Intraspecific and Interspecific Variation at the *y-ac-sc* Region of *Drosophila simulans* and *Drosophila melanogaster*

Jesús M. Martín-Campos,* Josep M. Comerón,* Naohiko Miyashita†,[1] and Montserrat Aguadé*

**Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain, and* †*Laboratory of Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709*

Manuscript received August 19, 1991
Accepted for publication December 13, 1991

ABSTRACT

A 2.2-kb region including the *ac* gene of *Drosophila simulans* has been sequenced. Interspecific divergence between *Drosophila melanogaster* and *D. simulans* was estimated as 0.0695 and 0.0558 for silent and for all sites, respectively. Estimated silent site divergence for the *ac* region is comparable to that estimated for other regions of the genome between these species, indicating that silent sites of the *ac* region are not under significantly stronger functional constraint. Intraspecific variation in both species was also investigated. Restriction-site and length polymorphism in the *ac* region of *D. simulans* has been investigated for 103 *X* chromosome lines sampled from three natural populations in Spain using eight four-cutter restriction enzymes. Neither restriction-site nor length variation was detected in the three populations surveyed. In *D. melanogaster* restriction-site and length polymorphism in all major transcription units of the *y-ac-sc* region (23.1-kb region) has been studied using four four-cutter restriction enzymes for 245 *X* chromosome lines sampled from 10 natural populations (seven from Europe, two from North America and one from Japan). Fourteen restriction-site and 28 length polymorphisms were detected. There was some indication of population subdivision for North American *vs.* European samples of *D. melanogaster*. The frequency spectrum of restriction-site polymorphisms in European populations was skewed toward rarer frequencies than predicted by the neutral theory. Comparison of silent site variation at this telomeric region with that in the *Adh* 5′-flanking region showed a reduced level of heterozygosity in the *y-ac-sc* region. Since interspecific silent divergence is not reduced in the *y-ac-sc* region as compared to other regions, the reduction in standing levels of variation at this telomeric locus in both *D. simulans* and *D. melanogaster* is most easily explained by a hitchhiking effect of linked selected substitutions.

THE substitution of selectively favorable mutations will change the frequencies of linked neutral polymorphisms in the surrounding region through a hitchhiking effect (MAYNARD SMITH and HAIGH 1974; KAPLAN, HUDSON and LANGLEY 1989). A prediction of this hitchhiking effect is a reduction in the levels of standing variation in natural populations. This reduction is sensitive to levels of recombination, the frequency of selected substitutions and the magnitude of selection. The effect is greatest when the recombination rate is smaller than the selection coefficient. On the other hand, in the absence of selection the level of recombination influences only the variance in the amount of polymorphism but not the mean. Therefore, no systematic reduction in variation in regions of reduced crossing over per physical length would be expected under pure neutrality. In regions with restricted recombination (the *y-ac-sc* and *Zw* regions, located close to the telomere and base of the *X* chromosome in *D. melanogaster*, respectively, and the *f* and *v* regions, located close to the base of the *X* chromosome in *D. ananassae*), relatively low heterozygosity has been observed for at least some populations (AGUADÉ, MIYASHITA and LANGLEY 1989a; MIYASHITA 1990; STEPHAN and LANGLEY 1989). In contrast, regions of the *X* chromosome with "normal" levels of recombination (SCHAEFFER, LANGLEY and AQUADRO 1988; MIYASHITA and LANGLEY 1988; AGUADÉ, MIYASHITA and LANGLEY 1989b) show levels of polymorphism similar to those found at autosomal loci.

Reduced levels of variation in regions of the autosomes or the *X* chromosome could be simply due to greater constraints, *i.e.*, stronger purifying selection in that region. If purifying selection was the main force reducing polymorphism in "recombination-supressed" regions, a similar reduction in the interspecies divergence would be expected under neutrality (KIMURA 1983), because interspecific divergence should parallel intraspecific variation. In order to evaluate this possibility in the telomeric region of the *X* chromosome, interspecific divergence between *D. simulans* and *D. melanogaster* was studied by sequencing a region including the *ac* gene in *D. simulans* and

[1] Present address: Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Japan.

comparing this sequence with that of the same region in *D. melanogaster* (VILLARES and CABRERA 1987). The rate of silent site divergence in this region was compared to that observed in other regions of the genome.

Whether the reduced variation in the *y-ac-sc* region in *D. melanogaster* was due to hitchhiking or to increased functional constraint, a similar reduction in variation might be expected in *D. simulans*, where this region is also located at the tip of the *X* chromosome. For this expectation to hold under hitchhiking, the mutation rate to selectively favorable variants should be high enough and the recombination rate low enough that there has been at least one substitution of a closely linked initially rare variant in each lineage since their divergence. Four-cutter restriction-map variation in a region including the *ac* gene has been studied in three natural populations of *D. simulans* in order to document levels of variation.

Finally, four-cutter restriction map variation in all major transcription units of the *y-ac-sc* region has been surveyed in 10 natural populations of *D. melanogaster*. Our original conclusion of reduced polymorphism in the *y-ac-sc* region (AGUADÉ, MIYASHITA and LANGLEY 1989a) was challenged by several authors (BEECH and LEIGH-BROWN 1989; EANES, LABATE and AJIOKA 1989; MACPHERSON, WEIR and LEIGH-BROWN 1990). This survey augments the quantity of data available and corroborates our first result, that the amount of polymorphism is low. This survey also documents substantial population differentiation between European and North American populations that was also seen in some previous studies (EANES, LABATE and AJIOKA 1989) but not in others (BEECH and LEIGH-BROWN 1989). As expected under the hitchhiking model, interspecific divergence in the *ac* region is typical of other regions of normal crossing over.

MATERIALS AND METHODS

Fly stocks: One hundred and three *X* chromosomes of *D. simulans* were extracted from three Spanish populations (52 from Barcelona, 26 from La Rábida, Huelva, and 25 from Tenerife, Canary Islands) by crossing single wild-caught males to virgin females from an attached-*X* chromosomes strain (*XXY, y w*^{a}*;XY, ++*) and collecting males either from the F_1 or F_2 of that cross.

Two hundred and forty five *X* isochromosomal lines of *D. melanogaster* were independently extracted from 10 natural populations: seven from the Old World, two from North America and one from Japan (Figure 1). Lines from North America and Japan were those from MIYASHITA *et al.* (1986). Lines from Old World populations were extracted by crossing single wild-caught males (XB) or single virgin females from each isofemale line (all other populations) with the balancer stock FM7a.

Cloning and sequencing: A random genomic library of a *D. simulans* strain from Putah Creek, California, was screened using as probe a 2.2-kb *Eco*RI fragment from *D. melanogaster* (clone 2.1RR101). Two positive phages were isolated and DNA purified. After digestion with several enzymes and analysis by Southern blot, a 2.2-kb *Eco*RI fragment from one of the positives with no *Kpn*I, *Hin*dIII, *Sac*I, *Sal*I, *Xho*I and *Xba*I sites was chosen for subcloning into the *Eco*RI site of the plasmid vector pBlueScript SK$^+$. A set of nested deletions was obtained for each strand according to HENIKOFF (1984), using restriction enzymes *Hin*dIII and *Kpn*I for one of the strands, and *Xba*I and *Sac*I for the complementary strand. From each of the clones double-stranded DNA was obtained by a modification of the alkaline procedure (SAMBROOK *et al.* 1989). One of the strands was sequenced using the universal M13 primer, and the complementary strand using the reverse M13 primer. Sequencing was performed by the dideoxy chain termination method (SANGER, NICKLEN and COULSON 1977) using T7 polymerase. The final sequence was assembled using STADEN's (1982) programs. Each nucleotide was on average sequenced 3.13 times.

Restriction map analysis: Procedures were as described in KREITMAN and AGUADÉ (1986a) using eight tetranucleotide-recognizing enzymes for all *D. simulans* and for a subset of *D. melanogaster* lines (53 lines from XB; and all lines from NC, TX and JPN): *Alu*I, *Dde*I, *Hae*III, *Hha*I, *Sau*3AI, *Msp*I, *Sau*96I and *Taq*I. All other *D. melanogaster* lines were digested with only four enzymes: *Alu*I, *Hae*III, *Sau*3AI and *Taq*I.

Table 1 gives a summary of all regions analyzed in *D. melanogaster* that include all major transcription units and flanking regions of the *y-ac-sc* region. Six of these regions had been previously sequenced either completely or partially (see Table 1) which facilitates analysis of the data. The set of four restriction enzymes used allows detection of 9–10% of all nucleotide changes in these regions, and nearly all insertions/deletions, as fragments as small as 70–80 bp can be reliably scored.

D. simulans samples were probed only with the 2.2-kb *Eco*RI fragment that includes the *ac* gene of this species (clone 2.2RRsim).

RESULTS

Interspecific sequence divergence: Figure 2 shows the sequence of the 2.2-kb *Eco*RI fragment of *D. simulans* compared to the published sequence of *D. melanogaster* (VILLARES and CABRERA 1987). Length differences are found only in the 5′- and 3′-flanking regions. In comparison to the sequence of *D. simulans*, the *D. melanogaster* sequence shows 4 insertions and 9 deletions in the 5′-region, and 6 insertions and 8 deletions in the 3′-region. Size differences ranged between 1 and 23 bp. Table 2 gives a summary of nucleotide changes between both species. The number of silent differences does not differ significantly among the three functional regions ($\chi^2 = 5.6$, d.f. = 2, $P > 0.05$). In the coding region there are four replacement differences, all conservative amino acid changes. The putative TATA box and two of the three putative polyadenylation signals in *D. melanogaster* are conserved in *D. simulans*. Estimated divergence between *D. melanogaster* and *D. simulans* is 0.0695 and 0.0558 for silent (K_s) and for all sites, respectively (JUKES and CANTOR 1969).

Within *D. simulans* variation: Neither restriction-site nor length variation was detected for the 2.2-kb region analyzed despite the fact that the technique

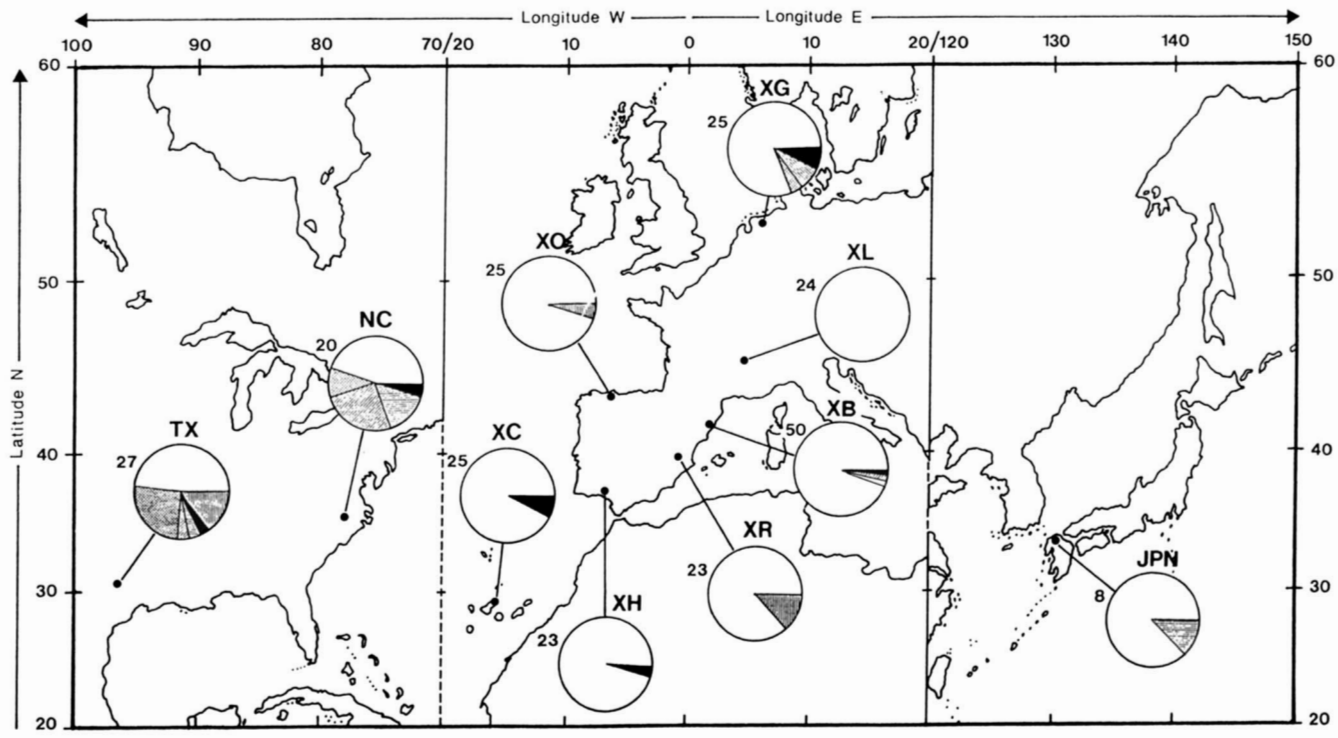

FIGURE 1.—Geographical distribution of *D. melanogaster* populations sampled, and distribution of restriction-site haplotypes. Figures indicate sample size. XG, Groningen (Holland); XL, Lyon (France); XO, Oviedo (Spain); XB, Barcelona (Spain); XR, Requena (Spain); XH, Huelva (Spain); XC, Canary Islands (Spain); NC, North Carolina (United States); TX, Texas (United States); JPN, Fukuoka (Japan).

used allows the detection of nearly all insertions/deletions in a given region, and for the region studied nucleotide variation was assessed for 390 site equivalents (KREITMAN and AGUADÉ 1986b). 64 restriction sites were scored. The haplotype fixed in the three Spanish populations surveyed is identical to that predicted from the DNA sequence of the American allele. Lack of variation within *D. simulans* precludes any further analysis of the data, although it is worth emphasizing the complete absence of variation in this large sample (103 chromosomes).

Within *D. melanogaster* variation: Tables 3 and 4 give a summary of the location and frequency of all restriction-site and length polymorphisms detected in the 245 lines analyzed for all transcription units of the *y-ac-sc* region. Fourteen out of 307 restriction sites scored were polymorphic, their pooled frequency over populations ranging between 0.131 and 0.004. The number of restriction-site polymorphisms segregating in any given population ranged between 0 in Lyon and 8 in both North Carolina and Texas. For each polymorphism frequencies have been compared between populations within a continent, using Fisher's exact test of independence for 2×2 contingency tables and a Monte-Carlo test for $2 \times n$ contingency tables (LEWONTIN and FELSENSTEIN 1965). No significant differences were detected in any of the 8 polymorphisms segregating in North America. Only 1 of the 12 polymorphisms segregating in European populations, #1, showed frequency heterogeneity among samples; when populations were pairwise compared for polymorphism #1, only one of the 6 pairwise comparisons (between XB and XR) indicated frequency heterogeneity between these populations ($P = 0.03$). Some multiply represented restriction-site polymorphisms show higher frequencies in the American than in the European populations (#s 12, 14, 16, 39 and 41); another polymorphism (#40) present both in the American and Japanese samples at frequencies higher than 0.10 has not been detected in any of the European populations sampled. The low number of restriction-site polymorphisms detected in any given transcription unit does not allow to test for heterogeneity in the distribution of polymorphic sites among different functional regions.

Twenty-seven length polymorphisms have been scored, 14 being present only once in all 245 lines analyzed. Most insertions/deletions detected are small, their size ranging between 2 and 40 bp. Length variants (either insertions or deletions) located within the same restriction fragment and differing from the consensus sequence by the same length have been considered the same, although they might be clustering different polymorphisms. All length polymorphisms are located in noncoding regions with the exception of #11 (ins4) located in the second exon of the *yellow* gene (estimated size is 3 bp). One insertion (ins1) and one deletion (del14) correspond to those previously

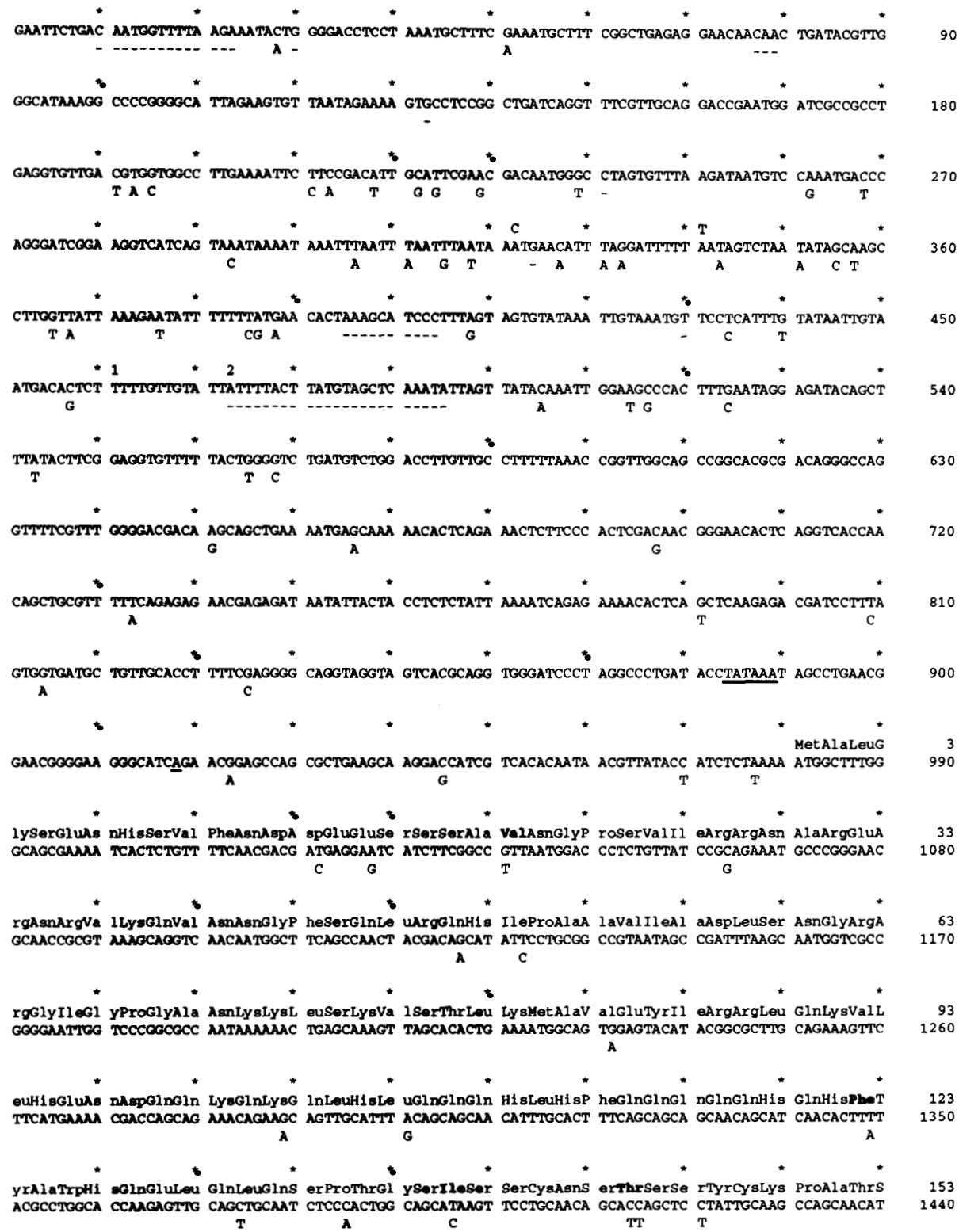

FIGURE 2.—Sequence of the *ac* region from *D. simulans* as compared to that of *D. melanogaster*. Location of the putative TATA box, capping site and polyadenylation sequences (underlined) have been ascertained from comparison with those previously described for *D. melanogaster*. The amino acid sequence of the *ac* gene is presented above the nucleotide sequence, beginning at +981 and ending at +1583. Nucleotide differences between *D. simulans* and *D. melanogaster* as well as deletions in this latter species (–) are shown below the sequence. Insertions one or more base pairs long are shown above the sequence either by the nucleotide inserted at that position or by a number. In the amino acid sequence, those amino acids that have changed between *D. simulans* and *D. melanogaster* are bold typed. The sequence of the *ac* region of *D. simulans* has been submitted to the EMBL/GenBank Data Libraries under the accession number X62400. 1, ATTCCACTAA; 2, TTG; 3, GCG; 4, GC; 5, GACACGCTTCCT; 6, GGTACATTCCTTTAAACGATCCT.

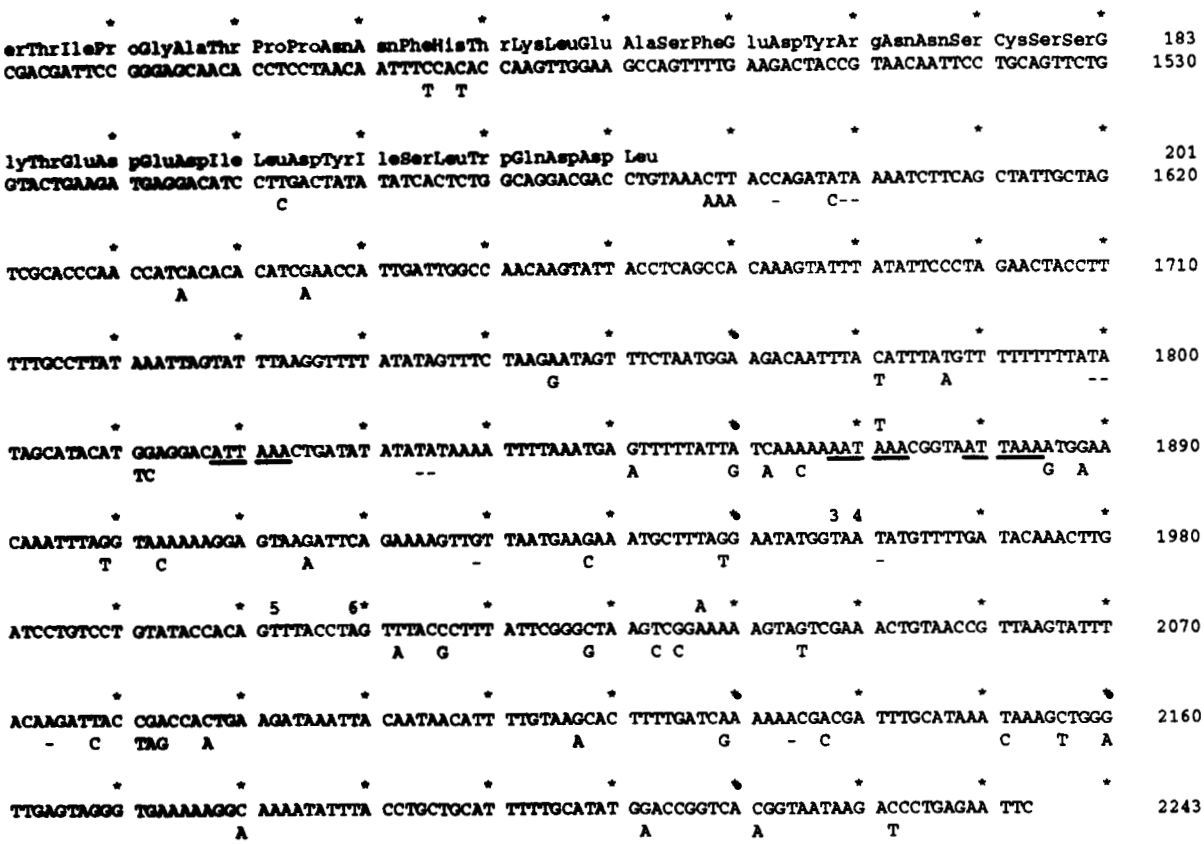

FIGURE 2.—Part 2

detected by six-cutter analysis in those same lines (ins2 and del4, respectively, in AGUADÉ, MIYASHITA and LANGLEY 1989a). In those samples digested with four additional enzymes (see MATERIALS AND METHODS), another length polymorphism (#15b) could be scored and was present both in Barcelona (7 lines) and in North Carolina (2 lines). Some length polymorphisms which reach rather high frequencies (#s 5, 21, 27 and 33) are present in all or most populations sampled; their frequency is however quite heterogeneous, specially among European populations.

Linkage disequilibrium between those restriction-site polymorphisms whose rarest variant shows a frequency higher than 0.10 has been estimated using the correlation coefficient r. This has only been possible for North American populations; lack of heterogeneity in the frequency of restriction-site variants between North Carolina and Texas has allowed pooling of the data. Significant departures from linkage equilibrium were detected in 23 out of 28 pairwise comparisons in the pooled American sample (Table 5). As in previous surveys of this same telomeric region (BEECH and LEIGH-BROWN 1989; AGUADÉ, MIYASHITA and LANGLEY 1989a; EANES, LABATE and AJIOKA 1989; MACPHERSON, WEIR and LEIGH-BROWN 1990) significant linkage disequilibria are found even for sites separated 100 kb.

When both restriction-site and length polymorphisms are considered, 59 different haplotypes have been detected among 245 lines. The number of haplotypes drops to 14 and to 46 when only restriction-site or length polymorphisms are considered, respectively. Given the high number of pairwise linkage disequilibria detected among restriction-site polymorphisms, haplotype analysis should be more informative than site by site analysis. Although there is a major restriction-site haplotype in all populations analyzed, the distribution of most frequent haplotypes differs between European and American samples (Figure 1), but not within continents: in European samples there is only one major haplotype (with frequency 0.921), while in the American samples besides this major haplotype (with frequency 0.468), there are two other rather frequent haplotypes (with frequencies 0.191 and 0.128, respectively); these other haplotypes are also present in Europe but at much lower frequencies (0.011 and 0.005, respectively). Restriction-site haplotype diversity (NEI and TAJIMA 1981) is accordingly much lower in European ($h = 0.1516$) than in American samples ($h = 0.7285$); estimated haplotype diversity for the Japanese sample is also rather low ($h = 0.2503$).

Table 6 gives estimates of nucleotide diversity for each individual population as well as for populations pooled according to continent [H, ENGELS (1981); π, NEI and TAJIMA (1981); θ, HUDSON (1982)]. The

TABLE 1

Description of the probed regions in *D. melanogaster*

	T6 (*y*)	T5 (*ac*)	T4 (*sc*α)	T3 (*l'sc*)	T2 (*sc*β)	T1a (*scγ* or *as*)	T1	Total
Plasmid	yscS/R	2.1RR101	9.0RRf53	3.2RR22	6.3BB14; 1.8RR14	4.9BB53	3.9RRf53	
Coordinates in map[a]	74.7, 66.0	59.6, 57.5	37.5, 28.8	19.9, 17.8	6.0, −0.4; 0, −1.8	−22.4, −27.0	−29.5, −33.5	
Region used as probe	*Bgl*II-*Bgl*II	*Eco*RI-*Eco*RI	*Xba*I-*Pvu*II	*Eco*RI-*Eco*RI	*Hin*cII-*Hin*cII-*Bam*HI *Eco*RI-*Eco*RI	*Xho*I-*Bam*HI	*Eco*RI-*Eco*RI	
Site in published sequence	−1868 to +4756[b,c]	1 to 2232[d]	1 to 2133[d]	1 to 2586[c,f]	1 to 2232[f]	1 to 1600[e]		
Not sequenced region			880 bp 3′	620 bp 3′	820 bp 3′	450 bp 5′, 1200 bp 3′	3.9 kb	
Probe size (in kb)	6.6	2.2	2.95	3.2	3.05	3.2	3.9	25.1
Exons	1623	603	1035	777	1218	1188	1971[a]	8415
Introns	2719							2719
Noncoding	2283	1629	1915	2423	1832	2062	1929	14073
No. sites scored	80	36 (62)[g]	38	28	38	36	51	307 (333)[g]

[a] CAMPUZANO *et al.* (1985).
[b] GEYER, SPANA and CORCÉS (1986).
[c] GEYER and CORCÉS (1987).
[d] VILLARES and CABRERA (1987).
[e] ALONSO and CABRERA (1988).
[f] GONZÁLEZ (1989).
[g] Figures in parentheses indicate the number of sites scored when eight four-cutter re

TABLE 2

Distribution of nucleotide divergence for the *ac* region between *D. simulans* and *D. melanogaster*

	5′	Coding[a]			3′
		Total	*ns*	*s*	
No. of changes	55	19	4	15	43
No. of nucleotides compared	925	603	472	131	646

[a] *ns*, nonsilent or replacement; *s*, silent or synonymous.

major difference between European and North American samples lies in the different haplotype distribution, and consequently of polymorphic variants in strong linkage disequilibrium in these populations. Estimates of heterozygosity are accordingly lower in European than in American samples, especially when using those estimators like NEI and TAJIMA's π (1981) and ENGELS' H (1981) that take into account not only the number but also the frequency of polymorphic sites.

In order to test for departures from the neutral theory TAJIMA's test (1989) has been used. The rationale of TAJIMA's test is that under the neutral mutation model no difference would be expected between the estimates of heterozygosity based on the number of segregating sites (WATTERSON 1975) and that based on the average number of nucleotide differences (TAJIMA 1983). Any departure of the frequency spectrum of variants from the neutral prediction will affect this latter estimate, as it takes into account frequencies, but not the former. Table 7 gives the estimates of D (equation 38 in TAJIMA 1989) for those *D. melanogaster* populations where the number of polymorphisms is greater than one. D values for all European populations are negative. In all four cases the probabilities associated with each D value are lower than 0.10, the estimated D value being significantly different from zero in two out of these four cases.

DISCUSSION

The relative importance of forces shaping nucleotide variation may vary across the genome. Differences in levels and patterns of intraspecific nucleotide variation in different regions might be due to differential positive or negative selection, different mutation rates and/or differential levels of recombination. One way to rule out differential purifying selection and different mutation rates as the main forces causing different levels of polymorphism in different regions is to compare estimates of interspecies divergence. The neutral mutation rate in a region decreases as the fraction of deleterious mutations increases. Divergence due to the substitution of neutral mutations in a region with a lower neutral mutation rate should also be reduced because the rate of fixation of

TABLE 3

Location and frequency of polymorphisms in *D. melanogaster*

#	Type	Location		Absolute frequency			
		Absolute	Functional	Europe (190)	America (47)	Japan (8)	Total (245)
T6							
1	*Taq*I (gain)	−1504	5′	4			4
2	*Taq*I (loss)	−1458, −1455	5′	2			2
3	del1 (120 bp)	−1018T, −755T	5′	5			5
4	ins1 (600 bp)	−755T, −495D	5′		1		1
5	del2 (20 bp)	1293–1335	Intron	96	28	5	129
6	ins2 (2 bp)	2002H–2044T	Intron	1			1
7	*Sau*3AI (gain)	2491	Intron	2	6		8
8	del3 (4–6 bp)	2648H–3096H	Introl	1			1
9	ins3 (20 bp)	2648H–3096H	Intron	1			1
10	*Hae*III (loss)	3097–3100	Intron	2			2
11	ins4 (3 bp)	4232A–4354H	Exon 2	1			1
12	*Hae*III (loss)	4442–4445 (4444 s)	Exon 2	5	11		16
13	ins5 (300 bp)	T4557–A4832	3′	1			1
T5							
14	*Taq*I (gain)	36	5′	5	11		16
15	del4 (20 bp)	49–85	5′	1 (2)			1 (2)
15b	del4b (2 bp)	245–471	5′	(7)	(2)		(9)
16	*Alu*I (gain)	2147	3′	3	11		14
T4							
17	*Sau*3AI (gain)	412	5′	1	1		2
18	del5 (4 bp)	430A–575S	5′	1	4	1	6
19	del6 (2 bp)	575S–620S	5′	1			1
20	*Taq*I (gain)	818/964	Coding		1		1
21	ins6 (12 bp)	2015S–2446T	3′	65	7		72
22	ins7 (6 bp)	2015S–2446T	3′	3	15		18
23	ins8 (10 bp)	2500–2650	3′	1			1
24	del7 (8–10 bp)	2500–2900	3′	2			2
T3							
25	*Hae*III (gain)	823	5′	1			1
26	del8 (4 bp)	956S–1088A	5′		2		2
27	del10 (10 bp)	1781A–2129S	3′	56	3		59
28	ins9 (30–40 bp)	2506H–2740S	3′	3	10		13
29	ins10 (10–20 bp)	2506H–2740S	3′	4	2		6
30	del9 (2–4 bp)	2746H–2940A	3′	1			1
T2							
31	*Taq*I (gain)	Tb340 to 145		1			1
T1a							
32	del11 (2–4 bp)	1345T–1511H	3′	2			2
33	del12 (10 bp)	1503–1523	3′	32	9	2	43
34	del13 (10 bp)	Ab420		1			1
35	del14 (150 bp)	Tb542	3′		1		1
T1							
36	del15 (220 bp)	Ab355		1			1
37	del16 (5 bp)	Ab355				1	1
38	del17 (10–20 bp)	Tb738				1	1
39	*Hae*III (gain)	b900	5′	6	23	1	30
40	*Hae*III (loss)	b735 + b160			8	1	9
41	*Hae*III (gain)	b280	3′	7	24	1	32

Polymorphisms are grouped according to the transcription unit where they are located (see Table 1). Absolute location is given according to published sequences as noted in Table 1. Nucleotide substitutions are indicated by a single nucleotide (in case of gain of site) or by an interval of four nucleotides (in case of loss of site). Insertions/deletions have been mapped to the minimal restriction fragment where they could be detected; limits are given either by nucleotide position of a given restriction site (A, *Alu*I; D, *Hin*dIII; H, *Hae*III; S, *Sau*3A1; T, *Taq*I) when located in a sequenced region, or by the size of a band for a given restriction enzyme. Polymorphism #15b has been scored only in a subsample of populations (see MATERIALS AND METHODS) and its frequency is given in parentheses.

neutral alleles equals the neutral mutation rate. When silent site divergence (K_s) for the *ac* region between *D. simulans* and *D. melanogaster* (0.0695) is compared with those values estimated for the *hsp82* (0.057, BLACKMAN and MESELSON 1986), *Mtn* (0.0753, LANGE, LANGLEY and STEPHAN 1990), *Adh* 5′-flank-

TABLE 4

Haplotype Distribution in *D. melanogaster* populations

Hap.	Polymorphisms 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	Populations XG	XL	XO	XB	XR	XH	XC	NC	TX	JPN	TOT
1	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	5	4	8	7	1	1	8	1		2	37
2	−	+	+	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	2			1			1				4
3	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	7	5	3	9	1	3		3			31
4	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	1								3		4
5	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	−	1						1				2
6	−	−	−	−	−	+	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1										1
7	−	+	−	−	−	−	−	−	−	+	−	−	−	+	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	−	+	1										1
8	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1		1	4	4	7	2	1			20
9	−	+	−	−	+	−	−	−	+	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	1										1
10	−	−	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1										1
11	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	1										1
12	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1	1	1	3	2	9	4	3	5	2	31
13	−	+	+	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	1										1
14	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	1										1
15	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−		2		2				1	7	2	14
16	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	−		4		4	4		4				16
17	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−		5	2	2	3	1	1				14
18	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−		1										1
19	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−		1		4	2						7
20	−	+	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−		1										1
21	+	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−			1		1						2
22	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−			2								2
23	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−			1								2
24	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	−			1								1
25	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−				1							1
26	−	+	−	−	−	−	+	−	−	+	−	−	−	+	−	+	−	−	−	−	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	−	+				1				2	1		4
27	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	−				3	2						5
28	−	+	−	−	+	−	−	+	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	−				1							1
29	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	+	−	+				1							1
30	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−				1							1
31	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	+	−	−	+	−	−	−	−	−	−	−	−				2							2
32	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−				1							1
33	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−				1							1
34	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−				1							1
35	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−				1							1
36	+	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−				2					1		2
37	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−					2						2
38	−	+	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	+	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	+					1						1
39	−	+	−	−	+	−	−	−	−	+	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−						1					1
40	−	+	−	−	−	−	+	−	−	+	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−	−	+	−	+							1				1
41	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−							1				1
42	−	+	−	−	−	−	−	−	−	+	−	+	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−							1				1
43	−	+	−	−	−	−	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	+	−	+							1				1
44	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	+	+							2				2
45	−	+	−	−	−	−	+	−	−	+	−	−	−	+	−	+	−	−	−	−	−	+	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	+	−	+							1	2			1
46	−	+	−	+	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+							1	1			1
47	−	+	−	−	−	−	+	−	−	+	−	−	−	+	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+							1	2			2
48	−	+	−	−	−	−	−	−	−	+	−	−	−	+	−	+	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+							1				1
49	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+							1				1
50	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+							1				1
51	−	+	−	−	−	−	−	−	−	+	−	−	−	+	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	+	+							1	1			1
52	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−							1				1
53	−	+	−	−	−	−	−	−	−	+	−	−	−	+	−	+	−	+	−	−	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	+	+							1	2			2
54	−	+	−	−	−	−	−	−	−	+	−	+	−	+	−	+	−	+	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+							1	3			3
55	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	+	−	+							1	1			1
56	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+							1	1			1
57	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+							1	1			1
58	−	+	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−										1	1
59	−	+	−	−	+	−	−	−	−	+	−	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+										1	1
																																										25	24	20	50	23	23	25	20	27	8	245

Polymorphisms are numbered according to Table 3. Populations are named according to Figure 1. As compared to published sequences all lines were monomorphic for absence of *Hae*III and *Taq*I sites at positions −1877 and 560, and for presence of *Sau*3AI site at position 1795 in the *yellow* region; they were all also monomorphic for absence of *Hae*III site at position 321 in T3 as in GONZÁLEZ (1989) but in contradiction with VILLARES and CABRERA (1987).

TABLE 5

Linkage disequilibrium between polymorphic sites expressed as the correlation coefficient *r*: American populations

7	12	14	16	39	40	41	
0.464***	0.671***	0.671***	0.671***	0.321*	0.204	−0.373*	5
	0.692***	0.692***	0.692***	0.391**	−0.173	−0.374*	7
		1.000***	1.000***	0.464***	0.284	−0.541***	12
			1.000***	0.464***	0.284	−0.541***	14
				0.464***	0.284	−0.541***	16
					0.463***	−0.958***	39
						−0.443***	40

Rows and columns indicate polymorphic sites numbered according to Table 3. * $0.01 < P < 0.05$, ** $0.001 < P < 0.01$, *** $P < 0.001$.

TABLE 6

Estimates of nucleotide diversity for *D. melanogaster* populations

Populations[a]	θ[b]	π[b]	H[b]
XG (25)	0.00076	0.00032	0.00032
XL (24)	0	0	0
XO (20)	0.00011	0.00004	0.00004
XB (50)	0.00064	0.00015	0.00015
XR (23)	0.00011	0.00009	0.00009
XH (23)	0.00056	0.00018	0.00018
XC (25)	0.00076	0.00035	0.00035
Pooled Europe (190)	0.00100	0.00017	0.00017
NC (20)	0.00093	0.00131	0.00131
TX (27)	0.00086	0.00102	0.00102
Pooled America (47)	0.00084	0.00115	0.00114
JPN (8)	0.00047	0.00031	0.00031
World pool (245)	0.00096	0.00043	0.00044

[a] Populations are named according to Figure 1. Figures in parentheses indicate sample size.
[b] See text for references.

TABLE 7

Estimates of *D* and its significance according to TAJIMA (1989)

Populations	*D* values[a]
XG (25)	−1.7911 (−1.807, −1.583)
XB (50)	−2.0417* (−1.800, −1.570)
XH (23)	−1.9921* (−1.806, −1.584)
XC (25)	−1.6688 (−1.807, −1.583)
NC (20)	1.3181 (2.001, 1.710)
TX (27)	0.5451 (2.001, 1.712)
JPN (8)	−1.4475 (−1.663, −1.521)

[a] Figures in parentheses are the *D* values with associated probability 0.05 and 0.10, respectively (TAJIMA 1989).

ing (0.0678), *Adh* coding (0.0368) and *Adh* 3′-flanking (0.0398) regions (COHN and MOORE 1988), no such reduction is observed. The silent sites of the *ac* region do not seem therefore to be under significant stronger constraint (or suffer a lower mutation rate) than similar sites in these other regions.

The present data on variation at the *ac* region in *D. simulans*, where silent site variation can be accurately estimated, have no counterpart at other loci with which to compare levels of polymorphism. The estimate of nucleotide variation in the *ac* region of *D. simulans* is slightly lower than that observed in the same region in *D. melanogaster* (0 *vs.* 2 silent site restriction-site polymorphisms detected in *D. simulans* and *D. melanogaster*, respectively, out of 298 silent sites analyzed). A previous six-cutter study of the *rosy* region in North American populations of both *D. melanogaster* and *D. simulans* found a higher heterozygosity per nucleotide in this latter species (0.019 in *D. simulans vs.* 0.003 in *D. melanogaster*, AQUADRO, LADO and NOON 1988). If this single estimate of nucleotide variation in *D. simulans* could be considered representative of that species [as seems to be confirmed by sequencing data at the *Adh* coding region (MACDONALD and KREITMAN 1991)], the observed level of DNA sequence polymorphism at the *ac* region seems to be reduced even more relative to other regions than it is in *D. melanogaster* (LANGLEY 1990). In fact, under the assumption of neutrality at the *ac* region of *D. simulans*, the maximum value of theta (θ_U) compatible at the 0.05 level with the observation of no variation in a sample of 103 chromosomes would be 0.63 for all 390 site equivalents (according to equation 12 in HUDSON 1990) and 0.002 (=0.63/390) per nucleotide; this latter estimated maximum value of theta for the *ac* region is an order of magnitude lower than the estimated theta for the *rosy* region of *D. simulans*. Given that interspecific divergence at the *ac* region is not reduced, this lower level of intraspecific variation in *D. simulans ac* region is consistent with the hitchhiking effect of a selectively favorable mutation in this region of reduced crossing over per physical length.

The present data for *D. melanogaster* confirm our previous results of low levels of variation in the *y-ac-sc* region (AGUADÉ, MIYASHITA and LANGLEY 1989a). Present estimates of heterozygosity, both for single and pooled population samples (Table 6), are in fact among the lowest estimates for any nuclear genomic region in Drosophila. Unlike for *D. simulans*, for *D. melanogaster* there is at least a region (*Adh* 5′-flanking) for which intraspecific silent site variation has been accurately estimated (KREITMAN and AGUADÉ 1986b;

TABLE 8

Estimates of the number of silent site equivalents in the *y-ac-sc* region of *D. melanogaster*

No. of sites	T6 (*y*)	T5 (*ac*)	T4 (*sc*α)	T3 (*l'sc*)	T2 (*sc*β)	T1a (*sc*γ or *ase*)	T1	Total
Noncoding								
Sites	4999	1630	2077 (880)	2433 (620)	1833 (820)	2062 (1650)	1929 (1929)	16439 (7961)
Site equivalents	408	166	182 (76)	207 (53)	150 (71)	182 (142)	166 (166)	1461 (408)
Coding								
Sites	1623	603	1035	771	1218	1188	1971 (1971)	8415 (1971)
Site equivalents	177	45	106	84	139	136	210 (210)	897 (210)
Silent site equivalents	42	10	25	18	32	31	48 (48)	206 (48)
Total								
Silent site equivalents	450	176	207	225	182	213	214	1667

Figures in parentheses indicate either the number of sites not sequenced in a given fragment or the number of site equivalents estimated for those nonsequenced regions (see text).

TABLE 9

HKA tests for silent site differences in the *y-ac-sc* and the *Adh* 5′-flanking regions

	Populations surveyed for the *y-ac-sc* region					
	European		American		Total	
	y-ac-sc	*Adh* 5′	*y-ac-sc*	*Adh* 5′	*y-ac-sc*	*Adh* 5′
n	190	81 (11)	47	81 (11)	245	81 (11)
S_i	12	9 (30)	9	9 (30)	14	9 (30)
m_i	1667	425 (1243)	1789	425 (1243)	1667	425 (1243)
D_i	113	210	113	210	113	210
n_i	2107	4052	2107	4052	2107	4052
X^2	3.68 (9.12)		3.70 (7.68)		3.01 (8.50)	

Within *D. melanogaster* sequencing data and X^2 values for the 5′ *Adh* region are given in parentheses (see text). n, sample size of population samples. S_i, observed number of segregating sites within populations. m_i, number of silent site equivalents (or of silent sites for sequencing data). D_i, observed number of silent site differences between species. n_i, number of silent sites compared between species. X^2, calculated according to HUDSON, KREITMAN and AGUADÉ (1987) with the modifications of BEGUN and AQUADRO (1991).

KREITMAN and HUDSON 1991), where silent site divergence has been estimated both between *D. melanogaster* and *D. simulans* (COHN and MOORE 1988) and between *D. melanogaster* and *Drosophila sechellia* (HUDSON, KREITMAN and AGUADÉ 1987), and where there is no evidence of selection or lower than average recombination. Present estimates of within *D. melanogaster* polymorphism and of interspecies divergence in the *y-ac-sc* region have been compared to those in the *Adh* 5′-flanking region in order to test for departures from the neutral theory using the conservative HUDSON, KREITMAN and AGUADÉ (1987) test (HKA test). In order to estimate the expected number of segregating sites under the infinite sites model with no selection or recombination (WATTERSON 1975), the number of silent sites studied needs to be estimated. When four-cutter enzymes are used to analyze variation at sequenced regions, the number of nucleotides that are being surveyed with a particular set of restriction enzymes (site equivalents) can be accurately estimated as can the percentage of silent site equivalents in coding regions. As shown in Table 1 sequence is available for 69 percent of the probed region. Given that no direct estimates of site equivalents can be obtained for nonsequenced regions, estimates for those regions have been obtained by extrapolating the average percentage over sequenced noncoding (8.61% of all sites) and coding (10.66% of all sites, being 23% silent) regions, respectively. Table 8 shows those estimates for the set of four restriction enzymes used. The number of silent site equivalents for the *achaete* region (T5) increases to 298 when all eight restriction enzymes are considered (see MATERIALS AND METHODS), and consequently the total number of silent site equivalents surveyed in those populations is 1789.

As shown in Table 9, for the *Adh* 5′-flanking region two estimates of polymorphism have been considered: that based on a four-cutter analysis of a large sample of two American populations (KREITMAN and AGUADÉ 1986b), and that based on sequencing data of a smaller world-wide sample (KREITMAN and HUDSON 1991). Silent site divergence for the larger *Adh* 5′-flanking region compared between *D. melanogaster* and *D. sechellia* (HUDSON, KREITMAN and AGUADÉ 1987) has been used to perform the HKA tests shown in Table

9. *D. sechellia* can be used here instead of *D. simulans* because *D. simulans* and *D. sechellia* are sister taxa and thus have the same time of divergence from *D. melanogaster*. In the present case where variation at an *X*-linked region is compared to that in an autosomal region, corrected expressions both for expected number of segregating sites and its variance have been used for the X-linked region taking into account its lower effective population size (see BEGUN and AQUADRO 1991). As already considered in the original application of the HKA test (HUDSON, KREITMAN and AGUADÉ 1987), the test also requires a slight modification since the estimates of polymorphism and of divergence are based on slightly different sets of sites. Table 9 shows the data used to perform the tests and the X^2 values obtained when comparing intra- and interspecific variation in the *y-ac-sc* and in the *Adh* 5′-flanking regions separately for samples from different continents and for the pooled data. The Japanese sample has been considered only in the total sample and not separately due to its small sample size and to the possible effect of this small sample size on the behavior of the X^2 statistic. In all tests performed using sequencing data for the *Adh* 5′-flanking region, the X^2 values show a significant departure from neutrality. When four-cutter data are used for the *Adh* 5′-flanking region, the X^2 values show an associated probability only slightly larger than 0.05 both for European and for American samples, and between 0.05 and 0.10 for the pooled data. In all cases the number of observed segregating sites for the *y-ac-sc* region is smaller than those that would be expected from interspecies divergence estimates. Under the hypothesis of hitchhiking, one would not only expect reduced levels of variation in the *y-ac-sc* region but also an excess of low frequency polymorphisms in that region. When restriction-site polymorphisms are considered, TAJIMA's statistic *D* is negative for all European populations (Table 7), an indication of a skewed frequency spectrum of polymorphisms consistent with a hitchhiking effect. The reduction in standing levels of variation at the *y-ac-sc* region in *D. melanogaster* can be therefore most easily explained by the hitchhiking effect of positive selection in a few rare sites.

Although for the region surveyed in *D. melanogaster* there is a major most common haplotype in all populations, there is some indication of population subdivision for North American *vs.* European samples. In North American populations there are indeed several sites with high frequency variants that are in strong linkage disequilibirum; American samples do show positive values of TAJIMA's statistic *D*, that might be reflecting or not some additional interesting feature of these populations. This differentiation may be compatible with the hitchhiking hypothesis if some populations are sufficiently isolated. The quantitative interpretation of hitchhiking and migration will require further theoretical analysis.

We would like to thank S. CAMPUZANO and J. MODOLELL for the different clones from the *ac-sc* region and for sharing unpublished results; V. CORCÉS for clone yscS/R; J. COYNE for the attached-*X* chromosome strain of *D. simulans*; R. ALLEMAND, W. VAN DELDEN, J. IZQUIERDO, A. GONZÁLEZ, D. OCHANDO and J. ROZAS for flies; and S. CIRERA for assistance. We also thank C. F. AQUADRO, J. HEY, R. R. HUDSON, M. KREITMAN, C. H. LANGLEY, M. TURELLI and J. ROZAS for critical comments, and B. JUDD and C. H. LANGLEY for facilities and support at the National Institute of Environmental Health Sciences during visits by M.A. This work was partly supported by predoctoral research grants (AR-87 and AR-88) from Comissió Interdepartamental de Recerca i Innovació Tecnològica to J.M.C. and to J.M.M-C., respectively, and by NATO grant CRG-900651 to M.A.

LITERATURE CITED

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

AGUADÉ, M., N. MIYASHITA and C. H. LANGLEY, 1989b Restriction-map variation at the *zeste-tko* region in natural populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6:** 123–130.

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

AQUADRO, C. F., K. M. LADO and W. A. NOON, 1988 The *rosy* region of *Drosophila melanogaster* and *D. simulans*. I. Contrasting levels of naturally occurring restriction map variation and divergence. Genetics **119:** 875–888.

BEECH, R. N., and A. J. LEIGH-BROWN, 1989 Insertion-deletion variation at the *yellow, achaete-scute* region in two natural populations of *Drosophila melanogaster*. Genet. Res. **53:** 7–15.

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.

BLACKMAN, R. K., and M. MESELSON, 1986 Interspecific nucleotide comparison used to identify regulatory and structural features of the *Drosophila hsp82* gene. J. Mol. Biol. **188:** 499–515.

CAMPUZANO, S., L. CARRAMOLINO, C. V. CABRERA, M. RUIZ-GOMEZ, R. VILLARES, A. BORONAT and J. MODOLELL, 1985 Molecular genetics of the *achaete-scute* gene complex of *D. melanogaster*. Cell **40:** 327–388.

COHN, V. H., and G. P. MOORE, 1988 Organization and evolution of the alcohol dehydrogenase gene in *Drosophila*. Mol. Biol. Evol. **5:** 154–166.

EANES, W. F., J. LABATE and J. W. AJIOKA, 1989 Restriction-map variation within the *yellow-achaete-scute* region in five populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6:** 492–502.

ENGELS, W. R., 1981 Estimating genetic divergence and genetic variability with restriction endonucleases. Proc. Natl. Acad. Sci. USA **78:** 6329–6333.

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

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

GONZÁLEZ, F., 1989 Estructura molecular de los genes del complejo *achaete-scute* de *Drosophila melanogaster*. Ph.D. thesis, Universidad Autónoma de Madrid.

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

HUDSON, R. R., 1982 Estimating genetic variability with restriction endonucleases. Genetics **100:** 711–719.

HUDSON, R. R., 1990 Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7:** 1–44.

HUDSON, R. R., KREITMAN, M. and M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.

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

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

KREITMAN, M., and M. AGUADÉ, 1986a Genetic uniformity in two natural populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA **86:** 3562–3666.

KREITMAN, M., and M. AGUADÉ, 1986b Excess polymorphism at the *Adh* locus in *Drosophila melanogaster*. Genetics **114:** 93–110.

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.

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

LANGLEY, C. H., 1990 The molecular population genetics of Drosophila, pp. 75–91 in *Population Biology of Genes and Molecules*, edited by N. TAKAHATA and J. F. CROW. Baifukan, Tokyo.

LEWONTIN, R. C., and J. FELSENSTEIN, 1965 The robustness of the homogeneity test in $2 \times N$ tables. Biometrics **21:** 19–33.

MACPHERSON, J. N., B. S. WEIR and A. J. LEIGH-BROWN, 1990 Extensive linkage disequilibrium in the *achaete-scute* complex of *Drosophila melanogaster*. Genetics **126:** 121–129.

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

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

MIYASHITA, N., 1990 Molecular and phenotypic variation at the *Zw* locus region in *Drosophila melanogaster*. Genetics **125:** 407–419.

MIYASHITA, N., and C. H. LANGLEY, 1988 Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. Genetics **120:** 199–212.

MIYASHITA, N., C. C. LAURIE-AHLBERG, A. N. WILTON, and T. H. EMIGH, 1986 Quantitative analysis of *X* chromosome effects on the activities of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases of *Drosophila melanogaster*. Genetics **113:** 321–335.

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

SAMBROOK, J., E. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning, A Laboratory Manual*, Ed.2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

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

SCHAEFFER, S. W., C. F. AQUADRO and C. H. LANGLEY, 1988 Restriction-map variation at the *Notch* region of *Drosophila melanogaster*. Mol. Biol. Evol. **5:** 30–40.

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.

TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

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

WATTERSON, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

Communicating editor: M. TURELLI