Divergence of the *yellow* Gene Between *Drosophila melanogaster* and *D. subobscura*: Recombination Rate, Codon Bias and Synonymous Substitutions

Agustí Munté, Montserrat Aguadé and Carmen Segarra

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

Manuscript received February 4, 1997
Accepted for publication May 22, 1997

ABSTRACT

The *yellow* (*y*) gene maps near the telomere of the *X* chromosome in *Drosophila melanogaster* but not in *D. subobscura*. Thus the strong reduction in the recombination rate associated with telomeric regions is not expected in *D. subobscura*. To study the divergence of a gene whose recombination rate differs between two species, the *y* gene of *D. subobscura* was sequenced. Sequence comparison between *D. melanogaster* and *D. subobscura* revealed several elements conserved in noncoding regions that may correspond to putative *cis*-acting regulatory sequences. Divergence in the *y* gene coding region between *D. subobscura* and *D. melanogaster* was compared with that found in other genes sequenced in both species. Both, *yellow* and *scute* exhibit an unusually high number of synonymous substitutions per site (p_s). Also for these genes, the extent of codon bias differs between both species, being much higher in *D. subobscura* than in *D. melanogaster*. This pattern of divergence is consistent with the hitchhiking and background selection models that predict an increase in the fixation rate of slightly deleterious mutations and a decrease in the rate of fixation of slightly advantageous mutations in regions with low recombination rates such as in the *y-sc* gene region of *D. melanogaster*.

IN *Drosophila melanogaster*, the *yellow* (*y*) gene is located at the tip of the *X* chromosome at cytological position *1B*, in a region where recombination is greatly reduced (DUBININ *et al.* 1937). Natural populations of *D. melanogaster* exhibit a low level of nucleotide polymorphism in the *y-achaete-scute* region (AGUADÉ *et al.* 1989; BEGUN and AQUADRO 1991; MARTÍN-CAMPOS *et al.* 1992). In this species a similar reduction in heterozygozity has been detected in other genomic regions with low recombination rates as is the case, for instance, of the *cubitus-interruptus* Dominant gene that maps on the fourth chromosome (BERRY *et al.* 1991) or the *suppressor of forked* region located near the centromere of the *X* chromosome (LANGLEY *et al.* 1993). Moreover, the same pattern of reduced variation has been observed in other Drosophila species, such as *D. simulans* (BEGUN and AQUADRO 1991; BERRY *et al.* 1991; MARTÍN-CAMPOS *et al.* 1992) and *D. ananassae* (STEPHAN and LANGLEY 1989; STEPHAN and MITCHELL 1992), in regions of restricted recombination.

The low level of variation detected in these genomic regions has been explained by two alternative selective models: the hitchhiking model and the background selection model. According to the hitchhiking model, the fixation of selectively favored mutations causes a reduction in neutral variation at closely linked sites (MAYNARD SMITH and HAIGH 1974); the extent of the reduction depends on the strength of selection and on the rate of recombination (KAPLAN *et al.* 1989). In the background selection model, the reduction of intraspecific variation is thought to be due to negative selection acting against deleterious alleles (CHARLESWORTH *et al.* 1993). The effect of background selection on neutral variation is also negatively related to the recombination rate.

In addition, both the hitchhiking and the background selection models predict, for regions with reduced recombination, an increase in the rate of fixation of slightly deleterious mutations and a decrease in the rate of fixation of slightly advantageous mutations (BIRKY and WALSH 1988; CHARLESWORTH 1994). Both predictions are related to the Hill-Robertson effect (HILL and ROBERTSON 1966) that proposes a reduction in the effective population size for a locus linked to a second locus under selection. This reduction implies that selection is relaxed in these regions, which is consistent with the observation that *D. melanogaster* genes located in genomic regions with a drastic reduction in the recombination rate tend to exhibit a low codon bias (KLIMAN and HEY 1993). In fact, AKASHI (1995) inferred weak selection acting against mutations that cause changes from preferred to unpreferred codons. Thus, the lower codon bias detected in regions with no recombination would seem to be the result of an increase in the rate of fixation of these slightly deleterious mutations. This increase would be reinforced further if we consider that codon bias is also affected by weak selection favoring backward mutations from unpreferred to

Corresponding author: Carmen Segarra, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain. E-mail: carme@porthos.bio.ub.es

preferred codons since the fixation rate of these slightly advantageous mutations would be reduced in regions of low recombination.

The predictions of the hitchhiking and background selection models can also be tested by comparing the levels of polymorphism for particular genes that, in different species, map at chromosomal sections with strong differences in the recombination rate. Moreover, predictions of divergence can be tested by comparing, for a particular pair of species, the levels of divergence between genes that map at chromosomal regions with strong differences in recombination rate and genes with similar recombination rates. Only a few data are available according to the first approach in which levels of variation between *D. melanogaster* and *D. ananassae* can be compared (STEPHAN and LANGLEY 1989; LANGE *et al.* 1990; MIYASHITA and LANGLEY 1994; STEPHAN *et al.* 1994).

In applying the second approach for testing predictions of the hitchhiking and background selection models, we cloned and sequenced the *yellow* gene in *D. subobscura* as the recombination rate of the *y* gene region differs between this species and *D. melanogaster*. As stated above, in *D. melanogaster*, the *y* gene maps very close to the telomere where recombination is strongly supressed. In contrast, in *D. subobscura* it maps at section *2B* of the *X* chromosome (SEGARRA *et al.* 1995). Although data on recombination rates in the latter species are sparse, this chromosomal region is separated from the centromere by more than one euchromatic section and, thus, a strong reduction of the recombination rate is not expected. In fact, in *D. melanogaster* the strong supression of recombination is limited to centromeric and, at least, certain telomeric regions and to the entire fourth chromosome. The reductional effect on recombination decreases as the distance from the centromere or telomere increases and, for instance, section 19 of *D. melanogaster* (one chromosomal section from the centromere) has recombination rates similar to those of other sections in the middle of the *X* chromosome (KLIMAN and HEY 1993). Therefore, there is no reason to suspect that recombination is supressed in the *y* gene region of *D. subobscura*, although accurate measures of recombination rates are not available in this species. Divergence of the *y* gene between *D. subobscura* and *D. melanogaster* was compared to divergence at other genes with similar (or also different) rates of recombination in the two species.

MATERIALS AND METHODS

The λ*Dsub*RA111 random genomic library of a *D. subobscura* strain from Raíces (Canary Islands) was screened according to BENTON and DAVIES (1977). Two fragments that partially contain the *y* gene of *D. melanogaster* were used as probes: a 3.1-kb *Bgl*II-*Cla*I fragment that includes the first exon and a 2.2-kb *Cla*I-*Bgl*II fragment that includes the complete second exon (GEYER and CORCES 1987). After isolation of the positive recombinant phages, their DNA was purified (MANIATIS *et al.* 1982). These DNAs were biotin labeled by nick translation and *in situ* hybridized to *D. subobscura* and/or *D. melanogaster* polytene chromosomes as described by SEGARRA and AGUADÉ (1992).

Restriction map analysis, Southern blotting, isolation of two cross-hybridizing fragments and cloning into the pBluescript II SK$^+$ vector were performed as described by MANIATIS *et al.* (1982). A set of nested deletions for both orientations of each of the two recombinant clones was obtained according to HENIKOFF (1984). The insert size of the resulting subclones was determined after their amplification by PCR using the universal primers of the vector polylinker (KILGER and SCHMID 1994). After ordering the subclones by decreasing lengths, they were sequenced by the dideoxy method (SANGER *et al.* 1977). Sequences of the overlapping subclones were assembled using the STADEN's program (STADEN 1982). The complete sequences of both strands of the two isolated fragments were obtained. As these fragments did not overlap, two primers (one from each of the fragments) were designed to obtain the complete sequence of the *y* gene region. A 253-bp fragment was PCR amplified using the DNA of one of the positive recombinant phages as template. Both strands of this fragment were sequenced after isolation of single stranded DNA by exonuclease digestion (HIGUCHI and OCHMAN 1989).

Sequence analysis and alignment of the *y* gene region of *D. subobscura* and *D. melanogaster* (GEYER *et al.* 1986; GEYER and CORCES 1987) and of the deduced Yellow proteins were performed using the GAP and the BESTFIT programs of the Wisconsin GCG package (DEVEREUX *et al.* 1984). Nucleotide divergence in the *y* coding region was estimated by using the DIVERGEN program (COMERON 1994). This program gives the number of synonymous substitutions per synonymous site (p_s) and the number of nonsynonymous substitutions per nonsynonymous site (p_a) according to the unweighted method of NEI and GOJOBORI (1986), and computes the corresponding corrected estimates K_s and K_a, respectively, using different correction methods.

Codon bias in the *y* gene of *D. melanogaster* and *D. subobscura* was estimated by four different methods: the scaled χ^2, CBI, F_{op} and GC_{3s}. The scaled χ^2 (SHIELDS *et al.* 1988) is the sum of χ^2 values obtained within each class of synonymous codons assuming equal codon usage divided by the number of codons analyzed; this measure was computed without applying Yates' correction. The codon bias index (CBI) proposed by MORTON (1993) is based on the relative use of codons within synonymous class. The relative use of a codon in a gene is calculated as the ratio between the number of ocurrences of that codon and the number of ocurrences of the most frequently used synonymous codon. Both, the scaled χ^2 and CBI methods measure departure from the random use of codons but do not take into account which codons are used preferentially. F_{op} gives the frequency of preferred or optimal codons in a gene (IKEMURA 1985) and was computed by using the CODONS program (LLOYD and SHARP 1992). As preferred codons for *D. subobscura* have not been established, those of *D. melanogaster* were considered when computing F_{op} in *D. subobscura* genes. In fact, when grouping different genes sequenced in *D. subobscura*, the codons used preferentially for a particular amino acid in *D. subobscura* agree with those identified as preferred codons in *D. melanogaster* (AKASHI 1995). There is only a partial deviation in the valine class of synonymous codons. In *D. melanogaster* GTC and GTG are classified as preferred codons though significance is higher for GTC. In the analyzed genes of *D. subobscura*, GTC and GTG are also used preferentially, but GTG exhibits a higher frequency of

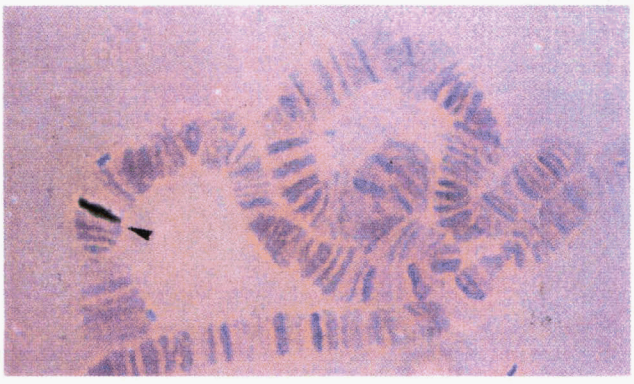

FIGURE 1.—*In situ* hybridization of λ*Dsub*RA4.1 on polytene chromosomes of *D.subobscura*. Hybridization signal at section *2B* of the *X* chromosome is shown by an arrowhead.

usage. Finally, GC_{3s} gives the G+C content for the third-codon positions that are synonymously variable. This is an indirect estimate strongly correlated with codon bias in *D. melanogaster* (SHIELDS *et al.* 1988) since in this species preferred codons end mostly in C though some end in G. GC_{3s} was also computed using the CODONS program (LLOYD and SHARP 1992). Codon usage for a particular gene was compared between *D. melanogaster* and *D. subobscura* by the distance measure (*D*) proposed by MORTON (1993). This measure is used for pairwise comparison of relative codon usage tables and was computed for a single gene as the sum over all codons of the difference (in absolute value) between the two species in the relative use of codons in that gene.

RESULTS

Three partially overlapping recombinant phages (λ*Dsub*RA2.1, λ*Dsub*RA3.1 and λ*Dsub*RA4.1) that putatively contained the *y* gene of *D. subobscura* were isolated. After *in situ* hybridization on *D. subobscura* polytene chromosomes, the three phages were mapped at section *2B* of the *X* chromosome (Figure 1), thereby corroborating that they were true positives. A 4.2-kb *Bam*HI fragment from λ*Dsub*RA4.1 and a 3.1-kb *Bam*HI fragment from λ*Dsub*RA2.1 were isolated, cloned and sequenced. These fragments are adjacent to each other in the genome and, thus, the complete sequenced region, which includes the *y* gene of *D. subobscura,* has a length of ~7.3 kb. This sequence is deposited in the EMBL data library under accession number Y13909.

The complete sequence of *D. subobscura* was compared with that of *D. melanogaster* (GEYER *et al.* 1986), which allowed identification of the putative coding region. The 7.3-kb sequenced region includes 1969 bp of the 5′ flanking region, the first exon (259 bp), a large intron with a length of 3238 bp, the second exon (1445 bp) and 446 bp of the 3′ flanking region. The general organization of the *y* gene is conserved between *D. subobscura* and *D. melanogaster*. The location of the intron in the coding region is identical in both species; however, a considerable difference in its length was detected (3238 bp in *D. subobscura vs.* 2721 bp in *D. melanogaster*).

The high number of nucleotide substitutions and indels accumulated in flanking regions and in the intron during divergence makes their alignment uncertain. However, some conserved stretches of nucleotides were identified in the 5′ flanking region. These conserved elements and their relative location are indicated in Figure 2. The conserved element II, which contains an imperfect palindrome, may be part of the putative enhancer identified by GEYER and CORCES (1987) and MARTIN *et al.* (1989) and roughly located by deletion analysis between positions −1868 and −700 in *D. melanogaster*. Moreover, the conserved element III may be related to the sequence also described by these authors in *D. melanogaster* spanning between −294 and −92 in *D. melanogaster,* which affects *yellow* gene expression in larvae. Element IV is the furthest downstream element conserved in the 5′ flanking region (Figure 2). This element contains the TATA box and transcription initiation site that are maintained in both species and are included in a longer conserved sequence. It is noteworthy that between sites −1620 and −1314 of the *D. subobscura* 5′ flanking region the trinucleotide CAG is repeated with an unusually high frequency (44%). The longest stretch of adjacent CAGs includes 10 repeats of this trinucleotide (Figure 3). CAA/CAG repeats are known as opa-boxes (WHARTON *et al.* 1985) and form part of different genes implicated in development in which they code for stretches of glutamine residues. The CAG-rich region of *D. subobscura* does not seem to be included within any other coding region as it is interspersed with inframe stop codons; however, this possibility cannot be completely discarded. A similar CAG-rich region has not been detected in *D. melanogaster*. Apart from the conserved elements in the 5′ flanking region, some conserved stretches of nucleotides were also identified in the large intron (Figure 4). In the 3′ flanking region the polyadenylation signal (AATAAA) was identified but no other conserved long stretches of nucleotides were detected.

Alignment of the *D. subobscura* and *D. melanogaster* *yellow* gene coding regions is presented in Figure 5. The putative coding region of *D. subobscura* encodes a protein 568 amino acids long, whose complete sequence is also shown in Figure 5. This protein has 27 residues more than that of *D. melanogaster*. The difference in length of both proteins is due to seven indels of sizes varying from one to 18 amino acids. Apart from this difference in length, a total of 54 amino acid replacements were detected, resulting in a protein identity of ~90%. The two putative N-linked glycosylation sites and the putative signal peptide described in the Yellow protein of *D. melanogaster* (KORNEZOS and CHIA 1992) are conserved in *D. subobscura*.

Estimates of the number of substitutions per site ac-

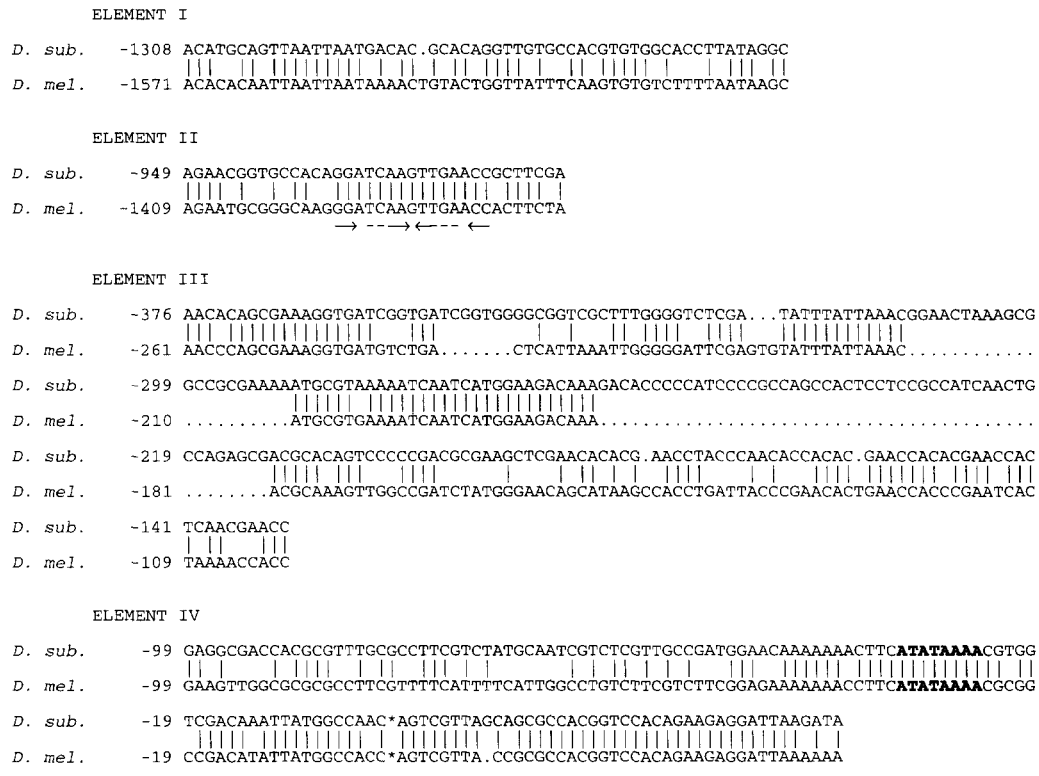

FIGURE 2.—Conserved elements in the 5′ flanking region of the *yellow* gene between *D. subobscura* (*D. sub.*) and *D. melanogaster* (*D. mel.*). Arrows in element II indicate a palindromic motif. Element IV includes the putative TATA box (in bold) and the transcription initiation site (indicated by an asterisk). Numbers are given considering the first transcribed nucleotide in the *D. melanogaster* sequence as position +1 (GEYER *et al.* 1986). Dots indicate missing nucleotides.

cumulated in the *y* coding region during the divergence of *D. subobscura* and *D. melanogaster* were obtained separately for synonymous and nonsynonymous sites. The number of synonymous substitutions per synonymous site is high (p_s = 0.8282). Divergence in the *y* coding region was compared with that found in a further 17 genes sequenced in both *D. subobscura* and *D. melanogaster* (Table 1). Data on the chromosomal location of these genes in both species are available except for *Sod*, which has not been mapped in *D. subobscura*. The close linkage between the *yellow* and *scute* genes in *D. melanogaster* is conserved in *D. subobscura* (BOTELLA *et al.* 1996) and, thus, the recombination rate for the *sc* region also differs between both species. None of the remaining genes is located in a chromosomal region that presumably presents a strong reduction in the recombination rate (as telomeric and centromeric regions) in either *D. subobscura* or *D. melanogaster*. The number of synonymous substitutions per site varies considerably among the different genes. For these other genes p_s values range from 0.3927 in *rp49* to 0.7919 in *sc*. Therefore, *y* and *sc* exhibit the highest number of synonymous substitutions per site. In fact, substitutions at synonymous sites are in both cases saturated and any method of correction is inapplicable or unreliable. A 2 × 2 independence test indicates that the number of synonymous substitutions does not differ significantly between *y* and *sc* (χ^2 = 1.504, 1 d.f., $P > 0.1$). Otherwise, when *Adhr* (as the third gene with a highest p_s value) is compared with *y* or *sc*, this number is significantly higher (χ^2 = 20.668, 1 d.f., $P < 0.005$ and χ^2 = 8.993, 1 d.f., $P < 0.01$, respectively).

In contrast to the high p_s value, the number of nonsynonymous substitutions per nonsynonymous site is rather low in the *y* gene (p_a = 0.0653). The corrected estimate, according to JUKES and CANTOR (1969), is K_a = 0.0683. K_a estimates for the other genes sequenced in *D. subobscura* and *D. melanogaster* range from 0.0285 for *rp49* to 0.4050 for the *chorion protein* gene *Cp15* (Table 1). Only *Adh, Adhr, Antp, rp49* and *Sxl* are more conserved at nonsynonymous positions than *y*.

To understand the high rate of synonymous substitutions per site detected in the *y* gene, its codon bias was estimated in *D. subobscura* and *D. melanogaster*, and compared with codon bias estimates for the other genes sequenced in both species. Codon bias was estimated

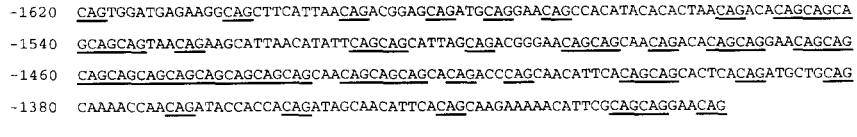

FIGURE 3.—Nucleotide sequence of the CAG-rich fragment detected in the 5′ flanking region of the *yellow* gene of *D. subobscura*. All CAG trinucleotides present in the sequence are underlined. Numbers are given as in Figure 2.

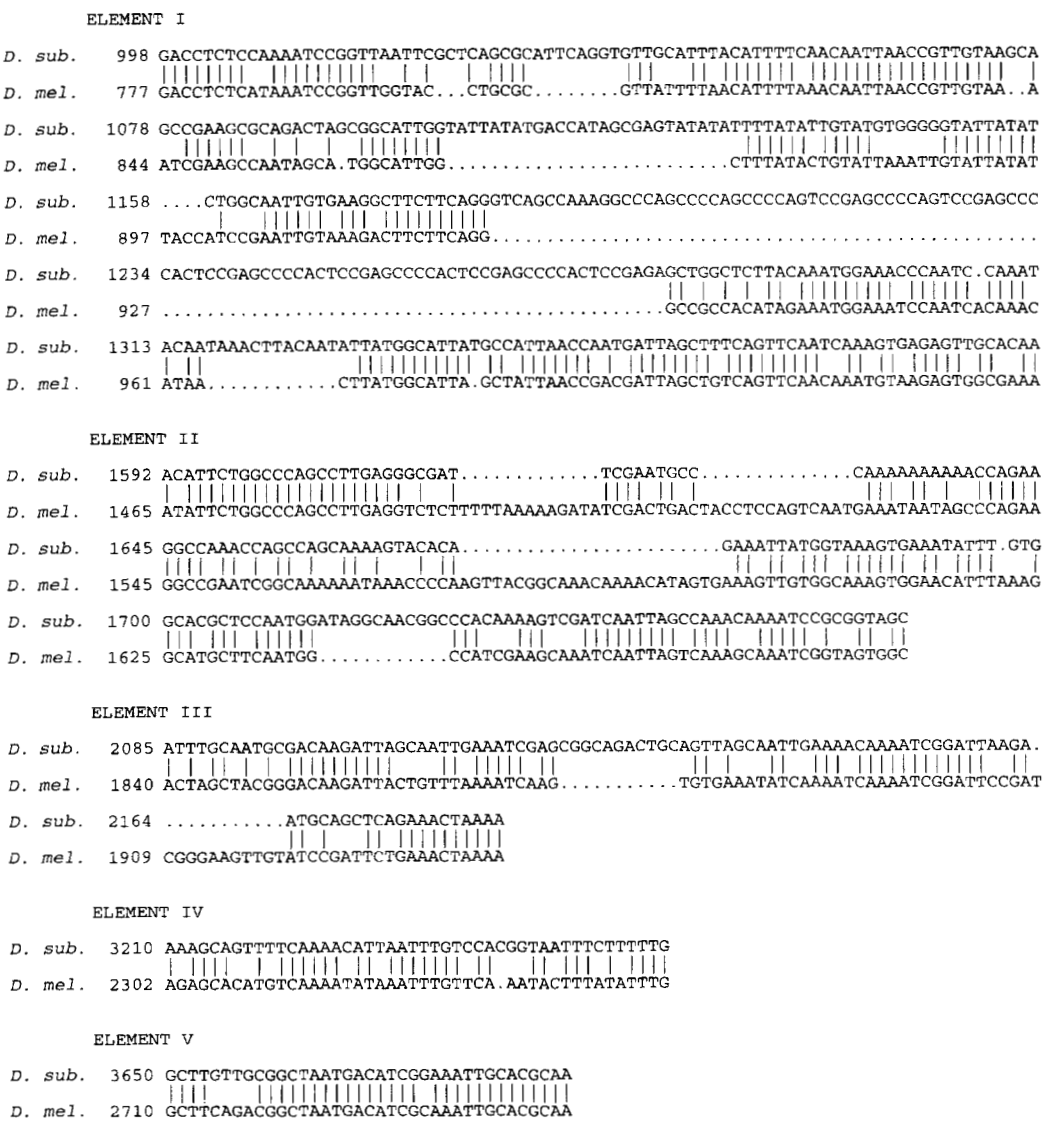

FIGURE 4.—Conserved elements in the intron of the *yellow* gene between *D. subobscura* (*D. sub.*) and *D. melanogaster* (*D. mel.*). Their relative position is indicated as in Figure 2. Dots indicate missing nucleotides.

as scaled χ^2, CBI, F_{op} and GC_{3s}. Table 2 shows these codon bias estimates, and for each gene the ratio between the *D. subobscura* and *D. melanogaster* estimates is also shown. The *y* gene presents the highest ratio for F_{op} and GC_{3s} and only *sc* exhibits a higher scaled χ^2 and CBI ratio than *y*.

The *y* and *sc* genes have, therefore, a much higher codon bias in *D. subobscura* than in *D. melanogaster*. This difference in codon bias causes *y* and *sc* to exhibit the highest *D* values (Table 2), a measure of codon usage difference between two species when comparing the relative use of codons for a particular gene (MORTON 1993). Moreover, the number of residues for each amino acid is relatively high due to the length of the Yellow protein. For this reason, equal codon usage for the *y* gene between *D. subobscura* and *D. melanogaster* was analyzed independently for each class of synonymous codons by χ^2 contingency tables. Sixteen out of 18 comparisons were significant at the $P < 0.05$ level. Only the usage of synonymous codons for cysteine and for phenylalanine did not differ significantly between both species ($P > 0.05$).

Codon bias is caused by the preferential use of particular synonymous codons. To relate synonymous substitutions to codon usage, third-codon positions that present a synonymous substitution between *D. subobscura* and *D. melanogaster* have been further analyzed considering whether they are included in a preferred or unpreferred codon in *D. subobscura* and/or *D. melanogaster*. A total of 189 out of 262 (72%) of these substituted sites indicate the use of an unpreferred codon in *D. melanogaster* and the use of a preferred codon in *D. subobscura*. Otherwise, only 11% of these sites are part of a preferred codon in the former species and of an unpreferred codon in the latter. The remaining sites are included in both species in a preferred (14%) or in an unpreferred codon (5%). This result indicates that most of these third-codon positions synonymously substituted involve the use of an unpreferred codon in *D. melanogaster* and the use of a preferred codon in *D. subobscura*. Therefore, the different

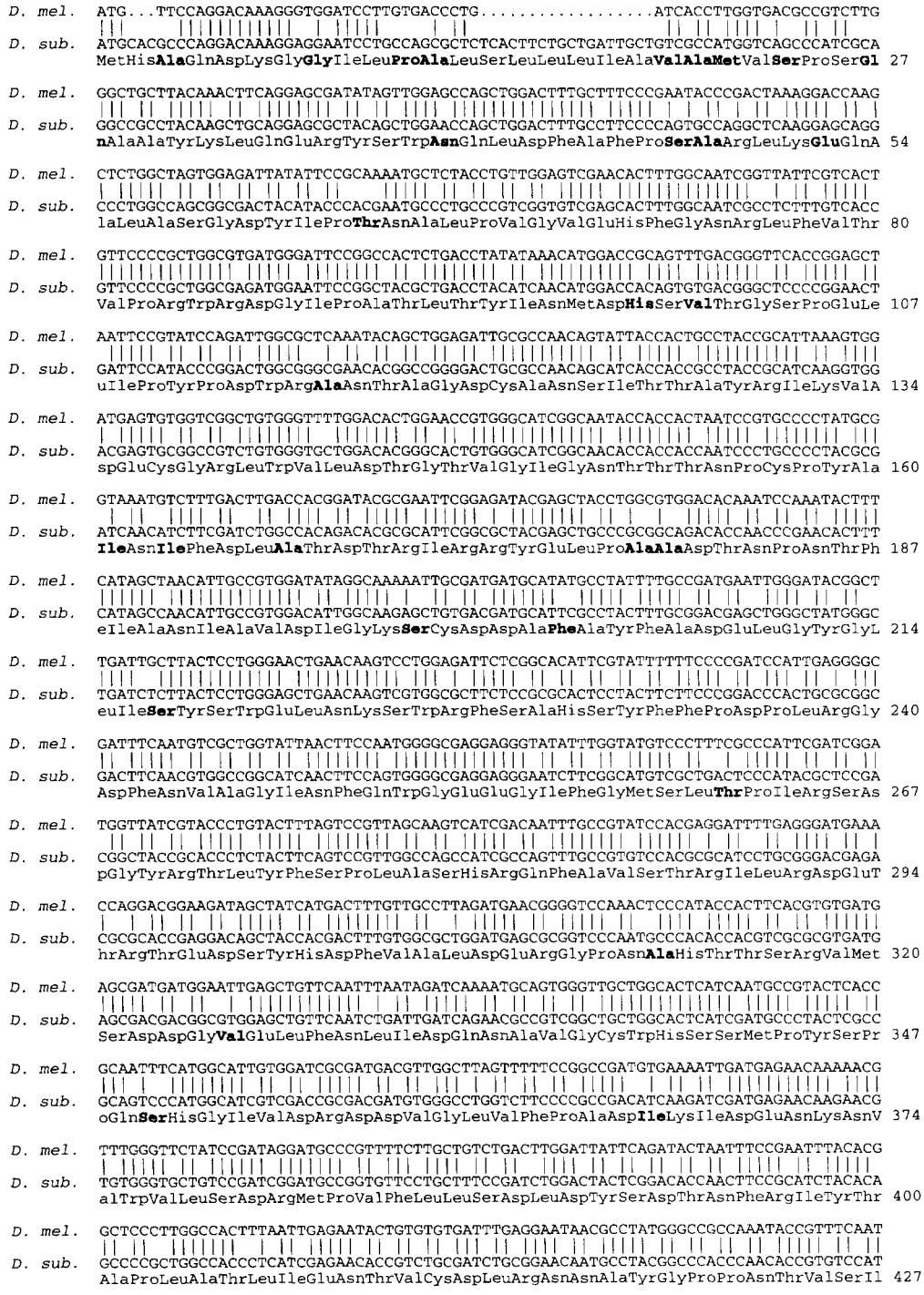

FIGURE 5.—Sequence alignment of the *yellow* gene coding region of *D. melanogaster* (*D. mel.*) (GEYER *et al.* 1986) and *D. subobscura* (*D. sub.*). The complete sequence of the Yellow protein of *D. subobscura* is also indicated under the nucleotide sequence of this species. Amino acid replacements between both species are shown in bold. Dots in nucleotide or amino acid sequences indicate missing nucleotides or amino acids, respectively. The first nucleotide of the ATG initiation codon is at position 172 in the *D. melanogaster* sequence. Numbers correspond to amino acid residues in the *D. subobscura* Yellow protein.

codon bias detected in the *y* gene of both species can be related to the synonymous substitutions accumulated during their divergence.

A negative correlation between K_s and codon bias has been detected for genes sequenced in *D. melanogaster* and species of the obscura group (SHARP and LI 1989). When applying the nonparametric tests of KENDALL and SPEARMAN, a significant ($P < 0.02$) negative association was also detected, for the 18 genes included in Table 2, between p_s and mean codon bias of *D. subobscura* and

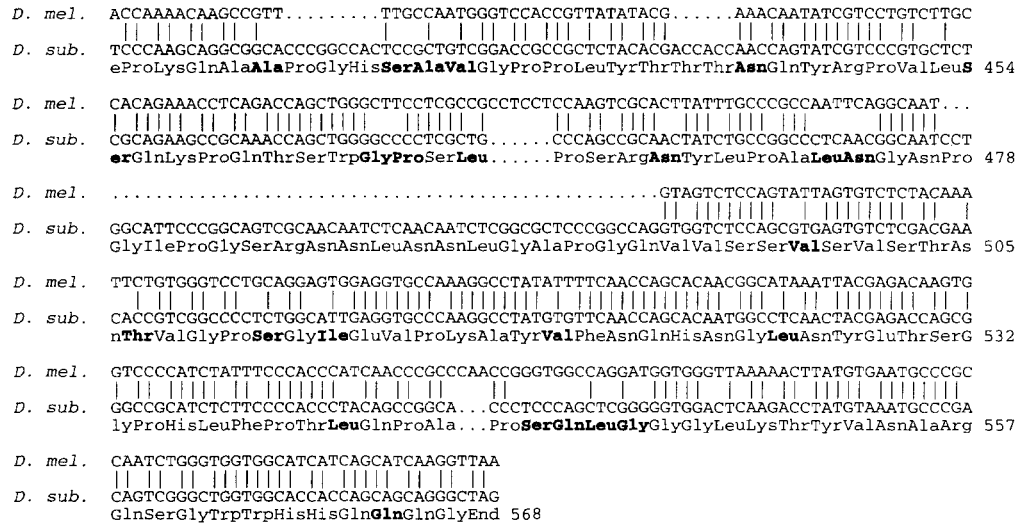

FIGURE 5.—*Continued*

D. melanogaster measured as scaled χ^2 and CBI. If the negative correlation is due to selection acting to maintain codon bias in both species preventing substitutions at synonymous sites, a positive correlation between codon bias of *D. subobscura* and *D. melanogaster* is expected. Significant ($P < 0.05$) or marginally significant ($P < 0.1$) KENDALL and SPEARMAN's coefficients of rank correlation were detected between codon bias (measured as CBI, scaled χ^2 or F_{op}) in *D. subobscura* and *D. melanogaster* but only when *y* and *sc* were excluded from the

TABLE 1

Divergence in different genes sequenced in *D. subobscura* and *D. melanogaster*

Gene[a]	n^b	p_s^c	K_s^d	K_a^e
Adh	253	0.4887 (0.0360)	0.7909 (0.1035)	0.0473 (0.0094)
Adhr	271	0.6595 (0.0347)	1.5860 (0.2879)	0.0586 (0.0099)
Antp	377	0.5047 (0.0319)	0.8383 (0.0975)	0.0218 (0.0050)
bcn92	91	0.5750 (0.0608)	1.0916 (0.2607)	0.0816 (0.0207)
Cp15	107	0.4534 (0.0527)	0.6957 (0.1332)	0.4050 (0.0523)
Cp16	136	0.5011 (0.0470)	0.8273 (0.1416)	0.1675 (0.0260)
Cp18	165	0.5262 (0.0427)	0.9069 (0.1430)	0.2378 (0.0292)
Cp19	161	0.4738 (0.0437)	0.7492 (0.1188)	0.2238 (0.0283)
rp49	133	0.3927 (0.0500)	0.5561 (0.1049)	0.0285 (0.0098)
sc	331	0.7919 (0.0263)	Inapplicable	0.2811 (0.0224)
Sod	114	0.5072 (0.0557)	0.8460 (0.1722)	0.0771 (0.0179)
Sry-α	521	0.6329 (0.0250)	1.3926 (0.1600)	0.2598 (0.0170)
Sry-β	341	0.5809 (0.0333)	1.1172 (0.1478)	0.1461 (0.0145)
Sry-δ	414	0.6484 (0.0294)	1.4990 (0.2168)	0.1289 (0.0122)
Sxl	342	0.3832 (0.0309)	0.5364 (0.0632)	0.0452 (0.0078)
Xdh	1333	0.6020 (0.0156)	1.2170 (0.0791)	0.0760 (0.0052)
y	537	0.8282 (0.0190)	Inapplicable	0.0683 (0.0078)
zen	343	0.6095 (0.0316)	1.2564 (0.1685)	0.1961 (0.0175)

[a] Accession numbers in EMBL/GenBank DNA sequence library or references are as follows: *Adh* and *Adhr* (*mel*: KREITMAN 1983 and KREITMAN and HUDSON 1991, respectively, *sub*: M55545), *Antp* (*mel*: X03790 and X03791, *sub*: X60995 and X62246), *bcn92* (*mel*: Z46608 and M80598, *sub*: Z46522), *Cp15, Cp16, Cp18 and Cp19* (*mel*: X02497 and X16715, *sub*: X12637), *rp49* (*mel*: X00848, *sub*: AGUADÉ 1988), *scute* (*mel*: M17119, *sub*: BOTELLA *et al.* 1996), *Sod* (*mel*: M24421, *sub*: S72444), *Sry-α-β-δ* (*mel*: X03121, *sub*: L19535 and FERRER *et al.* 1994), *Sxl* (*mel*: M59447, *sub*: PENALVA *et al.* 1996), *Xdh* (*mel*: Y00308, *sub*: COMERON and AGUADÉ 1996) and *zen* (*mel*: X68347, *sub*: TEROL *et al.* 1995). *mel* and *sub* refer to *D. melanogaster* and *D. subobscura*, respectively.

[b] Number of codons compared.

[c] Number of synonymous substitutions per synonymous site. Standard errors are in parentheses.

[d] p_s value corrected according to JUKES and CANTOR (1969). Standard errors are in parentheses.

[e] Number of nonsynonymous substitutions per nonsynonymous site corrected according to JUKES and CANTOR (1969). Standard errors are in parentheses.

TABLE 2

Codon bias in different genes sequenced in *D. subobscura* and *D. melanogaster*

	Scaled χ^2		CBI		F_{op}		GC_{3s}		
Gene[a]	sub[b]/mel[c]	Ratio[d]	sub/mel	Ratio	sub/mel	Ratio	sub/mel	Ratio	D^e
Adh	0.57/1.06	0.53	0.45/0.69	0.65	0.60/0.75	0.83	0.67/0.81	0.83	11.71
Adhr	0.34/0.19	1.82	0.39/0.23	1.66	0.49/0.39	1.26	0.58/0.53	1.09	13.90
Antp	0.46/0.41	1.12	0.46/0.46	1.02	0.59/0.61	0.97	0.73/0.75	0.97	10.77
bcn92	0.66/0.68	0.97	0.47/0.59	0.80	0.55/0.64	0.86	0.67/0.75	0.89	15.36
Cp15	0.88/0.76	1.16	0.56/0.63	0.90	0.54/0.63	0.86	0.60/0.65	0.82	9.71
Cp16	0.81/0.90	0.89	0.58/0.63	0.93	0.60/0.65	0.92	0.71/0.69	1.03	14.14
Cp18	0.55/0.80	0.68	0.51/0.66	0.77	0.56/0.64	0.87	0.61/0.70	0.87	9.52
Cp19	0.75/1.00	0.74	0.62/0.73	0.85	0.63/0.70	0.90	0.70/0.74	0.95	6.88
rp49	0.87/1.37	0.63	0.60/0.78	0.77	0.72/0.78	0.92	0.74/0.84	0.88	8.69
sc	0.74/0.12	6.17	0.64/0.19	3.30	0.64/0.40	1.60	0.83/0.51	1.63	21.16
Sod	0.75/0.72	1.04	0.61/0.57	1.07	0.66/0.66	1.00	0.76/0.74	1.03	8.55
Sry-α	0.33/0.42	0.78	0.31/0.43	0.72	0.49/0.57	0.86	0.66/0.70	0.94	13.76
Sry-β	0.48/0.77	0.62	0.48/0.68	0.70	0.62/0.73	0.85	0.74/0.88	0.84	12.32
Sry-δ	0.31/0.85	0.36	0.38/0.71	0.54	0.56/0.77	0.73	0.69/0.88	0.78	13.31
Sxl	0.39/0.30	1.30	0.33/0.32	1.05	0.50/0.49	1.02	0.60/0.59	1.02	9.44
Xdh	0.66/0.28	2.36	0.53/0.35	1.53	0.63/0.53	1.19	0.77/0.64	1.20	10.75
y	0.75/0.21	3.67	0.60/0.29	2.06	0.68/0.34	2.00	0.82/0.46	1.78	25.01
zen	0.46/0.31	1.48	0.44/0.38	1.14	0.52/0.54	0.96	0.68/0.65	1.05	11.96

[a] See Table 1 for data sources.
[b] *D. subobscura.*
[c] *D. melanogaster.*
[d] Ratio of the codon bias estimate between *D. subobscura* and *D. melanogaster.*
[e] Distance measure for comparisons between *D. subobscura* and *D. melanogaster* of the relative use of codons as proposed by MORTON (1993).

analysis. This result indicates that *y* and *sc* deviate from the expected trend exhibited by the other genes and further supports that the high p_s value of *y* and *sc* is not due to a low codon bias in both species but to the strong difference in codon bias between them.

DISCUSSION

The divergence time of the *melanogaster* and *obscura* groups of Drosophila has been estimated to be 30 million years (THROCKMORTON 1975). During this period of time, an extensive reorganization within the *X* chromosome occurred (SEGARRA and AGUADÉ 1992; SEGARRA *et al.* 1995). This reorganization has affected the *y* and *sc* genes, which are located very close to the telomere in *D. melanogaster* and more than one euchromatic chromosomal section from the centromere in *D. subobscura.* Although the ancestral location of these genes is unknown, their telomeric position in the melanogaster lineage can be traced back to at least six to 15 million years. This is the estimated divergence time of the species included in the *melanogaster* subgroup (LACHAISE *et al.* 1988), which present a similar banding pattern of the distal part of their polytene *X* chromosome (ASHBURNER 1989; F. LEMENIEUR, personal communication). Likewise, the nontelomeric position of *y* and *sc* in the *obscura* lineage is shared by *D. subobscura* and *D. pseudoobscura* (SEGARRA *et al.* 1995 and C. SEGARRA, unpublished result) and, thus, can be traced back to eight to 12 million years, the estimated divergence time between these two species in the obscura lineage (SEGARRA and AGUADÉ 1992; RAMOS-ONSINS *et al.* 1997). The chromosomal location of the *y* and *sc* genes in the *melanogaster* lineage is associated with a strong reduction in the recombination rate of both genes. This reduction due to the telomeric effect is not expected in the *obscura* lineage. Therefore, the study of the divergence in the *y* and *sc* genes between *D. subobscura* and *D. melanogaster* allows inferences to be made on the molecular evolution of two genomic regions that have differed in their recombination rate between the *melanogaster* and *obscura* lineages for a long period of time.

The pattern of divergence in the *y* and *sc* genes has peculiar characteristics that are not shared by the other genes sequenced in *D. subobscura* and *D. melanogaster.* First, p_s values for *y* and *sc* are sufficiently high that synonymous substitutions have saturated in both genes (Table 1). Secondly, codon bias in the *y* and *sc* genes is much higher in *D. subobscura* than in *D. melanogaster* (Table 2).

The high divergence at synonymous sites would *a priori* be in agreement with an increase in the fixation rate of synonymous substitutions in the *y* and *sc* genes in the *D. melanogaster* and/or the *D. subobscura* lineages.

A relative rate test (SARICH and WILSON 1973) might be useful in contrasting these alternatives. However, this approach cannot be carried out at present since the sequences of the *y* and *sc* genes in an outgroup-related species are not available.

The major differences in codon bias detected in the *y* and *sc* genes between *D. subobscura* and *D. melanogaster* cannot easily be explained by proposals to account for differences in codon bias among genes. A direct relationship between the level of gene expression and codon bias was suggested by SHIELDS *et al.* (1988) in *D. melanogaster* genes. Later AKASHI (1994) proposed that codon bias in highly expressed genes would enhance the accuracy of protein synthesis. Although the possibility that *y* and *sc* are under different constraints in the two species cannot be discarded, it seems very unlikely that both *y* and *sc* genes have a higher expression level and/or that they require a more efficient and accurate translation in *D. subobscura* than in *D. melanogaster.*

According to the nearly neutral theory of molecular evolution (reviewed by OHTA 1992), the differing codon bias in the *y* and *sc* genes in *D. subobscura* and *D. melanogaster* may also be due to a difference in the effective size (N_e) between both species provided that synonymous mutations are nearly neutral. In fact, AKASHI (1995) inferred the action of weak selection acting against mutations that cause changes from preferred to unpreferred codons and, thus, that these particular synonymous mutations can be considered slightly deleterious. Consequently, the low codon bias in *D. melanogaster* could be explained if this species had a smaller effective population size than *D. subobscura.* However, were this the case, the same deviation in codon usage should be detected in the other genes sequenced in both species, but this is not supported by the results presented in Table 2.

A reduction in effective size is also expected for sex-linked genes. Tables 1 and 2 include three *X*-linked genes: *y, sc* and *Sxl. Sxl* exhibits the lowest p_s value among the 18 genes and a similar codon bias in both species (Table 2). Thus, its pattern of divergence differs from *y* and *sc* indicating that sex-linkage of these two genes cannot account for the observed results. In fact, the reduced effective size of sex-linked genes would imply a lower codon bias in these genes but in both species. This, however, is not consistent with the strong differences in codon bias detected in *y* and *sc.*

Therefore, to explain the low codon bias of *y* and *sc* in *D. melanogaster,* the reduction in N_e has to have a differential effect on these two genes relative to the other genes (autosomic or sex-linked) included in Tables 2 and 3. This differential reduction in N_e is consistent with the location of *y* and *sc* in a region with a strong reduction in the recombination rate in *D. melanogaster.* The hitchhiking and background selection models predict an increase in the fixation rate of slightly deleterious mutations and a decrease in the fixation rate of slightly advantageous mutations in regions of reduced recombination. Synonymous changes from preferred to unpreferred codons can be considered slightly deleterious while reverse changes from unpreferred to preferred codons can be considered slightly advantageous (AKASHI 1995). Therefore, a higher fixation rate of slightly deleterious mutations (from preferred to unpreferred codons) and a lower fixation rate of slightly advantageous mutations (from unpreferred to preferred codons) are expected in the *D. melanogaster* than in the *D. subobscura* lineage for the *y* and *sc* genes. This difference in fixation rates would account for the strong difference in codon bias detected for *y* and *sc* between *D. melanogaster* and *D. subobscura,* which has as a consequence the high p_s value that both genes exhibit between these species.

According to this argument, a negative association between p_s and r values is expected. The lack of accurate measures of recombination rates in *D. subobscura* prevents performance of this analysis. However, when considering only recombination rates in *D. melanogaster* (KLIMAN and HEY 1993) a significant negative association between p_s and r was detected by the coefficient of rank correlation of KENDALL ($\tau = 0.379$, $P = 0.0207$) and SPEARMAN ($\rho = 0.513$, $P = 0.0284$). This significant increase of p_s values in regions with a strong reduction in the recombination rate is not consistent with previous results based on gene divergence between *D. melanogaster* and *D. simulans* (BEGUN and AQUADRO 1992; HILTON *et al.* 1994). We also failed to detect any association between K_s and r in 30 genes sequenced in these two species (result not shown). The most pausible explanation for this disagreement may be that the elapsed time since the divergence between *D. melanogaster* and *D. simulans* has not been long enough to detect the putative increase of synonymous substitutions in regions with no recombination.

Furthermore, the different rate of fixation of non-neutral mutations in regions of reduced recombination may also affect divergence at nonsynonymous sites provided that they are not neutral. The number of nonsynonymous substitutions per site between *D. subobscura* and *D. melanogaster* is moderate in the *y* gene but rather high in *sc.* However, the rate of nonsynonymous substitutions varies considerably among genes due to differences in the functional constraints on the encoded proteins. Therefore, to detect the putative effect of recombination on the fixation rate of nonsynonymous mutations, comparisons using an outgroup species are required, which would allow a relative rate test to be applied. Such studies may contribute to establishing if the rate of nonsynonymous substitutions decreases or increases in regions with no recombination.

We thank V. C. CORCES for the *y* clones of *D. melanogaster.* We also thank W. STEPHAN and J. ROZAS for critical comments on the

manuscript. This work was supported by a predoctoral fellowship from Ministerio de Educación y Ciencia, Spain, to A.M., and by grants PB94-923 from Comisión Interdepartamental de Ciencia y Tecnología, Spain, and 1995SGR-577 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Catalonia, Spain to M.A.

LITERATURE CITED

AGUADÉ, M., 1988 Nucleotide sequence comparison of the *rp49* gene region between *Drosophila subobscura* and *D. melanogaster*. Mol. Biol. Evol. **5:** 433–441.

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

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.

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

BEGUN, D. J., and C. F. AQUADRO, 1991 Molecular population genetics of the distal portion of the *X* chromosome in Drosophila: evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics **129:** 1147–1158.

BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **365:** 548–550.

BENTON, W. D., and R. W. DAVIS, 1977 Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196:** 180–183.

BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

BIRKY, C. W. JR., and J. B. WALSH, 1988 Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85:** 6414–6418.

BOTELLA, L. M., C. DOÑORO, L. SÁNCHEZ, C. SEGARRA and B. GRANADINO, 1996 Cloning and characterization of the *scute* (*sc*) gene of *Drosophila subobscura*. Genetics **144:** 1043–1051.

CHARLESWORTH, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63:** 213–227.

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

COMERON, J. M., 1994 Program for estimating the number of nucleotide substitutions using different methods. J. Hered. **85:** 493–494.

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.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DUBININ, N. P., N. N. SOKOLOV and G. G. TINIAKOV, 1937 Crossing over between the genes "yellow" and "scute." Dros. Inf. Serv. **8:** 76.

FERRER, P., M. CROZATIER, C. SALLES and A. VINCENT, 1994 Interspecific comparison of *Drosophila serendipity δ* and *β*: multimodular structure of these C_2 H_2 zinc finger proteins. J. Mol. Evol. **38:** 263–273.

GEYER, P. K., and V. C. CORCES, 1987 Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the *yellow* locus in *D. melanogaster*. Genes Dev. **1:** 996–1004.

GEYER, P. K., C. SPANA and V. C. CORCES, 1986 On the molecular mechanism of gypsy-induced mutations at the *yellow* locus of *Drosophila melanogaster*. EMBO J. **5:** 2657–2662.

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

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

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

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

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

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–120 in *Mammalian Protein Metabolism*, edited by H. W. MUNRO. Academic Press, New York.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

KILGER, C., and K. SCHMID, 1994 Rapid characterization of bacterial clones by microwave treatment and PCR. Trends Genet. **10:** 149.

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

KORNEZOS, A., and W. CHIA, 1992 Apical secretion and association of the *Drosophila yellow* gene product with developing larval cuticle structures during embryogenesis. Mol. Gen. Genet. **235:** 397–405.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

LACHAISE, D., M-L. CARIOU, J. R. DAVID, F. LEMENIEUR, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22:** 159–225.

LANGE, B. W., C. H. LANGLEY and W. STEPHAN, 1990 Molecular evolution of *Drosophila* metallothionein genes. Genetics **126:** 921–932.

LANGLEY, C. H., J. MACDONALD, N. MIYASHITA and M. AGUADÉ, 1993 Lack of correlation between interspecific divergence and intraspecific polymorphism at the *suppressor of forked* region of *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **90:** 1800–1803.

LLOYD, A. T., and P. M. SHARP, 1992 CODONS: a microcomputer program for codon usage analysis. J. Hered. **83:** 239–240.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, New York.

MARTIN, M., Y. B. MENG and W. CHIA, 1989 Regulatory elements involved in the tissue-specific expression of the *yellow* gene of *Drosophila*. Mol. Gen. Genet. **218:** 118–126.

MARTÍN-CAMPOS, J. M., J. M. COMERÓN, N. MIYASHITA and M. AGUADÉ, 1992 Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *D. melanogaster*. Genetics **130:** 805–816.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

MIYASHITA, N. T., and C. H. LANGLEY, 1994 Restriction map polymorphism in the *forked* and *vermilion* regions of *Drosophila melanogaster*. Jpn. J. Genet. **69:** 297–305.

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.

NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

OHTA, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23:** 263–286.

PENALVA, L. O. F., H. SAKAMOTO, A. NAVARRO-SABATÉ, B. GRANADINO, C. SEGARRA *et al.*, 1996 Regulation of the gene *Sex-lethal*: comparative analysis between *Drosophila melanogaster* and *D. subobscura*. Genetics **144:** 1653–1664.

RAMOS-ONSINS, S., C. SEGARRA, J. ROZAS and M. AGUADÉ, 1997 Molecular and chromosomal phylogeny in the obscura group of

Drosophila inferred from sequences of the *rp49* gene region. Mol. Phylogenet. (in press).

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

SARICH, M. V., and A. C. WILSON, 1973 Generation time and genomic evolution in primates. Science **179:** 1144–1147.

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.

SEGARRA, C., E. R. LOZOVSKAYA, G. RIBÓ, M. AGUADÉ and D. L. HARTL, 1995 P1 clones from *Drosophila melanogaster* as markers to study the chromosomal evolution of Muller's A element in two species of the *obscura* group of *Drosophila.* Chromosoma **104:** 129–136.

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.

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 C. H. LANGLEY, 1989 Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121:** 89–99.

STEPHAN, W., and S. J. MITCHELL, 1992 Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae.* Genetics **132:** 1039–1045.

STEPHAN, W., V. S. RODRIGUEZ, B. ZHOU and J. PARSCH, 1994 Molecular evolution of the Metallothionein gene *Mtn* in the *melanogaster* species group: results from *Drosophila ananassae.* Genetics **138:** 135–143.

TEROL J., M. PEREZ-ALONSO and R. DE FRUTOS, 1995 Molecular characterization of the *zerknüllt* region of the Antennapedia complex of *D. subobscura.* Chromosoma **103:** 613–624.

THROCKMORTON, L. H., 1975 The phylogeny, ecology and geography of *Drosophila,* pp. 421–469 in *Handbook of Genetics,* Vol. 3, edited by R. C. KING. Plenum Press, New York.

WHARTON, K. A., B. YEDVOBNICK, V. G. FINNERTY and S. ARTAVANIS-TSAKONAS, 1985 opa: a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D. melanogaster.* Cell **40:** 55–62.

Communicating editor: A. G. CLARK