Synonymous Rates at the *RpII215* Gene of Drosophila: Variation Among Species and Across the Coding Region

Ana Llopart and Montserrat Aguadé

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, 08071 Barcelona, Spain

Manuscript received September 29, 1998
Accepted for publication January 21, 1999

ABSTRACT

The region encompassing the *RpII215* gene that encodes the largest component of the RNA polymerase II complex (1889 amino acids) has been sequenced in *Drosophila subobscura*, *D. madeirensis*, *D. guanche*, and *D. pseudoobscura*. Nonsynonymous divergence estimates (K_a) indicate that this gene has a very low rate of amino acid replacements. Given its low K_a and constitutive expression, synonymous substitution rates are, however, unexpectedly high. Sequence comparisons have allowed the molecular clock hypothesis to be tested. *D. guanche* is an insular species and it is therefore expected to have a reduced effective size relative to *D. subobscura*. The significantly higher rate of synonymous substitutions detected in the *D. guanche* lineage could be explained if synonymous mutations behave as nearly neutral. Significant departure from the molecular clock hypothesis for synonymous and nonsynonymous substitutions was detected when comparing the *D. subobscura*, *D. pseudoobscura*, and *D. melanogaster* lineages. Codon bias and synonymous divergence between *D. subobscura* and *D. melanogaster* were negatively correlated across the *RpII215* coding region, which indicates that selection coefficients for synonymous mutations vary across the gene. The C-terminal domain (CTD) of the RpII215 protein is structurally and functionally differentiated from the rest of the protein. Synonymous substitution rates were significantly different in both regions, which strongly indicates that synonymous mutations in the CTD and in the non-CTD regions are under detectably different selection coefficients.

SYNONYMOUS mutations have been classically considered to behave near neutrality (Kimura 1983) because they do not contribute to variation of the primary structure of proteins. According to Ohta's nearly neutral theory (Ohta and Kimura 1971; Ohta 1972), mutations that comply $|N_e s| \approx 1$, where N_e is the effective population size and s is the selection coefficient, should be defined as nearly neutral. Contrarily, strictly neutral mutations should satisfy $|N_e s| \ll 1$ for any effective population size. Under a strictly neutral model (Kimura 1968, 1983; Kimura and Ohta 1971) the rate of substitutions per year (K_y) is equal to the neutral mutation rate per generation (μ_g) divided by the generation time (g): $K_y = \mu_g / g$. Then, assuming constant neutral mutation rate per generation, strictly neutral mutations should exhibit generation-time effects (Ohta and Kimura 1971; Gillespie 1991). For slightly deleterious mutations $K_y = \mu_g / 4N_e sg$ (Kimura and Ohta 1971), and consequently rate constancy will be achieved only if there is a negative correlation between N_e and g. Such a correlation has been reported by Chao and Carr (1993) for highly diverged species. A direct prediction from the nearly neutral theory is therefore that synonymous substitution rates among different lineages of closely related Drosophila species with equal generation times will depend on the effective population size. Also, testing the molecular clock hypothesis (Zuckerkandl and Pauling 1965) by comparing the synonymous rates among different lineages can shed some light on the magnitude of the selection coefficients of mutations at synonymous sites in Drosophila.

Synonymous substitution rates vary extensively among different genes in Drosophila, and this variation is negatively correlated with codon usage bias (Shields *et al.* 1988; Sharp and Li 1989). Kliman and Hey (1994) detected a small but significant correlation between the base composition of introns and codon bias among different loci of *Drosophila melanogaster*, which indicates a residual effect of mutational processes on base composition at synonymous sites. In Drosophila, variation in the strength of natural selection acting on synonymous mutations, usually related to the expression level, has been proposed to explain the observed pattern of variation of codon bias among genes (Shields *et al.* 1988; Moriyama and Hartl 1993; Kliman and Hey 1994). Indeed, selective constraints on synonymous sites to ensure amino acid incorporation accuracy and/or to enhance elongation rates in the translation process (Kurland 1987a,b; Precup and Parker 1987; Bulmer 1991; Akashi 1994; Comeron and Kreitman 1998) have also been proposed to modulate the codon bias of a particular gene. The secondary structure of mRNA could also have some effect on Drosophila synonymous

Corresponding author: Ana Llopart, Departament Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, 08071 Barcelona, Spain. E-mail: llopart@porthos.bio.ub.es

substitutions (Comeron and Aguadé 1996) as has been suggested in enterobacteria (Lawrence *et al.* 1991; Eyre-Walker and Bulmer 1993). Recently, Comeron *et al.* (1999) have proposed that both the recombinational environment and the length of the coding region also contribute significantly to synonymous divergence and codon bias of a particular gene in Drosophila.

The RNA polymerase II complex is responsible for transcription of protein-encoding genes (McKnight and Yamamoto 1992). The locus that encodes its largest subunit has been sequenced in several eukaryotic organisms. Extensive homology between prokaryotic and eukaryotic RNA polymerases II has been reported (Allison *et al.* 1985). The largest subunit of the RNA polymerase II complex has an unusual C-terminal domain (CTD) that consists of a motif of seven amino acids tandemly repeated (Corden 1990). The consensus sequence of the repeat is Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The number of repeats varies among species, but the sequence characteristics are highly conserved among eukaryotes. Its function is considered essential for the viability of a wide range of eukaryotes (Nonet *et al.* 1987; Allison *et al.* 1988), and it has been shown recently that it plays an important role in mRNA capping (McCracken *et al.* 1997).

In Drosophila the largest component of the RNA pol II complex is a 215-kD peptide that is encoded by the *RpII215* gene (Jokerst *et al.* 1989). This gene is supposed to be ubiquitously expressed in cells because of its critical role in the transcriptional process, which, on the other hand, leads to the prediction that it will be a very conserved gene at the amino acid level. The expression pattern and the expected amino acid constraints would support the *a priori* idea that the *RpII215* gene should exhibit a high codon bias. However, as predicted by Li (1987) for the case of absolute linkage among sites, smaller selection coefficients on individual synonymous mutations would be expected from its very long coding region (1889 codons in *D. melanogaster*; Comeron *et al.* 1999). These features make the *RpII215* gene an ideal candidate for the study of different aspects of synonymous substitution rates. In this sense, the comparison between the palearctic species *D. subobscura* and the insular species *D. madeirensis* and *D. guanche* provides an excellent opportunity to test the effects on synonymous rates of the expected reduced N_e of the latter species by contrasting the molecular clock hypothesis. On the other hand, we have addressed the question of whether the proposed reduction of N_e in *D. melanogaster* as compared to *D. simulans* (Akashi 1995, 1996) can also be detected in the *RpII215* gene when *D. melanogaster* is compared to *D. subobscura* and *D. pseudoobscura*. Moreover, the long coding region of the *RpII215* gene has allowed us to study the pattern of variation of codon bias and synonymous divergence across the gene, where possible sources of variation among genes (like mutational pattern, recombinational environment, or expression levels) can be overruled. Finally, the presence of the CTD region in this gene provides a unique chance to explore possible natural selection fingerprints in two regions of a gene that are structurally and functionally differentiated (Allison *et al.* 1988).

MATERIALS AND METHODS

DNA preparation, cloning, and sequencing strategy: Two recombinant phages were isolated from a random genomic library of *D. subobscura* (λ*sub*Ra111) from Raíces (Canary Islands). Two different sets of probes from *D. melanogaster* were used separately in the screening (Figure 1): (i) a 1.3-kb *Eco*RI-*Pst*I and a 1.2-kb *Pst*I-*Eco*RI fragment (isolated from a recombinant plasmid kindly provided by A. Greenleaf) that included the entire second exon of the *RpII215* gene; (ii) a 3.6-kb *Eco*RI-*Eco*RI fragment that covered the third and fourth exons of the gene. After Southern blot analysis of the two positive phages, three DNA fragments (a 3.4-kb *Eco*RI vector and a 4.8-kb *Eco*RI-*Eco*RI from one of the phages and a 5.2-kb *Sal*I-*Sal*I from the second phage) were subcloned in pBluescriptII vectors. A set of nested deletions was obtained for each orientation of each subclone (Henikoff 1984). The sequence of both strands of a 7.8-kb region was obtained by manually sequencing each subclone, entirely or partially, using double-stranded DNA and the dideoxy chain termination method (Sanger *et al.* 1977).

For *D. madeirensis* and *D. guanche*, genomic DNA was extracted from adult flies of isofemale lines using a CsCl gradient (Bingham *et al.* 1981) and a standard small-scale method (Ashburner 1989) with minor modifications, respectively. After digestion of the DNA with a set of restriction enzymes and Southern blot transfer, two different probes that encompassed the 5′ and 3′ ends of the *RpII215* gene of *D. subobscura* were used separately to perform hybridization: (i) a 0.7-kb *Cla*I-digested PCR fragment that included 0.6 kb of the noncoding 5′ region, the first exon, and a small part of the first intron; (ii) a 0.6-kb *Sal*I-vector fragment within the CTD. In both species, digestion with *Hin*dIII produced a single band that showed cross-hybridization with both probes, and this enzyme was then used to construct the corresponding libraries with the λDASH vector (Stratagene, La Jolla, CA). A 1.8-kb

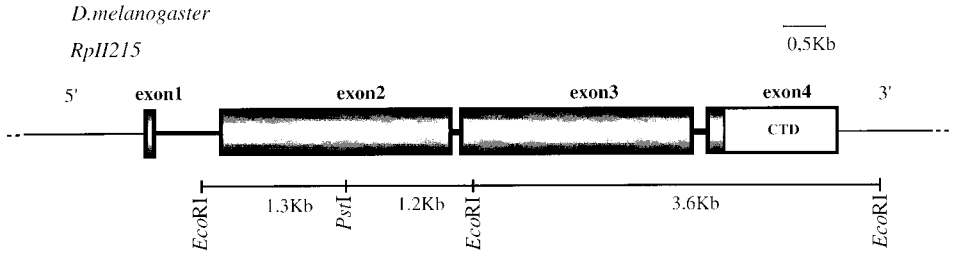

Figure 1.—Structure of the *RpII215* gene in *D. melanogaster*. Exons are shown as gray boxes. Black lines represent introns and flanking regions. The CTD included in exon 4 is depicted as a white box. Probes used to isolate the *RpII215* region in *D. subobscura* are presented below the gene.

*Sal*I-*Sal*I fragment that partially contained the third and fourth exons of the *RpII215* gene of *D. subobscura* was used as probe to perform screening of the libraries. In each case, several positive recombinant phages were isolated. After restriction map analysis, the entire *RpII215* region was cloned in two DNA fragments for *D. madeirensis* (7.8-kb *Eco*RI-*Hin*dIII and 4.8-kb *Eco*RI-*Eco*RI) and in three fragments for *D. guanche* (4-kb *Eco*RI-*Eco*RI, 0.3-kb *Eco*RI-*Eco*RI, and 4.5-kb *Eco*RI-*Xba*I). Synthetic oligonucleotides designed on the *D. subobscura* sequence were used to obtain the complete sequence of the *RpII215* region of *D. madeirensis* (7.1 kb) and *D. guanche* (6.2 kb) for both strands. For those regions with a high level of divergence, new specific primers were designed. The dideoxy method and double-stranded plasmid DNA were used to manually sequence one strand, while fluorescent dye-terminator chemistry (Perkin Elmer, Norwalk, CT) and an ABI 377 sequencer were used to obtain the sequence of the other strand by cycle sequencing.

For *D. pseudoobscura*, a highly inbred strain (kindly provided by C. Segarra) and a standard small-scale method (Ashburner 1989) were used to isolate genomic DNA from adult flies. Four overlapping fragments, 1.9, 2.3, 1.4, and 1.8 kb long, were amplified by PCR from genomic DNA. Most of the primers used for the amplifications and for sequencing were designed on the *D. subobscura* sequence, although some specific oligonucleotides were designed in highly diverged regions. The cycle sequencing method with fluorescent dye-terminator chemistry and an ABI 377 sequencer were used to obtain the sequence of both strands of a 5.7-kb region.

All sequences were assembled using Staden's programs (Staden 1982). Sequences newly reported in this study are deposited in the EMBL sequence database library under accession nos. Y18876, Y18877, Y18878, and Y18879.

Species divergence: The GCG Wisconsin package programs (v 7.3; Devereux *et al.* 1984) were used to align the sequences. For coding regions, insertions and deletions were placed by eye to minimize the number of amino acid replacements. The *K*-Estimator 4.4 program provided by J. Comeron was used to calculate synonymous (K_s) and nonsynonymous (K_a) divergence estimates per site according to Comeron (1995). This is a modification of Li's (1993) method that tends to minimize stochastic errors and quantifies in a more accurate way the number of transitions/transversions substitutions. In the sliding window analysis, confidence intervals of the estimated divergence values for each window were calculated according to Comeron and Aguadé (1996). For noncoding regions, Kimura's (1980) two-parameter method was applied to estimate the number of substitutions per site.

Codon bias: The codon bias index (CBI; Morton 1993) was used to estimate the degree of biased usage of synonymous codons (codon bias) of the *RpII215* coding region in each species. This measure exhibits a much lower dependence on the number of analyzed codons and a lower dispersion due to sampling than "Scaled χ^2" (χ^2/L; Shields *et al.* 1988). Moreover, it is independent of the length of the coding region (Comeron and Aguadé 1998). For the subobscura cluster species (*D. subobscura*, *D. madeirensis*, and *D. guanche*), codons were classified as preferred and unpreferred according to Akashi and Schaeffer (1997).

Tests of the molecular clock hypothesis: Two different kinds of tests were used to contrast the molecular clock hypothesis (Zuckerkandl and Pauling 1965). Tajima's relative rate tests with known outgroup (1D and 2D; Tajima 1993) were applied to test equal rates of evolution between lineages: *D. guanche* was used as the outgroup between *D. subobscura* and *D. madeirensis*, and *D. pseudoobscura* was used as the outgroup between *D. subobscura* and *D. guanche*. Both in the *D. subobscura*/*D. madeirensis*/*D. guanche* and in the *D. subobscura*/*D. pseudoobscura*/*D. melanogaster* comparisons, the index of dispersion (*R(t)*; Gillespie 1989, 1991) was calculated to test whether the number of substitutions on a lineage is Poisson-distributed and whether evolutionary rates across lineages are constant (Kimura and Ohta 1971). In the latter comparison and to avoid negative values, the number of substitutions on a given branch of the phylogeny built by the three species was calculated by comparison of each sequence with a generated ancestral sequence that was constructed using a parsimony criterion (Zeng *et al.* 1998). Computer simulations that considered the multiple hits effect were conducted to obtain significance levels of the estimated *R* values (Bulmer 1989; Gillespie 1989, 1991) in accordance with Zeng *et al.* (1998).

RESULTS

Molecular evolutionary rates: The *RpII215* region was sequenced in four species of the obscura group (Figure 2): *D. subobscura* (7816 bp), *D. madeirensis* (7103 bp), *D. guanche* (6220 bp), and *D. pseudoobscura* (5666 bp). The structure of the gene in *D. subobscura* consists of four exons separated by three introns whose lengths are 1029, 77, and 108 bp, respectively. The different exons presented the same length in the four species compared (81, 2244, 2245, and 1097 bp long, respectively). In *D. guanche*, the first large intron presented a 783-bp deletion in the central region relative to *D. subobscura* and *D. madeirensis*. This central region of *D. subobscura* showed extensive similarity to an 823-bp region that is defined as uncharacterized highly repetitive sequence of the same species (EMBL accession no. AF043638), and is also described in *D. madeirensis* (AF043637) and in *D. guanche* (AF043639). Table 1 gives a summary of divergence estimates between species for the *RpII215* region. Introns and flanking regions could only be aligned between the obscura group species. Divergence estimates in these noncoding regions were generally higher than the corresponding synonymous estimates.

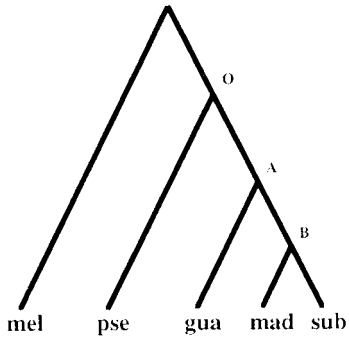

Figure 2.—Phylogenetic relationships among *D. subobscura* (sub), *D. madeirensis* (mad), *D. guanche* (gua), *D. pseudoobscura* (pse), and *D. melanogaster* (mel). Branch lengths do not represent real divergence distances. Node 0 indicates the inferred ancestral sequence (sequence 0) between *D. subobscura* and *D. pseudoobscura*. This phylogeny is based on data at the chromosomal (Krimbas and Loukas 1984), allozyme (Loukas *et al.* 1984), and nucleotide levels (Barrio *et al.* 1992; Barrio and Ayala 1997; Ramos-Onsins *et al.* 1998), and it is also supported by the present data.

TABLE 1

Substitutions per site in coding and noncoding regions of the *RpII215* gene

	D. subobscura	*D. madeirensis*	*D. guanche*	*D. pseudoobscura*	*D. melanogaster*
Coding[a]					
D. subobscura	–	0.021 [0.014–0.029]	0.058 [0.045–0.07]	0.28 [0.25–0.31]	0.91 [0.83–0.99]
D. madeirensis	0.0005 [0–0.001]	–	0.054 [0.042–0.066]	0.28 [0.25–0.31]	0.9 [0.82–0.99]
D. guanche	0.001 [0–0.002]	0.001 [0–0.003]	–	0.30 [0.27–0.34]	0.93 [0.84–1]
D. pseudoobscura	0.002 [0.001–0.004]	0.003 [0.001–0.004]	0.001 [0–0.003]	–	0.92 [0.84–1]
D. melanogaster	0.02 [0.015–0.025]	0.001 [0–0.003]	0.019 [0.014–0.023]	0.019 [0.014–0.023]	–
Noncoding[b]					
D. subobscura	–	0.047 [0.036–0.059]	0.076 [0.053–0.1]	0.34 [0.25–0.45]	
D. madeirensis	1372	–	0.076 [0.054–0.1]	0.35 [0.25–0.46]	
D. guanche	528	537	–	0.33 [0.24–0.43]	
D. pseudoobscura	196	196	203	–	

Ninety-five percent confidence intervals obtained by computer simulation using the *K*-Estimator 4.4 program are depicted between square brackets.

[a] Numbers of synonymous and nonsynonymous substitutions per site are depicted above and below the diagonal (upper part), respectively (1419.6 and 4219.6 average number of synonymous and nonsynonymous sites analyzed).

[b] Numbers of noncoding substitutions per site and numbers of positions analyzed are shown above and below the diagonal (lower part), respectively.

D. madeirensis and *D. guanche* are insular species geographically restricted to Madeira and to the Canary Islands, respectively. In contrast, *D. subobscura* is a palearctic species, and consequently it is expected to have a larger N_e than the insular species. We have analyzed the effect of the expected smaller N_e of *D. madeirensis* and *D. guanche* on the nucleotide substitution rates by applying Tajima's relative rate tests (Tajima 1993) between these species and *D. subobscura*. On the other hand, Akashi (1995, 1996) proposed that *D. melanogaster* exhibits the effects of a reduction of N_e on nucleotide substitutions when compared to *D. simulans*. We have addressed the question of whether those effects in *D. melanogaster* are also detectable when compared to the *D. subobscura* and *D. pseudoobscura* lineages. Relative rate tests, however, use only the outgroup species to assign nucleotide substitutions to each internal lineage, and substitution rates cannot be tested in the external branch. The mean-to-variance ratio (Kimura 1983; Gillespie 1989, 1991), also known as *R(t)*, allowed us to test the molecular clock hypothesis (Zuckerkandl and Pauling 1965) in a phylogeny built by *D. subobscura*, *D. pseudoobscura*, and *D. melanogaster*.

Relative rate tests between D. subobscura, D. madeirensis, and *D. guanche*: Tajima's relative rate tests (1D and 2D methods; Tajima 1993) were applied to coding (synonymous and nonsynonymous) and noncoding regions separately. For variable sites, only those positions with the same nucleotide in two of the three sequences compared were considered. As in the comparisons between *D. subobscura*, *D. madeirensis*, and *D. guanche*, variable codons were affected by a single change, and each observed substitution in the coding region of the *RpII215* gene could be classified unambiguously as synonymous or nonsynonymous. Results from the 1D method are shown in Table 2. No significant departure from the molecular clock hypothesis was detected between the *D. subobscura* and *D. madeirensis* lineages using *D. guanche* as the outgroup, either when the total number of the observed substitutions (1D method) was considered or when transitional and transversional changes were analyzed separately (2D method). In contrast, the *D. guanche* lineage exhibited a significantly larger number of substitutions in the coding region than the *D. subobscura* lineage using *D. pseudoobscura* as the outgroup, when both the 1D method ($P = 0.016$) and the 2D method ($P = 0.015$) were used. The different substitution rates in the coding region between both lineages can be attributed to synonymous substitutions ($P = 0.007$ and $P = 0.014$ for 1D and 2D methods, respectively). Contrarily, the noncoding region seems to fit the constant rate hypothesis, which suggests that noncoding and synonymous substitutions in the *RpII215* region behave differently.

Index of Dispersion Among D. subobscura, D. pseudoobscura, and D. melanogaster: We used Gillespie's *R(t)* to test the molecular clock hypothesis in the phylogeny

TABLE 2

Tajima's relative rate test in the subobscura cluster species of Drosophila

	Substitutions			*P* values[a]		
	Coding region			Coding region		
	Nonsyn[b]	Syn[c]	Noncoding	Syn	Total	Noncoding
D. subobscura	0	19	7	0.21	0.4	1
D. madeirensis *D. guanche*[d]	2	12	7			
D. subobscura	3	24	6	0.007	0.016	0.32
D. guanche *D. pseudoobscura*[d]	0	48	3			

[a] Calculated from Tajima's relative rate test (1D method).
[b] Nonsynonymous.
[c] Synonymous.
[d] The species used as the outgroup in each comparison.

built by *D. subobscura*, *D. pseudoobscura*, and *D. melanogaster* (see Figure 2). Analysis focused on the *RpII215* coding region because noncoding regions could not be aligned between any species of the obscura group and *D. melanogaster*. The numbers of synonymous and nonsynonymous substitutions in each lineage were calculated using the inferred ancestral sequence (sequence 0 in Figure 2) of both obscura species according to Zeng *et al.* (1998). The lineage from the *D. subobscura*/*D. pseudoobscura* split to *D. melanogaster* was defined as the *D. melanogaster* lineage. Langley and Fitch (1974) pointed out the important contribution of lineage and residual effects on *R* values. Lineage effects, like the generation-time effect and different branch lengths, were considered constant within a lineage and were removed by the method proposed by Gillespie (1989, 1991). In the present study, we used the synonymous and nonsynonymous weights calculated in Zeng *et al.* (1998) from sequences of 24 genes in the same three species. The *R* values estimated for synonymous and nonsynonymous substitutions in the *RpII215* gene are shown in Table 3.

A significant departure from the expected Poisson process was detected for synonymous ($P = 0.025$) and nonsynonymous ($P = 0.008$) substitutions in the *RpII215* gene. The *D. melanogaster* lineage presented an excess of both kinds of substitutions. A similar result was obtained when the numbers of synonymous and nonsynonymous substitutions on each branch were estimated directly from the corrected distances between real sequences (Gillespie 1989), as the calculated *R(t)* values (9.74 and 6.36 for synonymous and nonsynonymous substitutions, respectively) were even higher than those obtained by means of the constructed ancestral sequence.

Analysis of preferred and unpreferred codons: It has been proposed that mutations at synonymous sites behave near neutrality (Ohta and Kimura 1971; Ohta 1972; Kimura 1983), and their fate in the population would therefore depend on the effective population size. Codons within a synonymous family can be classified as major (preferred) and nonmajor (unpreferred) codons according to Akashi (1995). Selection coefficients for mutations from preferred to unpreferred codons (unpreferred changes) were calculated by Akashi (1995, 1997), who suggested deleterious and beneficial effects on fitness of unpreferred and preferred changes, respectively. The observed numbers of preferred and unpreferred changes at the *RpII215* coding region in the *D. subobscura*, *D. madeirensis*, and *D. guanche* lineages were studied and the results are summarized in Table 4. Preferred and unpreferred changes were assigned to one lineage by comparison to an outgroup sequence.

TABLE 3

Estimated numbers of synonymous and nonsynonymous substitutions and indexes of dispersion [*R*(*t*)]

	Substitutions			*R* (*t*)
	0-sub[a]	0-pse[b]	0-mel[c]	*RpII215*
Synonymous	186.30	186.43	908.19	4.95*
Nonsynonymous	6.50	3.01	72.73	4.63**

The $R(t)$ values were calculated using the synonymous ($W_{sub} = 0.475$; $W_{pse} = 0.535$; $W_{me1} = 1.99$) and nonsynonymous ($W_{sub} = 0.429$; $W_{pse} = 0.412$; $W_{me1} = 2.159$) weights proposed by Zeng *et al.* (1998) for each lineage. Asterisks indicate the level of significance achieved by computer simulations for one-tailed tests: $*P < 0.05$; $**P < 0.01$.
[a] The *D. subobscura* lineage.
[b] The *D. pseudoobscura* lineage.
[c] The *D. melanogaster* lineage.

TABLE 4

Observed numbers of unpreferred and preferred changes at the *RpII215* coding region in the subobscura cluster species of Drosophila

	D. guanche[a]		*D. pseudoobscura*[a]	
	D. subobscura	*D. madeirensis*	*D. subobscura*	*D. guanche*
Pref → unpref	3	5	9	25
Unpref → pref	6	4	8	4
P values	0.36		0.017	

The observed numbers of unpreferred (pref → unpref) and preferred (unpref → pref) changes in the *D. subobscura*, *D. madeirensis*, and *D. guanche* lineages are shown. Probability values (*P* values) were calculated using a *G*-test with Williams' correction (2 × 2 contingency table with the above numbers; Sokal and Rohlf 1997).

[a] For each comparison, the species used as the outgroup.

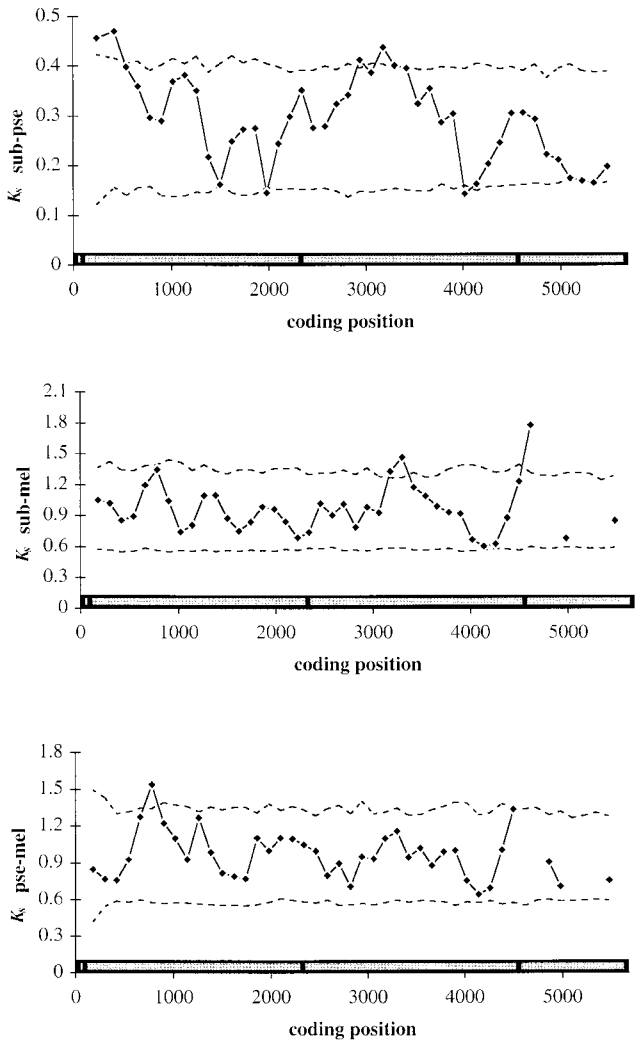

Figure 3.—Sliding window analysis across the *RpII215* coding region (in the *x* axis) of the number of synonymous substitutions per site (K_s in the *y* axis). Exons of the *RpII215* gene are depicted as black rectangles below the graph. Ninety-five percent confidence intervals of the synonymous divergence estimates are shown as dotted lines. Window size, 360 nucleotides. sub, *D. subobscura*; pse, *D. pseudoobscura*; mel, *D. melanogaster*.

For each variable site, the ancestral nucleotide was the one present in two of the three sequences compared. Positions with different nucleotides in the three sequences were not considered in this analysis. In the comparison between *D. subobscura* and *D. guanche*, the number of unpreferred changes in the *D. guanche* lineage was six times larger than the number of preferred changes. In contrast, preferred and unpreferred changes were equally frequent in the *D. subobscura* lineage. We tested the ratio of preferred to unpreferred changes between *D. subobscura* and each of the insular species (*D. madeirensis* and *D. guanche*) by applying a *G*-test of independence (Table 4). There was a significant excess ($P = 0.017$) of unpreferred changes in the *D. guanche* lineage as compared to the *D. subobscura* lineage. Otherwise, the *D. madeirensis* and *D. subobscura* lineages showed an equivalent ratio of preferred to unpreferred changes. Both results are consistent with our previous analysis using Tajima's relative rate test and support a reduction in the intensity of selection on synonymous mutations at the *D. guanche* lineage caused by its smaller N_e.

Divergence and codon bias across the coding region: Synonymous codon usage (codon bias) for the entire *RpII215* coding region was studied using the CBI (Morton 1993). The CBI values for the species *D. subobscura*, *D. madeirensis*, *D. guanche*, and *D. pseudoobscura* were 0.505, 0.502, 0.475, and 0.522, respectively. For *D. melanogaster* (Jokerst *et al.* 1989) the CBI value was 0.411. The distribution of synonymous substitutions across the *RpII215* coding region was studied in the comparisons between *D. subobscura*, *D. pseudoobscura*, and *D. melanogaster*. We performed a sliding window analysis using five different window sizes: 360, 450, 540, 630, and 720 nucleotides. Figure 3 shows the analysis for a window size of 360 nucleotides. In the comparisons between *D. melanogaster* and either *D. subobscura* or *D. pseudoobscura*, for all window sizes there was a variable number of windows for which synonymous divergence could not be calculated because of the saturation of the observed synonymous substitutions per site (NA windows). All these windows encompassed, totally or partially, the CTD re-

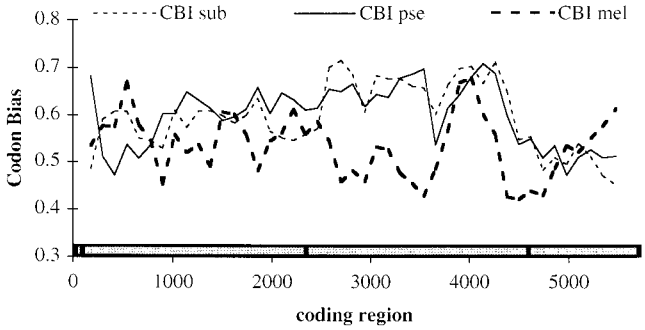

Figure 4.—Sliding window analysis of the codon bias (CBI in *y* axis) across the *RpII215* gene (exons are depicted as black rectangles below the graph) of the *D. subobscura* (sub), *D. pseudoobscura* (pse), and *D. melanogaster* (mel) species. Window size, 360 nucleotides.

gion that begins at nucleotide 4741 of the *RpII215* coding region of *D. melanogaster* (nucleotide 4735 in the obscura group species). We generated a null distribution to calculate the probability of detecting the observed number of NA windows by chance. Codon positions of the *D. melanogaster-D. subobscura* comparison were randomized and the same sliding window analysis of synonymous divergence that was applied to the original data set was performed. The number of NA windows from the randomized sequences was compared to the observed number obtained from the data set. For all window sizes, the probability of obtaining by chance a number of NA windows equal or higher than that observed was ≤ 0.002, which indicates a heterogeneous distribution of synonymous substitutions across the *RpII215* coding region.

Figure 4 shows the codon usage bias across the *RpII215* coding region of *D. melanogaster*, *D. subobscura*, and *D. pseudoobscura*. The distribution across the coding region of the codon bias estimator CBI was fairly similar in both species of the obscura group. The sequence of *D. melanogaster*, otherwise, showed a region (from approximately nucleotide 2500 to 4000 of the coding region) with a different pattern than that observed in the obscura species. The negative correlation between synonymous divergence and codon bias among genes is well known (Shields *et al.* 1988; Sharp and Li 1989). This correlation has been usually associated with differences in the expression level (Shields *et al.* 1988; Moriyama and Hartl 1993). However, the possible correlation between synonymous divergence and codon bias across a coding region cannot be explained by differences in the level of expression (Comeron and Aguadé 1996). For adjacent windows of 540 nucleotides, synonymous divergence between *D. melanogaster* and *D. subobscura* across the *RpII215* coding region was strongly correlated with the *D. melanogaster* codon bias (Kendall's nonparametric correlation $\tau = -0.911$; $P = 0.0002$) but not with that of *D. subobscura* ($\tau = 0.0$; $P > 0.99$; see Figure 5). The significant negative correlation held for window sizes of 360 and 630 nucleotides with probability values of 0.0041 and 0.013, respectively. These results are consistent with the observation that most synonymous substitutions between *D. subobscura* and *D. melanogaster* were located preferentially in the *D. melanogaster* lineage. In fact, there was a negative correlation across the coding region between the *D. melanogaster* codon bias and the number of synonymous substitutions per site in the *D. melanogaster* lineage ($\tau = -0.644$; $P = 0.0095$). Neither in the *D. subobscura* nor in the *D. pseudoobscura* lineages was there a significant relationship between codon bias and K_s across the *RpII215* coding region ($P = 0.53$ and $P = 0.94$, respectively). These results support the possible acceleration of the *D. melanogaster* lineage detected by the significant index of dispersion.

Comparison between the CTD and non-CTD regions: Gillespie's index of dispersion of synonymous substitutions was calculated separately for the CTD and the non-CTD regions of the *RpII215* gene. Initially, two different sets of synonymous weights were used to compensate for lineage effects: (i) weights calculated by Zeng *et al.* (1998), using 24 genes sequenced in the same three species (*D. subobscura*, *D. pseudoobscura*, and *D. melanogaster*), and (ii) weights of the entire *RpII215* gene. The

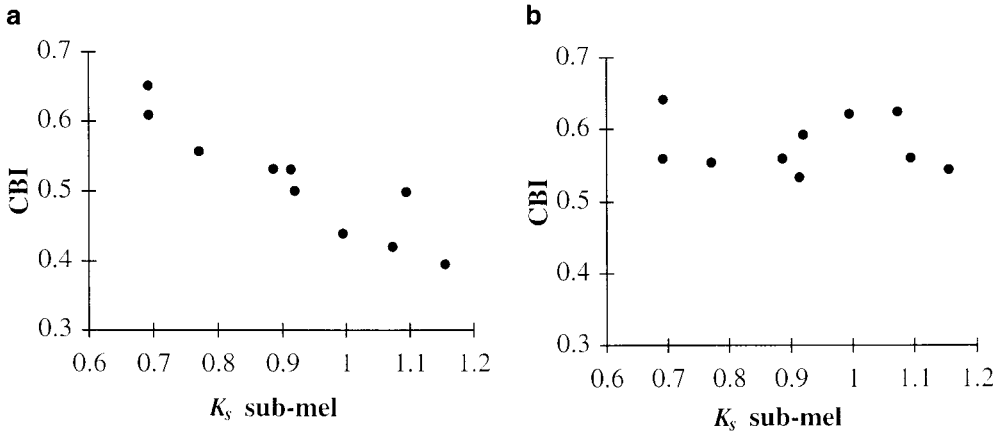

Figure 5.—Relationship between the *D. subobscura-D. melanogaster* synonymous divergence per site (K_s) and the codon bias (CBI) of (a) *D. melanogaster* and of (b) *D. subobscura* across the *RpII215* coding region. Probability values from Kendall's nonparametric correlations are $\tau = -0.911$, $P = 0.0002$ and $\tau = 0.0$, $P > 0.99$ for the (a) *D. melanogaster* and (b) *D. subobscura* CBI values, respectively. Size of adjacent windows, 540 nucleotides.

TABLE 5

Estimated numbers of synonymous substitutions and indexes of dispersion [*R*(*t*)] for the non-CTD and CTD regions of the *RpII215* gene

	Substitutions			*R*(*t*)			
	0-sub[a]	0-pse[b]	0-mel[c]	*W*	$W_{RpII215}$	W_{CTD}	$W_{non\text{-}CTD}$
Non-CTD	163.38	160.06	718.54	2.05	0.62	19.31***	
CTD	23.32	26.65	199.32	8.56**	3.91*		5.73*

The *R*(*t*) values were calculated using the following: synonymous weights (W) proposed by Zeng *et al.* (1998), $W_{sub} = 0.475$, $W_{pse} = 0.535$, $W_{mel} = 1.99$; the *RpII215* synonymous weights ($W_{RpII215}$), $W_{sub} = 0.436$, $W_{pse} = 0.437$, $W_{mel} = 2.127$; the CTD synonymous weights (W_{CTD}), $W_{sub} = 0.281$, $W_{pse} = 0.321$, $W_{mel} = 2.399$; and the non-CTD synonymous weights ($W_{non\text{-}CTD}$), $W_{sub} = 0.470$, $W_{pse} = 0.460$, $W_{mel} = 2.069$. Asterisks indicate significance levels based on computer simulations for one-tailed tests: $*P < 0.05$; $**P < 0.01$; and $***P < 0.001$.

[a] The *D. subobscura* lineage.

[b] The *D. pseudoobscura* lineage.

[c] The *D. melanogaster* lineage.

numbers of synonymous substitutions in each lineage and the calculated *R* values for the two regions are summarized in Table 5. The CTD region has accumulated more synonymous substitutions in the *D. melanogaster* lineage than expected according to both the Zeng *et al.* (1998) survey ($R = 8.56$, $P = 0.002$) and the whole *RpII215* coding region ($R = 3.91$, $P = 0.042$). In contrast, the non-CTD region did not show a significant departure from the general tendency described in Zeng *et al.* (1998). Finally, we addressed the possible incompatibility of synonymous rates between the non-CTD and the CTD regions in the three lineages studied. The CTD region showed a significant *R* value ($P = 0.018$) when the weights of the non-CTD region were used. The probability value was much lower ($R = 19.31$, $P < 0.001$) for the non-CTD region when the CTD weights were used. Forces with different intensity seem to have driven the synonymous evolution of both regions of the *RpII215* gene.

DISCUSSION

The observed low level of nonsynonymous divergence in the *RpII215* gene indicates that purifying selection plays an important role in the evolution of the corresponding protein. In contrast, the synonymous substitution rate is moderately high. The average numbers of nonsynonymous (K_a) and synonymous (K_s) substitutions per site between *D. melanogaster* and the obscura species are 0.02 and 0.905, respectively, while the reported averages for 24 genes are 0.08 and 0.81, respectively (Zeng *et al.* 1998). In the comparison between *D. subobscura* and *D. pseudoobscura*, which is not affected by the smaller N_e of *D. melanogaster*, the K_s estimate for the *RpII215* gene is usually higher than those observed among genes with low levels of nonsynonymous divergence. In fact, of the 14 genes with $K_a = 0.0094$ or lower (Zeng *et al.* 1998), 11 showed K_s values lower than the *RpII215* estimate. The *RpII215* gene encodes the largest component of the RNA pol II complex and, *a priori*, it is therefore expected to have a ubiquitous expression. Moreover, the reduced number of amino acid replacements in its coding region suggests that accuracy acting at the translational level may play a significant role in shaping its codon bias (Akashi 1994). Unexpectedly, however, the *RpII215* gene of *D. melanogaster* showed low bias in codon usage. Recently, a positive correlation between synonymous divergence and the length of the coding region (Comeron 1997; Comeron *et al.* 1999) as well as a negative correlation between the degree of codon bias and the length of the coding region (Comeron 1997; Moriyama and Powell 1998; Comeron *et al.* 1999) have been found. Comeron *et al.* (1999) have proposed two different models to explain the correlation between the length of the coding region, the codon bias, and the synonymous divergence for the entire range of recombination rates in Drosophila. In these models selection on synonymous mutations would act less efficiently on long genes than on short genes. The *RpII215* gene has a very long coding region (1889 codons in *D. melanogaster*), within the 5% longest coding regions sequenced in this species. Our observation of low codon bias and high synonymous divergence in the *RpII215* gene would be consistent with those predicted for very long coding regions (Comeron *et al.* 1999).

Evolutionary rates of the *RpII215* gene: If we assume that substitutions in noncoding regions are neutral, they should show generation-time effects (Ohta and Kimura 1971; Ohta 1993) and be independent of changes in N_e. According to our results from Tajima's (1993) relative rate tests, the numbers of substitutions at the *RpII215* noncoding region are not significantly different between the *D. subobscura* and *D. guanche* lineages. It is therefore likely that both species have a similar generation time and mutation rate. On the other hand, according to Ohta (1972), nearly neutral mutations have selection coefficients close to the inverse of the effective population size ($|N_e s| \approx 1$), while effectively neutral mu-

tations satisfy the inequality $|N_e s| \ll 1$. Mutations on synonymous sites in Drosophila are under the influence of weak selection (Akashi 1995, 1996). The amount of synonymous mutations that could be maintained in a population and their probability of fixation would therefore depend on N_e (Kimura 1983). A reduction of the effective population size would increase the fraction of mutations that are considered strictly neutral.

D. guanche is restricted to very specific locations of the Canary Islands. As detected by the significant relative rate tests, the rate of synonymous substitutions at the *RpII215* gene was higher in the *D. guanche* than in the *D. subobscura* lineage. Also, the *R* value for synonymous substitutions ($R = 5.51$, $P = 0.01$), estimated using weights from the five genes available in *D. subobscura*, *D. madeirensis*, and *D. guanche* species (*Adh*, *Adhr*, Marfany and Gonzàlez-Duarte 1993; *Gpdh*, *Sod*, Barrio and Ayala 1997; and *rp49*, Ramos-Onsins *et al.* 1998), was consistent with the relative rate test results. We propose, therefore, that this higher rate of synonymous substitutions in the *D. guanche* lineage may be caused by a smaller effective population size. This smaller N_e, probably associated with both the origin of *D. guanche* and its current distribution, could have increased the fixation rate of the nearly neutral synonymous mutations in the *RpII215* gene. As argued above, the fate of mutations in noncoding regions would not have been affected because their selection coefficients would be much smaller ($s \ll 1/N_e$) and therefore their fate would not depend on N_e. A higher rate of synonymous substitutions in the *D. guanche* lineage as compared to the *D. subobscura* lineage might also be caused by a lower recombination rate at the *RpII215* region, and therefore be equivalent to a lower N_e (Hill and Robertson 1966; Felsenstein 1974) in *D. guanche*. Changes in the recombinational map between closely related species (True *et al.* 1996) in centromere or telomere-proximal regions have been recently described. The cytological location of the *RpII215* region on the sexual chromosome is equivalent (band 10A) in the three species of the subobscura cluster (Segarra and Aguadé 1992) far from the centromeric and telomeric regions. Although we cannot exclude changes in the recombination rate between these species, the smaller N_e of *D. guanche* seems a more plausible explanation of the higher synonymous substitution rate in this lineage.

The proposed smaller N_e of *D. guanche* should have affected other regions of the genome. Analysis of the other five genes sequenced in the three species of the subobscura cluster revealed that the numbers of synonymous differences between *D. subobscura* and *D. guanche* were low (12, 15, 8, 5, and 2 for *Adh*, *Adhr*, *Gpdh*, *Sod*, and *rp49*, respectively) as compared to the 72 synonymous substitutions found in the *RpII215* gene. *Adhr*, the gene with the highest number of synonymous substitutions, exhibited the same tendency (Tajima's 1D method, $P = 0.075$) to accumulate more synonymous changes in the *D. guanche* than in the *D. subobscura* lineage.

The smaller effective population size of *D. guanche* as compared to *D. subobscura* would have reduced the effectiveness of natural selection on synonymous mutations in that lineage. An equivalent effect in *D. melanogaster* as compared to *D. simulans* was reported by Akashi (1996). A reduction in the effectiveness of natural selection on synonymous sites would have caused the observed significant excess of unpreferred changes at the *RpII215* gene in the *D. guanche* lineage. Our results, on the basis of the analysis of preferred and unpreferred codons, are consistent with measures of codon bias in the entire *RpII215* coding region in both *D. subobscura* and *D. guanche*, which indicates that natural selection on synonymous mutations has been less effective in the *D. guanche* lineage. The constancy of the preferred to unpreferred changes ratio (see Table 4) was contrasted with a *G*-test using the number of nucleotide differences between sequences. Although this ratio is expected to be different for fixed and polymorphic changes, the ratio of total preferred to unpreferred changes (fixed and polymorphic) is predicted to be the same for lineages with equivalent branch lengths, N_e's, and generation times. Under the assumption of equal branch lengths and generation times, as would be the case for the *D. subobscura* and the *D. guanche* lineages when compared to *D. pseudoobscura*, the significant result from the *G*-test would also point to a different N_e between *D. subobscura* and *D. guanche*.

In the *D. madeirensis* lineage neither noncoding nor synonymous substitutions departed from molecular clock expectations when compared to the *D. subobscura* lineage. These results, though consistent with those found for the *rp49* region (Ramos-Onsins *et al.* 1998), are quite unexpected, considering that *D. madeirensis*, as *D. guanche*, is currently an insular species. Although the A (=X) chromosome shows some structural reorganizations between *D. subobscura* and *D. madeirensis* (Papaceit and Prevosti 1989), the cytological location of the *RpII215* gene does not change between these two species. The short divergence time between *D. subobscura* and *D. madeirensis* ($0.6–1 \times 10^6$ years according to Ramos-Onsins *et al.* 1998) could explain our results, considering that the average fixation time for neutral mutations is $4N_e$ generations and that this time increases exponentially for deleterious mutations (Kimura 1983). The much longer divergence time between *D. subobscura* and *D. guanche* would have allowed us, therefore, to detect the effect of a smaller N_e on synonymous substitution rates.

The mean-to-variance ratios (R) calculated for nonsynonymous and synonymous substitutions at the *RpII215* gene were significantly higher than one. A reduction or fluctuation of N_e in the *D. melanogaster* lineage (Zeng *et al.* 1998) could have caused the observed overdispersion. Congruently, the CBI measures of codon

bias of the *RpII215* gene in the obscura group species were systematically higher than in *D. melanogaster*, suggesting that the effectiveness of natural selection acting on synonymous sites might be different between the *D. melanogaster* and the obscura lineages. We conclude, therefore, that mutations on synonymous sites at the *RpII215* gene are indeed nearly neutral and hence sensitive to changes in N_e.

Different selection coefficients on synonymous mutations across the *RpII215* gene: It is widely accepted that in Drosophila there is a negative correlation between synonymous substitution rates and codon bias (Shields *et al.* 1988; Sharp and Li 1989). Shields *et al.* (1988) suggested that in some Drosophila genes there would be a positive correlation between codon bias and expression levels, which would indicate a stronger selection on highly expressed genes. Differences in codon bias among genes could then be explained by different expression levels, which suggests a wide range of selection coefficients on synonymous mutations. The length of the *RpII215* coding region (5667 bp) and the high level of synonymous divergence were the most adequate to study variation of codon bias and synonymous divergence across the coding region. Synonymous divergence (K_s) between *D. subobscura* and *D. melanogaster* correlates negatively with the *D. melanogaster* codon bias across the *RpII215* coding region, which suggests that the selection coefficients on synonymous mutations may vary not only among genes but also within a particular gene. This variation of selection coefficients within the *RpII215* gene is supported by the observed heterogeneous distribution of the synonymous divergence across the coding region and, obviously, cannot be explained by different levels of expression. The intragenic analysis of codon bias and synonymous divergence allowed us to study at what levels natural selection would most probably act on synonymous mutations at this gene. According to a mutation-selection-drift theory (Bulmer 1991), natural selection could modulate the synonymous codons usage to enhance translational efficiency (translational accuracy and/or elongation rates) or to maintain the mRNA secondary structure (Hasegawa *et al.* 1979; Stephan and Kirby 1993; Parsch *et al.* 1997). Conflicting selection pressures on synonymous mutations, as predicted by selection on mRNA structure, would prevent a negative correlation between K_s and codon bias (Eyre-Walker and Bulmer 1993). The observed correlation between codon bias and synonymous divergence across the *RpII215* coding region indicates that selection acts to enhance translational efficiency rather than to maintain the mRNA secondary structure. Bulmer (1991) suggested that selection at the level of translational accuracy would generate a negative correlation between codon bias and the rate of nonsynonymous divergence (K_a). Equivalent results are then expected along a given coding region if there is a heterogeneous efficiency of selection on synonymous mutations (Comeron and Aguadé 1996). There is a close to significant negative correlation between codon bias of *D. melanogaster* and nonsynonymous divergence between *D. melanogaster* and *D. subobscura* along the *RpII215* coding region ($\tau = -0.45$; $P = 0.052$; window size, 540 nucleotides). In agreement with the general low rate of nonsynonymous substitutions, this correlation suggests that selection would contribute to shaping of the codon bias of the *RpII215* gene by enhancing the accuracy of translation.

We have already pointed out that changes in the effective size of populations could affect the fate of mutations differently, depending on their selection coefficients. We have also shown that synonymous selection coefficients vary across the *RpII215* coding region. The estimated mean-to-variance ratios (index of dispersion; Gillespie 1989, 1991) for the CTD and the non-CTD regions of the *RpII215* gene support the proposal that selection coefficients of synonymous mutations are detectably different in these two regions. Indeed, the synonymous substitution rates of the non-CTD region did not show a significant departure from the tendency described in Zeng *et al.* (1998). In contrast, the proposed reduction or fluctuation of the *D. melanogaster* N_e would have had a stronger effect on the CTD region. The CTD is a highly conserved structure with an essential function among a wide range of organisms (Nonet *et al.* 1987). It contains several amino acids (primarily serine residues but also threonine and tyrosine residues to a lesser degree) that can be phosphorylated (Zhang and Corden 1991; Baskaran *et al.* 1993; Yuryev and Corden 1996). An important interaction between the phosphorylated CTD of the largest subunit of the RNA polymerase II complex and the enzyme responsible for mRNA capping has been reported recently (McCracken *et al.* 1997). A stronger selective constraint at the translational level on synonymous mutations in the CTD region than in the non-CTD region may explain the significantly detected different effect of the N_e change in the *D. melanogaster* lineage in these two regions of the *RpII215* gene.

We are grateful to J. M. Comeron for computer simulations, helpful discussion, and comments on the manuscript. We thank Serveis Científico-Tècnics of Universitat de Barcelona for automatic sequencing facilities. A.Ll. was a predoctoral fellow from Ministerio de Educación y Ciencia, Spain, during this study. This work was supported by grants PB94-0923 from Dirección General de Investigación Científica y Técnica, Ministerio de Educación y Ciencia, Spain, and 1997 SGR-59 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Generalitat de Catalunya, to M.A.

LITERATURE CITED

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

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

Akashi, H., 1996 Molecular evolution between *Drosophila melanogas-*

ter and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics **144:** 1297–1307.

Akashi, H., 1997 Codon bias evolution in Drosophila. Population genetics of mutation-selection-drift. Gene **205:** 269–278.

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

Allison, L. A., M. Moyle, M. Shales and C. J. Ingles, 1985 Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell **42:** 599–610.

Allison, L. A., J. K. Wong, V. D. Fitzpatrick, M. Moyle and C. J. Ingles, 1988 The C-terminal domain of the largest subunit of RNA polymerase of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. Mol. Cell. Biol. **8:** 321–329.

Ashburner M., 1989 *Drosophila: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Barrio, E., and F. J. Ayala, 1997 Evolution of the *Drosophila obscura* species group inferred from the *Gpdh* and *Sod* genes. Mol. Phylogenet. Evol. **7:** 79–93.

Barrio, E., A. Latorre, A. Moya and F. J. Ayala, 1992 Phylogenetic reconstruction of the *Drosophila obscura* group on basis of mitochondrial DNA. Mol. Biol. Evol. **9:** 621–635.

Baskaran, R., M. E. Dahmus and J. Y. Wang, 1993 Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. Proc. Natl. Acad. Sci. USA **90:** 11167–11171.

Bingham, P. M., R. Levis and G. M. Rubin, 1981 Cloning of DNA sequences from *white* locus of *D. melanogaster* by a novel and general method. Cell **25:** 693–704.

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

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

Chao, L., and D. E. Carr, 1993 The molecular clock and the relationship between population size and generation time. Evolution **47:** 688–690.

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

Comeron, J. M., 1997 Estudi de la variabilitat nucleotídica a Drosophila: Regió *Xdh* a *D. subobscura*. PhD thesis. Universitat de Barcelona, Barcelona, Spain.

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

Comeron, J. M., and M. Aguadé, 1998 An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. **47:** 268–274.

Comeron, J. M., and M. Kreitman, 1998 The correlation between synonymous and nonsynonymous substitutions in Drosophila: Mutation, selection, or relaxed constraints? Genetics **150:** 767–775.

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

Corden, J. L., 1990 Tails of RNA polymerase II. Trends Biochem. Sci. **15:** 383–387.

Devereux, J., P. Haeberli and O. Smithies, 1984 A comprehensive set of sequence analysis programs for VAX. Nucleic Acids Res. **12:** 387–395.

Eyre-Walker, A., and M. Bulmer, 1993 Reduced synonymous substitution rate at the start of enterobacterial genes. Nucleic Acids Res. **21:** 4599–4603.

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

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

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford Series in Ecology and Evolution, Oxford University Press, New York.

Hasegawa, A. M., T. Yasunaga and T. Miyata, 1979 Secondary structure of MS2 phage RNA and bias in code word usage. Nucleic Acids Res. **7:** 2073–2079.

Henikoff, S., 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:** 351–359.

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

Jokerst, R. S., J. R. Weeks, W. A. Zehring and A. L. Greenleaf, 1989 Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila. Mol. Gen. Genet. **215:** 266–275.

Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

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

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

Kimura, M., and T. Ohta, 1971 On the rate of molecular evolution. J. Mol. Evol. **1:** 1–17.

Kliman, R. M., and J. Hey, 1994 The effects of mutations and natural selection on codon bias in the genes of Drosophila. Genetics **137:** 1049–1056.

Krimbas, C. B., and M. Loukas, 1984 Evolution of the obscura group Drosophila species. I. Salivary chromosomes and quantitative characters in *D. subobscura* and two closely related species. Heredity **53:** 469–482.

Kurland, C. G., 1987a Strategies for efficiency and accuracy in gene expression. 1. The major codon preference: a growth optimization strategy. Trends Biochem. Sci. **12:** 126–128.

Kurland, C. G., 1987b Strategies for efficiency and accuracy in gene expression. 2. Growth optimized ribosomes. Trends Biochem. Sci. **12:** 169–171.

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.

Lawrence, J. G., D. L. Hartl and H. Ochman, 1991 Molecular considerations in evolution of bacterial genes. J. Mol. Evol. **33:** 241–250.

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

Li, W.-H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

Loukas, M., C. B. Krimbas and Y. Vergini, 1984 Evolution of the obscura group Drosophila species. II. Phylogeny of ten species based on electrophoretic data. Heredity **53:** 483–493.

Marfany, G., and R. Gonzàlez-Duarte, 1993 Characterization and evolution of the *Adh* genomic region in *Drosophila guanche* and *Drosophila madeirensis*. Mol. Phylogenet. Evol. **2:** 13–22.

McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers *et al.* 1997 5′-capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. **11:** 3306–3318.

McKnight, L., and K. R. Yamamoto, 1992 *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Moriyama, E. N., and J. P. Powell, 1998 Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res. **26:** 3188–3193.

Morton, B. R., 1993 Chloroplast DNA codon use: Evidence for selection at the *psb A* locus based on tRNA availability. J. Mol. Evol. **37:** 273–280.

Nonet, M. L., D. Sweetser and R. A. Young, 1987 Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell **50:** 909–915.

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

Ohta, T., 1993 An examination of the generation-time effect on molecular evolution. Proc. Natl. Acad. Sci. USA **90:** 10676–10680.

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

Papaceit, M., and A. Prevosti, 1989 Differences in chromosome A arrangement between *Drosophila madeirensis* and *Drosophila subobscura*. Experientia **45:** 310–312.

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

Precup, J., and J. Parker, 1987 Missenses misreading of asparagine codons as a function of codon identity and context. J. Biol. Chem. **262:** 11351–11356.

Ramos-Onsins, S., C. Segarra, J. Rozas and M. Aguadé, 1998 Mo-

lecular and chromosomal phylogeny in the obscura group of Drosophila inferred from sequences of the *rp49* gene region. Mol. Phylogenet. Evol. **9:** 33–41.

Sanger, F., S. Nicklen and A. R. Coulson, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

Segarra, C., and M. Aguadé, 1992 Molecular organization of the X chromosome in different species of the obscura group of Drosophila. Genetics **130:** 513–521.

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.

Sokal, R. R., and F. J. Rohlf, 1997 *Biometry*. W. H. Freeman & Co., New York.

Staden, R., 1982 Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10:** 4731–4751.

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

Tajima, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **123:** 597–601.

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.

Yuryev, A., and J. L. Corden, 1996 Suppression analysis reveals a functional difference between the serines positions two and five in the consensus sequence of the C-terminal domain of yeast RNA polymerase II. Genetics **143:** 661–671.

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

Zhang, J., and J. L. Corden, 1991 Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. J. Biol. Chem. **266:** 2290–2296.

Zuckerkandl, E., and L. Pauling, 1965 Evolutionary divergence and convergence in proteins, pp. 97–166 in *Evolving Genes and Proteins*, edited by V. Bryson and H. J. Vogel. Academic Press, New York.

Communicating editor: W. Stephan