furthermore, we found that the C content in the nucleotide sequences of pseudogenes in *Drosophila* was lower than that in functional homologues. On the other hand, the A content of the former was higher than that of the latter.

We consider the following two possible explanations for the observed relationships between synonymous substitution rates and base composition. The first is that a biased pattern of spontaneous mutation in the *Drosophila* nuclear genome may cause the observed relationships. Spontaneous mutation directed from C to A or T, particularly to A, may frequently occur in the *Drosophila* nuclear genome. As a result, A and T have accumulated and C content has been reduced at the sites under low functional constraints. From the comparisons between pseudogenes and the functional homologues, we obtained results consistent with this explanation. At the sites with reduced functional constraint, more nucleotide substitutions, particularly those directed from C to A, may occur, and thus the substitution rate becomes higher, and at those sites nucleotides A would accumulate and C would reduce.

TICHER and GRAUR (1989) have stated that such a mutation bias can be ruled out because in mammalian genomes, there is a predominance of transitional mutations over transversional ones (BROWN *et al.* 1982). However, in the *Drosophila* nuclear genome, transitional and transversional substitutions seem to occur equally (E. N. MORIYAMA and T. GOJOBORI, in preparation). Therefore, it is possible that in *Drosophila*, mutations from C to A or T occur frequently, although it is still unclear why only the G content did not correlate with the substitution rates.

TABLE 5
Number of synonymous substitutions per site for *Drosophila* nuclear genes

Genes	No. of synonymous substitutions per site between ^a				
	<i>melanogaster</i> - its sibling species	<i>melanogaster</i> - <i>yakuba</i> complexes	<i>melanogaster</i> - <i>obscura</i> groups	<i>Drosophila</i> - <i>Sophophora</i> subgenera	<i>grimshawi</i> - <i>planitibia</i> subgroups
Group I					
<i>Amy</i>		0.143	0.401 (6.7)		
<i>Cp36</i>				0.568 (7.1)	
<i>rp49</i>			0.585 (9.8)		
<i>Adh</i>	0.045 (9.0) ^b	0.152	0.611 (10.2)	1.096 (13.7)	0.125 (14.2)
<i>hsp82</i>	0.057 (11.4)		0.621 (10.4)	0.791 (9.9)	
<i>Sod</i>	0.109 (21.8)			0.847 (10.6)	
<i>Ubx</i>			0.719 (12.0)	0.861 (10.8)	
<i>ArrB</i>			0.724 (12.1)		
<i>Cp19</i>			0.806 (13.4)	1.048 (13.1)	0.116 (13.2)
<i>mam</i>				1.052 (13.2)	
Group II					
<i>en</i>				1.197 (15.0)	
<i>Cp15</i>			0.955 (15.9)	1.363 (17.0)	0.099 (11.3)
<i>hb</i>				1.281 (16.0)	
<i>per</i>		0.282	0.988 (16.5)	1.282 (16.0)	
<i>Xdh</i>			1.162 (19.4)		
<i>Pcp</i>			1.166 (19.4)		
<i>sev</i>				1.580 (19.8)	
<i>bcd</i>			1.206 (20.1)		
<i>Gld</i>			1.221 (20.4)		
Group III					
<i>Gart</i>			1.483 (24.7)		
<i>β2 tubulin</i>				2.047 (25.6)	
3' orf of <i>Adh</i>			1.580 (26.3)		
<i>Est-6</i>			1.984 (33.1)		
<i>sal</i>	0.185 (37.0)	0.525			

^a Values represent the average number of synonymous substitutions per site between each pair of species.

^b Values in parentheses represent the rate of synonymous substitution ($\times 10^{-9}$ /site/year) calculated by using the divergence times as follows: 2.5 million years (between *D. melanogaster* and its sibling species), 30 million years (between the *melanogaster* and *obscura* groups), 40 million years (between the subgenera *Drosophila* and *Sophophora*), 4.4 million years (between the *grimshawi* and *planitibia* subgroups). The divergence times were obtained from THROCKMORTON (1975), LACHAISE *et al.* (1988), MORIYAMA (1987) and ROWAN and HUNT (1991).

The second possible explanation is that mutations which cause an accumulation of A, and possibly T too, in the coding regions of *Drosophila* nuclear genes may be subjected to some kind of selection. Such selection would lead to reduce A and T and to accumulate C at the synonymous sites of the genes. Consequently, the gene under higher selective constraints may have a more biased codon usage, as described by SHARP and LI (1989). However, it is again unclear why only G content does not increase or decrease at the synonymous sites according to the degree of selective constraints.

Alternatively, there is a possibility that the biased codon usage in *Drosophila* is a result of selection at the translational level. tRNA abundance is one factor affecting codon usage. However, we cannot test this possibility because of a lack of data for tRNA abundance in *Drosophila*. GROSJEAN and FIERs (1982) proposed the optimal codon-anticodon interaction energy

hypothesis. They suggested the following two rules; (1) codons in which the first two nucleotides are T and/or A stabilize their interaction with anticodons if a C is used in the wobble position, and conversely, (2) if the first two nucleotides of the codon are C and/or G, a preference would exist for T as the third letter to avoid too sticky a codon-anticodon interaction. The nuclear genes in *Drosophila* follow only the first rule. In *Drosophila*, codons ending in C are preferred most even if the first two nucleotides are C and/or G (Table 6). Therefore, our findings that substitution rates varied considerably among *Drosophila* lineages and that the same tendency was shown for several nuclear genes may not be explained only by selective constraints at the translational level of each gene.

For rodents, WOLFE, SHARP and LI (1989) have shown that genes with intermediate G + C content have the highest substitution rates and the substitution rates of GC-rich and AT-rich genes are correlated

FIGURE 3.—Relationships between the number of synonymous substitutions per site and divergence time. The regression lines for the genes in groups I and III are represented. See Table 2 for the abbreviations of genes. The synonymous substitution rates for rodents and artiodactyls were taken from LI and TANIMURA (1987).

negatively and positively with G + C content, respectively. They proposed that these differences arise because mutation patterns vary with the timing of the replication of different chromosomal regions in the germline. Their hypothesis is related to the existence of isochores in mammalian genomes. On the other hand, substitution rates in *Drosophila* nuclear genes were correlated simply and negatively with G + C or C content. This simplicity may be related to the fact that there is no distinct isochore structure in the *Drosophila* genome and, as mentioned before, DNA polymerases of this organism have characteristics different from those of mammals. The relationships between substitution rates and base composition in the *Drosophila* nuclear genome may represent only one aspect of the relationships in the mammalian gene, namely that in the GC-rich gene region in mammalian genomes.

To elucidate the mechanisms governing evolutionary rates and base composition in *Drosophila* and other organisms, we should proceed to more detailed analyses on the pattern and/or direction of mutations in the *Drosophila* genome. We need to compare the pattern of mutations in *Drosophila* with those in other organisms (rodents, primates, or other invertebrates). In *Drosophila*, transversal substitutions, probably substitutions between C and A, seem to occur more

TABLE 6

Codon usage for the nuclear genes of *D. melanogaster*

Codons	Frequency (%) ^a		Codons	Frequency (%) ^a	
	Genes with higher C% ^b	Genes with lower C% ^b		Genes with higher C% ^b	Genes with lower C% ^b
TTT (Phe)	0.31	1.38	CCT (Pro)	0.32	0.79
TTC (Phe)	2.48	2.62	CCC (Pro)	3.32	2.15
			CCA (Pro)	1.02	1.04
TTA (Leu)	0.04	0.28	CCG (Pro)	1.06	1.98
TTG (Leu)	0.59	1.53			
			GCT (Ala)	1.97	1.52
ATT (Ile)	1.41	1.53	GCC (Ala)	6.04	3.46
ATC (Ile)	3.38	1.84	GCA (Ala)	0.71	1.10
ATA (Ile)	0.16	0.91	GCG (Ala)	1.08	1.16
TAT (Tyr)	0.82	1.27	CGT (Arg)	1.19	0.71
TAC (Tyr)	3.83	2.11	CGC (Arg)	2.66	1.25
			CGA (Arg)	0.32	0.79
AAT (Asn)	1.38	2.23	CGG (Arg)	0.16	0.71
AAC (Asn)	4.24	2.57			
			GGT (Gly)	2.85	2.09
AAA (Lys)	0.38	0.90	GGC (Gly)	5.19	3.58
AAG (Lys)	5.13	3.53	GGA (Gly)	2.63	3.63
			GGG (Gly)	0.19	0.58

^a Frequency of each codon for all codons used.

^b Nuclear genes of *D. melanogaster* with C contents more than or equal to 40% (higher C%), and less than 40% (lower C%), respectively.

frequently than in mammals, as shown here. Such a difference would affect the mutation rate in *Drosophila* and, thus, this could be one cause for the higher evolutionary rates in *Drosophila* than in mammals.

We are grateful to M. KIMURA and T. IKEMURA, National Institute of Genetics at Mishima, for their valuable comments. This study was supported by a grant-in-aid from the Ministry of Education, Science and Culture, Japan.

LITERATURE CITED

- AGUADE, M., 1988 Nucleotide sequence comparison of the *rp49* gene region between *Drosophila subobscura* and *D. melanogaster*. *Mol. Biol. Evol.* **5**: 433-441.
- ATKINSON, P. W., L. E. MILLS, W. T. STARMER and D. T. SULLIVAN, 1988 Structure and evolution of the *Adh* genes of *Drosophila mojavensis*. *Genetics* **120**: 713-723.
- BERLETH, T., M. BURRI, G. THOMA, D. BOPP, S. RICHSTEIN, G. FRIGERIO, M. NOLL and C. NUSSLEIN-VOLHARD, 1988 The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* **7**: 1749-1756.
- BLACKMAN, R. K., and M. MESESON, 1986 Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila hsp82* gene. *J. Mol. Biol.* **188**: 499-515.
- BODMER, M., and M. ASHBURNER, 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature* **309**: 425-430.
- BOER, P. H., and D. A. HICKEY, 1986 The alpha-amylase gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motif. *Nucleic Acids Res.* **14**: 8399-8411.
- BOWTELL, D. D. L., M. A. SIMON and G. M. RUBIN, 1988 Nucleotide sequence and structure of the *sevenless* gene of *Drosophila melanogaster*. *Genes Dev.* **2**: 620-634.

- BRADY, J. P., R. C. RICHMOND and J. G. OAKESHOTT, 1990 Cloning of the *esterase-5* locus from *Drosophila pseudoobscura* and comparison with its homologue in *D. melanogaster*. *Mol. Biol. Evol.* **7**: 525–546.
- BRITTEN, R. J., 1986 Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**: 1393–1398.
- 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.
- BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* **18**: 225–239.
- CACCONI, A., G. D. AMATO and J. R. POWELL, 1988 Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. *Genetics* **118**: 671–683.
- CACCONI, A., and J. R. POWELL, 1990 Extreme rates and heterogeneity in insect DNA evolution. *J. Mol. Evol.* **30**: 273–280.
- CITRI, Y., H. V. COLOT, A. C. JACQUIER, Q. YU, J. C. HALL, D. BALTIMORE and M. ROSBASH, 1987 A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. *Nature* **326**: 42–47.
- COLLET, C., K. M. NIELSEN, R. J. RUSSELL, M. KARL, J. G. OAKESHOTT and R. C. RICHMOND, 1990 Molecular analysis of duplicated esterase genes in *Drosophila melanogaster*. *Mol. Biol. Evol.* **7**: 9–28.
- COLOT, H. V., J. C. HALL and M. ROSBASH, 1988 Interspecific comparison of the *period* gene of *Drosophila* reveals large blocks of non-conserved coding DNA. *EMBO J.* **7**: 3929–3937.
- COYNE, J. A., and M. KREITMAN, 1986 Evolutionary genetics of two sibling species, *Drosophila simulans* and *D. sechellia*. *Evolution* **40**: 673–691.
- FISHER, J. A., and T. MANIATIS, 1985 Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. *Nucleic Acids Res.* **13**: 6899–6917.
- GODDARD, K., A. CACCONI and J. R. POWELL, 1990 Evolutionary implications of DNA divergence in the *Drosophila obscura* group. *Evolution* **44**: 156–1670.
- GROSJEAN, H., and W. FIERS, 1982 Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* **18**: 199–209.
- HENIKOFF, S., and M. K. EGHTEADARZADEH, 1987 Conserved arrangement of nested genes at the *Drosophila Gart* locus. *Genetics* **117**: 711–725.
- HENIKOFF, S., M. A. KEENE, K. FECHTEL and J. W. FRISTROM, 1986 Gene within a gene: nested *Drosophila* genes encode unrelated proteins on opposite DNA strands. *Cell* **44**: 33–42.
- HICKEY, D. A., L. BALLY-CAUF, S. ABUKASHAWA, V. PAYANT and B. F. BENKEL, 1991 Conserved evolution of duplicated protein-coding genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **88**: 1611–1615.
- JEFFS, P., and M. ASHBURNER, 1991 Processed pseudogenes in *Drosophila*. *Proc. R. Soc. Lond. Ser. B* **244**: 151–159.
- KASSIS, J. A., S. J. POOLE, D. K. WRIGHT and P. H. O'FARRELL, 1986 Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. *EMBO J.* **5**: 3583–3589.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- KIMURA, M., 1991 The neutral theory of molecular evolution: a review of recent evidence. *Jpn. J. Genet.* **66**: 367–386.
- KONSOLAKI, M., K. KOMITOPOULOU, P. P. TOLIAS, D. L. KING, C. SWIMMER and F. C. KAFATOS, 1990 The chorion genes of the medfly, *Ceratitis capitata*. I. Structural and regulatory conservation of the *s36* gene relative to two *Drosophila* species. *Nucleic Acids Res.* **18**: 1731–1737.
- KRASNEY, P. A., C. CARR and D. R. CAVENER, 1990 Evolution of the glucose dehydrogenase gene in *Drosophila*. *Mol. Biol. Evol.* **7**: 155–177.
- KRISHNAN, R., and R. GANGULY, 1990 Nucleotide sequence of the arrestin-like 49 Kd protein gene of *Drosophila miranda*. *Nucleic Acids Res.* **18**: 5894.
- KWIATOWSKI, J., and F. J. AYALA, 1989 *Drosophila virilis* Cu-Zn superoxide dismutase gene sequence. *Nucleic Acids Res.* **17**: 2133.
- KWIATOWSKI, J., F. GONZALEZ and F. J. AYALA, 1989 *Drosophila simulans* Cu-Zn superoxide dismutase gene sequence. *Nucleic Acids Res.* **17**: 6735.
- KWIATOWSKI, J., M. PATEL and F. J. AYALA, 1989 *Drosophila melanogaster* Cu, Zn superoxide dismutase gene sequence. *Nucleic Acids Res.* **17**: 1264.
- LACHAISE, D., M.-L. CARIU, J. R. DAVID, F. LEMEUNIER, L. TSACAS and M. ASHBURNER, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol. Biol.* **22**: 159–225.
- LI, W.-H., and M. TANIMURA, 1987 The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326**: 93–96.
- MARFANY, G., and R. GONZALEZ-DUARTE, 1991 The *Adh* genomic region of *Drosophila ambigua*: evolutionary trends in different species. *J. Mol. Evol.* **32**: 454–462.
- MARTINEZ-CRUZADO, J. C., 1990 Evolution of the autosomal chorion cluster in *Drosophila*. IV. The Hawaiian *Drosophila*: rapid protein evolution and constancy in the rate of DNA divergence. *J. Mol. Evol.* **31**: 402–423.
- MARTINEZ-CRUZADO, J. C., C. SWIMMER, M. G. FENERJIAN and F. C. KAFATOS, 1988 Evolution of the autosomal chorion cluster in *Drosophila*. I. General organization of the locus and sequence comparisons of genes *s15* and *s19* in evolutionarily distant species. *Genetics* **119**: 663–677.
- MENOTTI-RAYMOND, M., W. T. STARMER and D. T. SULLIVAN, 1991 Characterization of the structure and evolution of the *Adh* region of *Drosophila hydei*. *Genetics* **127**: 355–366.
- MICHAEL, W. M., D. D. L. BOWTELL and G. M. RUBIN, 1990 Comparison of the sevenless genes of *Drosophila virilis* and *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **87**: 5351–5353.
- MICHIELS, F., D. FALKENBURG, A. M. MULLER, U. HINZ, U. OTTO, R. BELLMANN, K. H. GLATZER, R. BRAND, S. BIALOJAN and R. RENKAWITZ-POHL, 1987 Testis-specific $\beta 2$ tubulins are identical in *Drosophila melanogaster* and *D. hydei* but differ from the ubiquitous $\beta 1$ tubulin. *Chromosoma* **95**: 387–395.
- MORIYAMA, E. N., 1987 Higher rates of nucleotide substitution in *Drosophila* than in mammals. *Jpn. J. Genet.* **62**: 139–147.
- NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- NEWFIELD, S. J., D. A. SMOLLER and B. YEDVOBNICK, 1991 Interspecific comparison of the unusually repetitive *Drosophila* locus *mastermind*. *J. Mol. Evol.* **32**: 415–420.
- O'CONNELL, P., and M. ROSBASH, 1984 Sequence, structure, and codon preference of the *Drosophila* ribosomal protein 49 gene. *Nucleic Acids Res.* **12**: 5495–5513.
- OHTA, T., and H. TACHIDA, 1990 Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* **126**: 219–229.
- POOLE, S. J., L. M. KAUFAR, B. DRESS and T. KORNBERG, 1985 The *engrailed* locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell* **40**: 37–43.
- POWELL, J. R., A. CACCONI, G. D. AMATO and C. YOON, 1986 Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. *Proc. Natl. Acad. Sci. USA* **83**: 9090–9093.
- REUTER, D., R. SCHUH and H. JACKLE, 1989 The homeotic gene

- spalt (*sal*) evolved during *Drosophila* speciation. *Proc. Natl. Acad. Sci. USA* **86**: 5483–5486.
- RILEY, M. A., 1989 Nucleotide sequence of the *Xdh* region in *Drosophila pseudoobscura* and an analysis of the evolution of synonymous codons. *Mol. Biol. Evol.* **6**: 33–52.
- ROWAN, R. G., and W. J. DICKINSON, 1988 Nucleotide sequence of the genomic region encoding alcohol dehydrogenase in *Drosophila affinisdisjuncta*. *J. Mol. Evol.* **28**: 43–54.
- ROWAN, R. G., and J. A. HUNT, 1991 Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of Hawaiian *Drosophila*. *Mol. Biol. Evol.* **8**: 49–70.
- SAKAGUCHI, K., and J. B. BOYD, 1985 Purification and characterization of a DNA polymerase β from *Drosophila*. *J. Biol. Chem.* **260**: 10406–10411.
- SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- SCHAEFFER, S. W., and C. F. AQUADRO, 1987 Nucleotide sequence of the *Adh* gene region of *Drosophila pseudoobscura*: evolutionary change and evidence for an ancient gene duplication. *Genetics* **117**: 61–73.
- SEGER, M. A., and T. C. KAUFMAN, 1990 Molecular analysis of the *bicoid* gene from *Drosophila pseudoobscura*: identification of conserved domains within coding and noncoding regions of the *bicoid* mRNA. *EMBO J.* **9**: 2977–2987.
- SHARP, P. M., and W.-H. LI, 1989 On the rate of DNA sequence evolution in *Drosophila*. *J. Mol. Evol.* **28**: 398–402.
- SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS and F. WRIGHT, 1988 "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5**: 704–716.
- SMOLLER, D., C. FRIEDEL, A. SCHMID, D. BETTLER, L. LAM and B. YEDVOBNICK, 1990 The *Drosophila* neurogenic locus *mastermind* encodes a nuclear protein unusually rich in amino acid homopolymers. *Genes Dev.* **4**: 1688–1700.
- SNYDER, M., M. HUNKAPILLER, D. YUEN, D. SILVERT, J. FRISTROM and N. DAVIDSON, 1982 Cuticle protein genes of *Drosophila*: structure, organization and evolution of four clustered genes. *Cell* **29**: 1027–1040.
- SPRADLING, A. C., D. V. DE CICCO, B. T. WAKIMOTO, J. F. LEVINE, L. J. KALFAYAN and L. COOLEY, 1987 Amplification of the X-linked *Drosophila* chorion gene cluster requires a region upstream from the s38 chorion gene. *EMBO J.* **6**: 1045–1053.
- STARMER, W. T., and D. T. SULLIVAN, 1989 A shift in the third-codon-position nucleotide frequency in alcohol dehydrogenase in the genus *Drosophila*. *Mol. Biol. Evol.* **6**: 546–552.
- TAUTZ, D., R. LEHMANN, H. SCHNURCH, R. SCHUH, E. SEIFERT, A. KIENLIN, K. JONES and H. JACKLE, 1987 Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. *Nature* **327**: 383–389.
- THACKERAY, J. R., and C. P. KYRIACOU, 1990 Molecular evolution in the *Drosophila yakuba period* locus. *J. Mol. Evol.* **31**: 389–401.
- THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of *Drosophila*, pp. 421–469 in *Handbook of Genetics*, Vol. III, edited by R. KING. Plenum Press, New York.
- TICHER, A., and D. GRAUR, 1989 Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. *J. Mol. Evol.* **28**: 286–298.
- TREIER, M., C. PFEIFLE and D. TAUTZ, 1989 Comparison of the gap segmentation gene *hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. *EMBO J.* **8**: 1517–1525.
- VILLARROYA, A., and E. JUAN, 1991 ADH and phylogenetic relationships of *Drosophila lebanonensis* *Scaptodrosophila*. *J. Mol. Evol.* **32**: 421–428.
- WANG, T. S.-F., 1991 Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **60**: 513–552.
- WEAVER, J., J. M. ANDREWS and D. T. SULLIVAN, 1989 Nucleotide sequence of the *Adh-1* gene of *Drosophila navojoa*. *Nucleic Acids Res.* **18**: 7524.
- WEINZIERL, R., J. M. AXTON, A. GHYSEN and M. AKAM, 1987 *Ultrabithorax* mutations in constant and variable regions of the protein coding sequence. *Genes Dev.* **1**: 386–397.
- WILDE, C. D., and M. AKAM, 1987 Conserved sequence elements in the 5' region of the *Ultrabithorax* transcription unit. *EMBO J.* **6**: 1393–1401.
- WOLFE, K. H., P. M. SHARP and W.-H. LI, 1989 Mutation rates differ among regions of the mammalian genome. *Nature* **337**: 283–285.
- WU, C.-L., 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.
- YAMADA, T., Y. TAKEUCHI, N. KOMORI, H. KOBAYASHI, Y. SAKAI, Y. HOTTA and H. MATSUMOTO, 1991 A 49-kilodalton phosphoprotein in the *Drosophila* photoreceptor is an arrestin homologue. *GenBank* **70**.
- YUM, J., W. T. STARMER and D. T. SULLIVAN, 1991 The structure of the *Adh* locus of *Drosophila mettleri*: an intermediate in the evolution of the *Adh* locus in the repleta group of *Drosophila*. *Mol. Biol. Evol.* **8**: 857–867.

Communicating editor: W.-H. LI