The *rosy* Region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting Levels of Naturally Occurring DNA Restriction Map Variation and Divergence

Charles F. Aquadro, Karen M. Lado and William A. Noon

Section of Genetics and Development, Cornell University, Ithaca, New York 14853

Manuscript received August 7, 1987
Revised copy accepted April 20, 1988

ABSTRACT

A 40-kb region around the *rosy* and *snake* loci was analyzed for restriction map variation among 60 lines of *Drosophila melanogaster* and 30 lines of *Drosophila simulans* collected together at a single locality in Raleigh, North Carolina. DNA sequence variation in *D. simulans* was estimated to be 6.3 times greater than in *D. melanogaster* (heterozygosities per nucleotide of 1.9% *vs.* 0.3%). This result stands in marked contrast to results of studies of phenotypic variation including proteins (allozymes), morphology and chromosome arrangements which are generally less variable and less geographically differentiated in *D. simulans*. Intraspecific polymorphism is not distributed uniformly over the 40-kb region. The level of heterozygosity per nucleotide varies more than 12-fold across the region in *D. simulans*, being highest over the *hsc2* gene. Similar, though less extreme, variation in heterozygosity is also observed in *D. melanogaster*. Average interspecific divergence (corrected for intraspecific polymorphism) averaged 3.8%. The pattern of interspecific divergence over the 40-kb region shows some disparities with the spatial distribution of intraspecific variation, but is generally consistent with selective neutrality predictions: the most polymorphic regions *within* species are generally the most divergent *between* species. Sequence-length polymorphism is observed for *D. melanogaster* to be at levels comparable to other gene regions in this species. In contrast, no sequence length variation was observed among *D. simulans* chromosomes (limit of resolution approximately 100 bp). These data indicate that transposable elements play at best a minor role in the generation of naturally occurring genetic variation in *D. simulans* compared to *D. melanogaster*. We hypothesize that differences in species effective population size are the major determinant of the contrasting levels and patterns of DNA sequence and insertion/deletion variation that we report here and the patterns of allozyme and morphological variation and differentiation reported by other workers for these two species.

DEVELOPMENTS in DNA technology have allowed population geneticists to initiate comparisons of the level of protein variation with the underlying DNA sequence variation. Significant levels of polymorphism exist both in the coding and the noncoding region of the genome (*e.g.*, KREITMAN 1983; AQUADRO *et al.* 1986; KREITMAN and AGUADÉ 1986; SCHAEFFER, AQUADRO and ANDERSON 1987; LANGLEY and AQUADRO 1987), with recent analyses of within and between species variation suggesting a departure from selective neutrality for both sequence length and base pair substitution variants (*e.g.*, AQUADRO *et al.* 1986; GOLDING, AQUADRO and LANGLEY 1986; KREITMAN and AGUADÉ 1986; HUDSON, KREITMAN and AGUADÉ 1987). Interest in the distribution of this variation along the chromosome, and in the existence of nonrandom associations between variable sites has been rejuvenated. In addition, many of the "point" mutations studied in laboratory stocks of *Drosophila melanogaster* have been found to be caused by insertion and/or deletion events mediated by transposable elements, rather than by simple nucleotide substitution (reviewed by RUBIN 1983). This finding has led to the speculation that transposable elements have played a major role in the evolution of species (*e.g.*, BINGHAM, KIDWELL and RUBIN 1982; ROSE and DOOLITTLE 1983; GINZBURG, BINGHAM and YOO 1984).

In *Drosophila*, most of the studies of DNA sequence variation have focused on the single species *D. melanogaster*. It is important to determine whether the patterns observed in *D. melanogaster* are representative of *Drosophila* in general. In addition, KREITMAN and AGUADÉ (1986) and HUDSON, KREITMAN and AGUADÉ (1987) have demonstrated that comparisons of within-species polymorphism to between-species divergence can provide a powerful test of the influence of drift versus natural selection in determining patterns of DNA sequence variation. *Drosophila simulans* provides a particularly interesting comparison because it is closely related to *D. melanogaster* and shows a similar, worldwide geographic distribution. *D. simulans* and *D. melanogaster* are often sympatric and appear to occupy almost the same ecological niche (DAVID and TSACAS 1981). Previous studies of protein

polymorphism have found that *D. simulans* possesses a lower average level of protein polymorphism within populations and geographic differentiation between populations than *D. melanogaster* (O'BRIEN and MACINTYRE 1969; BERGER 1970; GONZALEZ *et al.* 1982; OHNISHI *et al.* 1982; HYYTIA *et al.* 1985; SINGH, CHOUDHARY and DAVID 1987; CHOUDHARY and SINGH 1987b; WATADA, TOBARI and OHBA 1986). Moreover, ASHBURNER and LEMEUNIER (1976), in an analysis of chromosome inversion polymorphisms, found 57 inversions in 67 collections of *D. melanogaster*, of which seven were cosmopolitan and often present at high frequency. In contrast, no inversions were found in 27 collections of *D. simulans*. A study by DOWSETT and YOUNG (1982) suggested that *D. simulans* may possess significantly lower levels of dispersed middle repetitive DNA than *D. melanogaster*. This latter finding is of interest in light of ASHBURNER and LEMEUNIER's (1976) results since dispersed middle repetitive DNA includes transposable elements, which have been implicated as causes of inversions and other chromosome rearrangements (*e.g.*, ENGELS 1983; GOLDBERG *et al.* 1983; DAVIS, SHEN and JUDD 1987).

In this study, we have examined naturally occurring restriction-site polymorphism around the *rosy* locus on the third chromosome in *D. melanogaster* and *D. simulans*. This region has been well characterized at the level of fine structure genetics, genomic DNA maps, location of transcripts and chromatin structure (*e.g.*, BENDER, SPIERER and HOGNESS 1983; BOSSY, HALL and SPIERER 1984; COTÉ *et al.* 1986; GAUSZ *et al.* 1986 and references therein). The *rosy* locus is located on the right arm of chromosome *3* between polytene chromosome bands 87D8-12 of *D. melanogaster* (CHOVNICK, GELBART and MCCARRON 1977). This region is present in an inverted orientation on the *D. simulans* third chromosome due to a large inversion fixed between the two species, having breakpoints at 84-85A and 93F (*D. melanogaster* coordinates) (OHNISHI and VOELKER 1979). *Rosy* encodes the enzyme xanthine dehydrogenase (XDH), which is a homodimer of two 150,000-dalton subunits (GELBART *et al.* 1974; EDWARDS, CANDIDO and CHOVNICK 1977). XDH catalyzes the reaction of hypoxanthine to xanthine, which is then converted to uric acid (CHOVNICK, GELBART and MCCARRON 1977). *Rosy* mutants lacking in XDH activity cannot synthesize the red drosopterin eye pigments and consequently have brownish eye color. XDH allozyme polymorphism can be readily examined using protein electrophoresis and has been the object of extensive study in natural populations of *Drosophila* (*e.g.*, SINGH, HICKEY and DAVID 1982; BUCHANAN and JOHNSON 1983; KEITH 1983). The *rosy* locus (LEE *et al.* 1987; KEITH *et al.* 1987) is flanked on the 5′ end by the lethal gene *l(3)s12*, and on the 3′ end by *snake*, a maternal effect gene involved in the establishment of the dorsal-ventral axis in development (DELOTTO and SPIERER 1986), and by *hsc2*, a heat shock cognate gene (CRAIG, INGOLIA and MANSEAU 1983).

We report here an examination of restriction map variation and divergence in a 40-kb region around the *rosy* locus in 60 lines of *D. melanogaster* and 30 lines of *D. simulans* sampled from a single locality in Raleigh, North Carolina. Surprisingly, *D. simulans* shows a sixfold higher level of nucleotide heterozygosity, and no insertion/deletion variation compared to *D. melanogaster*. This difference in restriction site polymorphism is in marked contrast to results from surveys of protein (allozyme) polymorphism which have shown *D. simulans* to be no more variable, and often substantially less variable, then *D. melanogaster*. These results lead to the hypothesis of larger effective population size, and a resulting stronger purifying selection, in *D. simulans* relative to *D. melanogaster*. In addition, the pattern of intraspecific polymorphism compared with interspecific differentiation at the restriction site level indicates that the underlying substitution rate is significantly higher in the region including *rosy*, *snake* and *hsc2* than in the adjacent 20-kb region, particularly in *D. simulans*. Levels and patterns of linkage disequilibrium among restriction map variants and haplotype relatedness will be discussed in a subsequent paper.

MATERIALS AND METHODS

Samples: Flies were collected using sweep nets over fermenting bananas in buckets during a 4-day period in November 1984 from the Raleigh (North Carolina) Farmer's Market. Isofemale lines were established for both species. Sixty lines of *D. melanogaster* were established that are homozygous for independent third chromosomes from males of the isofemale lines by standard crosses to *T(3)MKRS/TM2 ry Ubx* (obtained from A. CHOVNICK and S. CLARK) [see HILLIKER *et al.* (1980) for a description of this balancer stock]. The *MKRS* chromosome, marked by *Sb*, was the balancer used in these extractions.

Thirty inbred lines of *D. simulans* were established by brother-sister mating for 10 generations (balancers were not available for the *D. simulans* third chromosome).

Restriction map analysis: Genomic DNA from each line was prepared using a modification of the protocol of BINGHAM, LEVIS and RUBIN (1981). Lines were mapped using four restriction endonucleases: *Bam*HI, *Eco*RI, *Hin*dIII and *Sal*I. One microgram of total genomic DNA was digested with each of the enzymes and separated by size on 0.8% agarose gels using gel and electrode buffers containing 0.8 mM Tris, 0.4 mM acetic acid and 0.04 mM EDTA, pH 8.0. DNA fragments were transferred to nylon filters (Zetabind, commercially available from AMF Cuno) according to the method of REED and MANN (1985).

Filters were probed with *D. melanogaster* genomic DNA previously cloned into the phage Charon-4 by BENDER, SPIERER and HOGNESS (1983). Three overlapping clones, L2848, R2841 and R2838, were used to examine a 40-kb region around the *rosy* locus. Probes were nick-translated

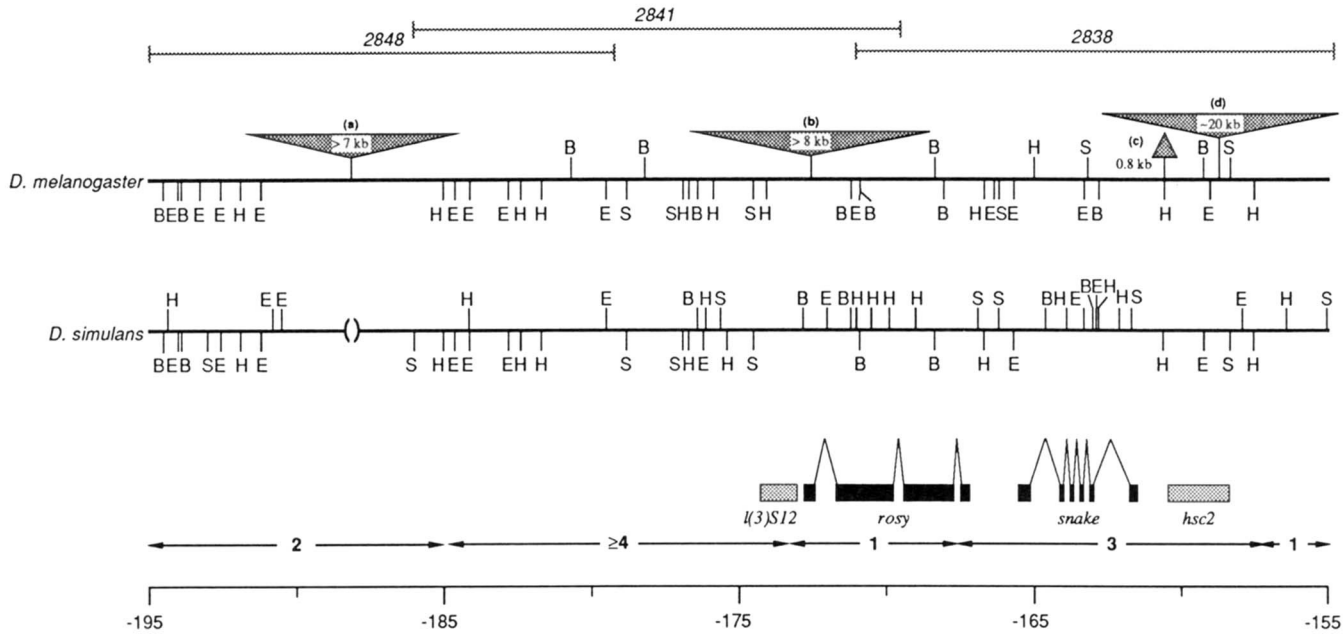

FIGURE 1.—Summary restriction site map of the *rosy* region in *D. melanogaster* (upper map) aligned with that of *D. simulans* (lower map). Lines above the maps labeled L2848, R2841 and R2838 indicate the regions probed by these three *D. melanogaster* DNA phage clones (BENDER, SPIERER and HOGNESS 1983). Polymorphic restriction sites are shown above the maps, fixed sites below the maps. B, E, H and S refer to the restriction endonucleases *Bam*HI, *Eco*RI, *Hin*dIII and *Sal*I. Map coordinates for each site were assigned by aligning the *D. simulans* map with the *D. melanogaster* map using the *D. melanogaster* clone coordinates (BENDER, SPIERER and HOGNESS 1983). Intraspecific insertions and deletions are indicated by triangles pointing toward or away from the map, respectively, and are shown approximately to scale. Insert locations are known only to the fragment shown. The deletion "c" eliminated *Hin*dIII site −160.5 and thus must bracket that location. Note that the *D. simulans* map is actually 200 bp shorter than that of *D. melanogaster* in the −190 to −186 region (indicated on the *D. simulans* map by parentheses). We have, however, retained the *D. melanogaster* coordinates to make the maps more easily comparable. Several minor differences between our genomic maps and those published by BENDER, SPIERER and HOGNESS (1983) were confirmed by remapping the appropriate phage clones. The map in Figure 1 represents the revised map. Genes and transcripts detectable from the 40-kb region are shown below the maps. Specific transcripts are indicated in cases when such information is available (DELOTTO and SPIERER 1986; KEITH *et al.* 1987) with known exons indicated by black boxes. Hatched boxes indicate transcripts for which exon locations are not available. The number of transcripts for the indicated segments are also shown (data from BOSSY, HALL and SPIERER 1984). Direction of transcription is from left to right for *rosy* as shown in the figure, and from right to left for *snake*.

using enzymes and protocol from Bethesda Research Labs. Filters were hybridized, washed and autoradiographed as suggested by the filter manufacturer (AMF Cuno) with the exception that final washes were carried out at 50°.

A restriction site map of the *rosy* region extending from *D. melanogaster* clone coordinates −155 to −195 kb (BENDER, SPIERER and HOGNESS 1983) was constructed for each of the 90 lines using single and double digests. In several cases it was necessary to probe with small fragments from the above phage (isolated in low melting point agarose and labeled by random priming; FEINBERG and VOGELSTEIN 1984) to confirm restriction site positions in *D. simulans*, particularly with respect to those in *D. melanogaster*. Minor disparities with the map for clones of this region from Canton S and Oregon R strains of *D. melanogaster* (BENDER, SPIERER and HOGNESS 1983), particularly in the −195 to −185 region, were confirmed by remapping of the phage clones (data not shown).

RESULTS

Summaries of restriction map variation in the 40-kb *rosy* region among 60 lines of *D. melanogaster* and 30 lines of *D. simulans* from Raleigh, North Carolina, are presented in Figure 1 and Tables 1 and 2. Below we describe the patterns of intraspecific variation revealed for each species and contrast them with the pattern of interspecific differentiation.

Intraspecific variation

Insertions and deletions: In contrast to *D. simulans*, where no evidence of scoreable intraspecific sequence length variation was found, 8% of the *D. melanogaster* lines differed from the most common restriction map by one or two sequence length variants. (We should have been able to detect all sequence length variation greater than about 100 bp.) Three large insertions (>7 kb to approximately 20 kb) and one 0.8-kb deletion were observed (Figure 1). Insertions "b" and "d" were unique in the sample, while insertion "a" and deletion "c" were both observed in three of the 60 *D. melanogaster* lines. That insertion "a" is identical in all three lines is suggested by identical sizes and patterns of restriction fragments observed for the insert region for all four enzymes. In addition, all three lines also have deletion "c" and are of the identical multisite genotype (haplotype 4, Table 1) suggesting that we have sampled three recently derived copies of the same chromosome. The enormous size of insertion

TABLE 1

Polymorphic restriction sites scored in the *rosy* locus region of *D. melanogaster*

Variant	Haplotype																Variant frequency
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Ins (a) -188.0	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	0.05
*Bam*HI -180.7	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	0.03
*Bam*HI -178.2	−	+	+	−	+	−	−	−	+	+	−	−	−	−	−	−	0.43
Ins (b) -172.5	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	0.02
*Bam*HI -168.35	+	+	+	+	+	+	+	+	−	−	+	+	+	+	+	−	0.05
*Hin*dIII -165.0	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	0.02
*Sal*I -163.2	−	−	+	−	−	+	+	−	+	−	+	−	+	+	−	+	0.30
Del (c) -160.5	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	0.05
*Bam*HI -159.2	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	−	0.05
Ins (d) -158.8	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	0.02
*Sal*I -158.3	+	+	+	+	−	+	+	−	+	+	+	+	+	+	−	+	0.10
Haplotype frequency	0.30	0.22	0.13	0.05	0.05	0.05	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	

The 60 lines of *D. melanogaster* are grouped and listed by distinct haplotype (multisite genotype). Restriction map coordinates are from Figure 1 and are listed in linear order. The positions given for the insertions (Ins) and deletion (Del) represent only approximate locations, generally just the midpoint of the smallest fragment to which they could be localized. The presence of a restriction map variant is indicated by a "+" and the absence by a "−". Variant frequency refers to the frequency of the least common state for a restriction map variant, either presence or absence. Haplotype frequency indicates the frequency of the haplotype in the sample.

"d" (~20 kb) was indicated by the apparent absence of any *Eco*RI sites in the inserted sequence (there are internal *Bam*HI, *Hin*dIII and *Sal*I sites).

While we have no direct evidence here, previous experience where large insertions were cloned and characterized (*e.g.*, LEIGH BROWN 1983; AQUADRO *et al.* 1986) would suggest that all of the insertions observed in the *rosy* region of *D. melanogaster* are transposable elements. The 0.8-kb segment removed by deletion "c" appears to be unique sequence since those sequences are contained in the DNA cloned in phage 2838 (see Figure 1). Use of this phage as a probe on genomic DNA Southern blots does not indicate hybridization to fragments from outside this region.

Insertion "b" as well as deletion "c" and insertion "d" are of interest since they appear to be located very close to or possibly within the transcriptional units of *rosy* or *l(3)S12* and *hsc2*, respectively. These lines are all homozygous and do not exhibit any obvious eye color or severe fitness differences, although more careful study of the expression of these genes in these lines is warranted. Insertion "a" is located in a region noticeably devoid of restriction sites and that also shows a 200-bp sequence length difference between *D. melanogaster* and *D. simulans*. The significance of this observation is unclear.

Restriction site polymorphism: A total of 41 sites were mapped in *D. melanogaster*, of which seven (17%) were variable in our sample. In contrast, 28 (50%) of the 56 restriction sites mapped in *D. simulans* were variable among the lines sampled. Both the number per region and frequency of restriction site variants reflect the differences in sequence polymorphism between the two species (Tables 1 and 2). The few polymorphic sites in *D. melanogaster* make it difficult to detect any specific patterns. However the large number of polymorphic sites in *D. simulans*, ranging in frequency from 0.03 for sites present only once, to 0.47 (Table 2) show an interesting spatial pattern. Sites located in the region from map coordinates −195 to approximately −175 show more skewed frequencies than sites in the region from −175 to −155; the average frequency in the former region being 0.10 (range of 0.03 to 0.23) and in the latter region being 0.23 (range of 0.03 to 0.47). This means that expected heterozygosity per restriction site is roughly twofold higher in the −175 to −155 region (heterozygosity calculated as $2pq$, where p is the variant frequency from Table 2 and $q = 1 - p$).

The frequency spectra for variable restriction sites for the entire region in *D. melanogaster* and *D. simulans* are contrasted in Figure 2 (also included for *D. melanogaster* are the frequencies of insertions and deletions, these being absent in *D. simulans*). The patterns observed are qualitatively similar to those observed for other regions in *D. melanogaster* (*e.g.*, AQUADRO *et al.* 1986; LANGLEY and AQUADRO 1987). In particular, large insertions and deletions are in low frequency, while restriction sites are observed at all frequencies.

Levels of heterozygosity per nucleotide (π, NEI and TAJIMA 1981) are presented in Table 3 for the entire region, and then separately for the regions from −195 to −175 kb, and from −175 to −155 kb (dividing the 40-kb region in half along what appears from visual inspection of Figure 1 to be a natural break in the data). Heterozygosity (per nucleotide) over the entire *rosy* region in *D. simulans* was estimated to be 1.9% compared with 0.3% for *D. melanogaster*. However,

TABLE 2

Polymorphic restriction sites scored in the *rosy* locus region of *D. simulans*

Variant	Haplotype																							Variant frequency
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
*Hin*dIII -194.35	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	0.13
*Eco*RI -190.8	−	−	−	−	−	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0.10
*Eco*RI -190.5	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	0.03
*Hin*dIII -184.15	−	−	−	−	+	+	−	−	−	+	+	+	−	−	−	−	−	+	−	−	−	−	−	0.20
*Eco*RI -179.5	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	0.03
*Bam*HI -176.4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	−	+	+	+	0.07
*Hin*dIII -176.1	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	0.03
*Sal*I -175.6	−	+	−	−	−	−	+	−	−	−	−	−	+	−	−	−	−	−	+	−	+	+	−	0.23
*Bam*HI -172.8	−	−	+	−	−	−	−	+	+	−	−	−	−	+	+	−	−	−	−	−	−	−	−	0.20
*Eco*RI -172.0	+	−	−	−	−	−	−	−	−	−	+	−	−	+	−	+	−	−	−	−	−	−	−	0.27
*Bam*HI -171.2	+	−	−	−	−	−	−	−	−	+	−	−	−	+	−	−	−	−	−	−	+	+	−	0.30
*Hin*dIII -171.0	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	0.03
*Hin*dIII -170.5	+	+	−	+	+	+	+	−	−	+	+	+	+	+	−	+	+	+	+	+	+	+	+	0.17
*Hin*dIII -169.9	+	−	+	+	+	+	−	+	+	+	+	−	+	+	+	+	+	+	−	+	−	−	−	0.27
*Hin*dIII -169.0	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	0.03
*Sal*I -166.9	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	0.20
*Sal*I -166.2	+	+	−	+	+	+	+	−	−	+	+	+	−	+	−	+	+	+	+	+	+	+	+	0.20
*Bam*HI -164.6	+	−	+	−	−	−	−	+	+	+	+	−	−	+	+	+	+	+	−	+	+	+	−	0.37
*Hin*dIII -163.9	−	−	+	−	−	−	−	+	+	−	−	−	−	−	+	−	−	−	−	−	−	−	−	0.17
*Eco*RI -163.3	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	0.20
*Eco*RI -163.0	−	−	−	−	+	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	+	−	−	0.10
*Hin*dIII -162.9	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	−	−	−	0.10
*Bam*HI -162.8	+	−	−	−	−	−	−	−	−	+	+	−	+	+	−	+	+	+	−	+	+	+	+	0.47
*Hin*dIII -162.1	−	+	−	+	+	+	+	−	−	−	−	+	+	−	−	−	+	+	+	+	+	+	+	0.47
*Sal*I -161.7	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	0.20
*Eco*RI -157.9	−	+	−	−	−	−	+	−	−	−	−	−	+	−	−	−	+	−	+	−	−	−	+	0.23
*Hin*dIII -156.4	−	−	+	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	0.13
*Sal*I -155.0	+	+	−	−	−	−	+	−	−	+	+	−	−	+	−	+	+	+	+	+	−	−	+	0.43
Haplotype frequency	0.17	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	

The 30 lines of *D. simulans* are grouped and listed by distinct haplotype as in Table 1.

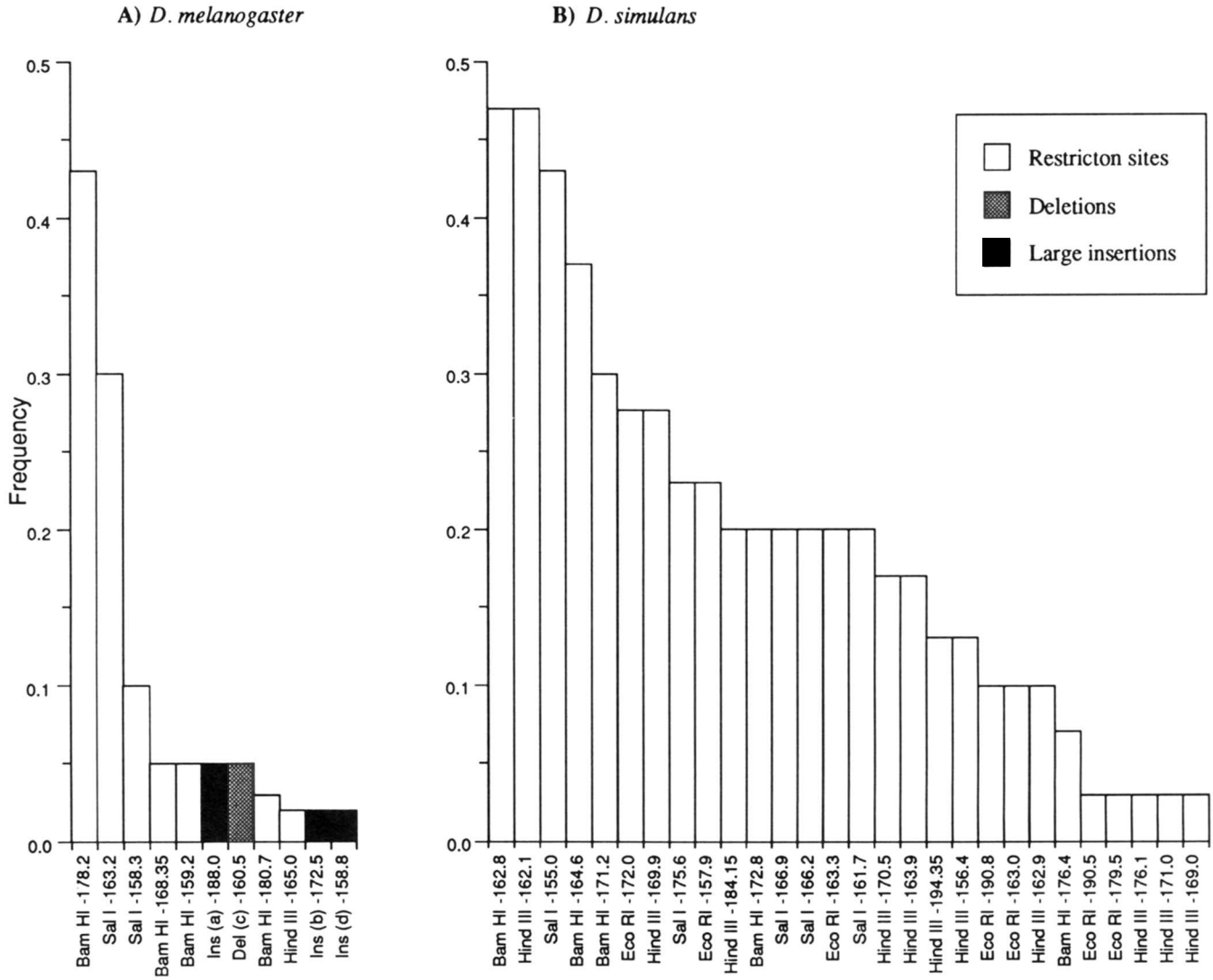

FIGURE 2.—Frequency spectra for naturally occurring restriction map variants in the 40-kb *rosy* region of (A) *D. melanogaster* and (B) *D. simulans*. Shown is the least frequent character at each site (refer to Figure 1 and Tables 1 and 2).

variation was not distributed evenly over the region, but rather was concentrated in the region from −175 to −155. The level of heterozygosity in this region was 3.9%, more than six times the 0.6% estimated for the region from −195 to −175. A similar, though less dramatic, trend is seen for *D. melanogaster* (Table 3).

Interspecific divergence

We made a special effort to align the *rosy* region restriction map of *D. simulans* with that of *D. melanogaster*. Sites within roughly 200 bp on our maps were conservatively scored as homologous due to the inherent difficulties in restriction mapping and since subtle differences in sequence length due to unscoreable insertion/deletion variation makes it likely that truely homologous sites of identical sequence may not always exactly line up. A strict consideration of only perfectly aligned sites gives virtually identical estimates and conclusions for this and the following analyses.

Average nucleotide divergence (NEI and TAJIMA 1981) across the 40-kb *rosy* region between the 30 lines of *D. simulans* and the 60 lines of *D. melanogaster* is 0.050 (Table 3). Taking into account intraspecific polymorphism indicates that the net nucleotide divergence that has accumulated between the two species in the 40-kb region since divergence from a common (presumably polymorphic) ancestor is 0.038 [= 0.049 − (0.019 + 0.003)/2; following NEI and TAJIMA (1981); see Table 3]. Considering two nonoverlapping 20-kb segments, divergence appears roughly twofold higher in the *rosy/snake/hsc2* region than the adjacent 20 kb (Table 3).

Of particular interest is the distribution of interspecific divergence compared to the distribution of intraspecific variation across the 40-kb *rosy* region. We have contrasted intraspecific heterozygosity with interspecific divergence for the entire 40-kb region, 10 kb at a time, in Figure 3. Divergence is highest over

the *rosy* transcript while heterozygosity peaks over *snake* and *hsc2*. This divergence appears to represent the accumulation of substitutions since speciation together with the complete sorting of ancestral polymorphism since none of the 67 mapped sites were observed to be variable in both species. Of the 35 sites polymorphic in one or the other species 10 are located in the −195 to −175 region while 25 are in the adjacent −175 to −155 region. Surprisingly, of the ten sites unique and fixed in one or the other species, six of them are in the more conserved and less polymorphic −195 to −175 region with only 4 located in the −175 to −155 region.

KREITMAN and AGUADÉ (1986) noted a strong disparity between the pattern of nucleotide polymorphism in the alcohol dehydrogenase (*Adh*) gene region of *D. melanogaster*, compared to divergence from a single sequence of *D. simulans*. These authors pointed out that under the hypothesis of selective neutrality (KIMURA 1983), interspecific divergence is simply and positively correlated with intraspecific polymorphism; highly polymorphic regions are expected to be rapidly evolving due to a higher mutation rate to selectively neutral alternatives. HUDSON, KREITMAN and AGUADÉ (1987) have formalized this concept into a test of departure from selective neutrality expectation given intraspecific and interspecific estimates of DNA sequence variation. We have applied their test to our data in the following manner.

We divided the 40-kb *rosy* region into two 20-kb segments (coordinates −195 to −175 we will call region 1, and −175 to −155 called region 2; see Figure 1). The number of alleles sampled for *D. melanogaster* (n_A) and *D. simulans* (n_B) were 60 and 30, respectively. The numbers of segregating restriction sites (S_i) in regions 1 and 2 of *D. melanogaster* were 2 and 5, respectively, and 8 and 20 for *D. simulans*. The average number of sites differing between pairs of chromosomes from the two species (D_i) for regions 1 and 2 were 7.30 and 11.92, respectively. Solution of the simultaneous equation model of HUDSON, KREITMAN and AGUADÉ (1987) leads to estimates of $T = 9.95$, $f = 4.71$, $\theta_1 = 0.48$ and $\theta_2 = 1.02$ (for regions 1 and 2, respectively). T is an estimate of the number of generations since divergence of the two species divided by twice the effective population size; f is the factor by which the effective population size (N_e) for *D. simulans* is larger than that of *D. melanogaster*; and θ_i is an estimate of $4N_e\mu$, where μ is an estimate of the neutral mutation rate for region i. The observed and expected values of segregating sites and divergence are not significantly different from the neutral expectation [$\chi^2 = 0.42$ with 2 d.f.; see HUDSON, KREITMAN and AGUADÉ (1987) for further details]. This analysis does indicate a twofold difference in θ between the two regions suggesting variation in the neutral mutation rate. Overall, the data are also consistent with a 4.7-fold larger effective population size for *D. simulans vs. D. melanogaster*, as reflected in the high number of segregating sites observed in the former *vs.* the latter species.

DISCUSSION

D. melanogaster and *D. simulans* are closely related and morphologically almost indistinguishable species (STURTEVANT 1920). Previous comparisons of DNA restriction maps and sequences between *D. melanogaster* and *D. simulans* in the 87A and 87C heat-shock gene region (LEIGH BROWN and ISH-HOROWICZ 1981) and the *Adh* gene region (LANGLEY, MONTGOMERY and QUATTLEBAUM 1982; COHN, THOMPSON and MOORE 1984; BODMER and ASHBURNER 1984) as well as a study of DNA-DNA hybridization between single copy DNA in *D. simulans* and *D. melanogaster* (ZWIEBEL *et al.* 1982) have suggested that the two species differ at the nucleotide level by 2–4%. These results are consistent with our use of *D. melanogaster* DNA clones of the *rosy* region to probe genomic DNA from *D. simulans* and with our estimate of net sequence divergence of 3.8% over the 40-kb *rosy* region. Although our filters were probed under relatively stringent conditions, the probes hybridized strongly and the filters were easy to score. Moreover, the fact that we were able to construct a coherent restriction site map of the region indicates that the *rosy* region in *D. simulans* has the same general structure as the *rosy* region in *D. melanogaster*.

Previous studies of genetic and morphological variation in these two species have led to two generalizations that are relevant to our data (reviewed by CHOUDHARY and SINGH 1987b). First, protein sequence (allozymes), morphology and chromosomal arrangement are generally much less variable within and geographically differentiated between populations of *D. simulans* compared to *D. melanogaster*. Second, copy number of transposable elements has been hypothesized to be as much as eightfold lower in the genome of *D. simulans* than in *D. melanogaster* (DOWSETT and YOUNG 1982). Our data support the hypothesis of fewer transposable elements per genome in *D. simulans* compared to *D. melanogaster*. However, a more than sixfold *higher* level of heterozygosity per nucleotide in *D. simulans* relative to *D. melanogaster* stands in apparent strong contrast to previous data. Below we first discuss the patterns of insertion/deletion variation, then consider the levels and patterns of DNA versus protein variation in the two species. Finally, we examine possible explanations for these results.

Species differences in insertions and deletions: Eight percent of the *D. melanogaster* lines compared (5 out of 60) had a large insertion and/or deletion

TABLE 3

DNA sequence variation and divergence estimated for the *rosy* locus region

	Region examined		
	−195 to −175	−175 to −155	Entire 40 kb
Heterozygosity (π)			
D. melanogaster	0.002 ± 0.0002	0.004 ± 0.0006	0.003 ± 0.0003
D. simulans	0.006 ± 0.0009	0.040 ± 0.0027	0.019 ± 0.0009
Divergence between			
D. melanogaster and *D. simulans*	0.033 ± 0.0007	0.073 ± 0.0019	0.050 ± 0.0007
(net divergence)	(0.029 ± 0.0006)	(0.052 ± 0.0019)	(0.038 ± 0.0007)

Estimates (±SE) of nucleotide heterozygosity and divergence estimated from restriction site variation among lines of *D. melanogaster* and *D. simulans* from Raleigh, North Carolina. Standard errors given are simply the square root of the variance. Heterozygosity (π) and its sampling variance were estimated using Equations 18 and 11 in NEI and TAJIMA (1981). Divergence was estimated for all pairwise comparisons of the 30 *D. simulans* and 60 *D. melanogaster* chromosomes following NEI and TAJIMA (1981). Net divergence takes into account intraspecific nucleotide diversity (π) for each species (see text). Standard errors for total and net divergence are the square root of sampling variances given by equations 26 and 25, respectively, in NEI and TAJIMA (1981).

relative to the most common restriction map haplotype. This falls on the low end of the range observed in other regions of the *D. melanogaster* genome [*rosy*: 0.003 insertions or deletions over 100 bp in length per chromosome examined per kilobase surveyed; compared to 0.024 for *Adh*: AQUADRO *et al.* (1986) and CROSS and BIRLEY (1986); 0.008 for *white*: LANGLEY and AQUADRO (1987); 0.037 for 87A heat shock locus: LEIGH BROWN (1983); and 0.002 for *Notch*: SCHAEFFER, AQUADRO and LANGLEY (1988)].

In contrast, no sequence length variation was detected among the 30 *D. simulans* chromosomes. We have also examined restriction map variation in the 12-kb *Adh* region in these same lines of *D. simulans* and, like in the *rosy* region, found a complete absence of detectable sequence length variation [limit of resolution approximately 100 bp (C. AQUADRO, K. SYKES and P. NELSON, unpublished data)]. The lack of large insertions detected in the *rosy* and *Adh* regions of *D. simulans* supports the generality of fewer copies per genome of dispersed repetitive, nomadic DNA compared to *D. melanogaster* suggested by the studies of DOWSETT and YOUNG (1982). As transposable elements have been implicated as causes of gene rearrangements in *D. melanogaster* (*e.g.*, ENGELS 1983; GOLDBERG *et al.* 1983; DAVIS, SHEN and JUDD 1987), this may contribute to the lack of inversion polymorphism in *D. simulans* (ASHBURNER and LEMEUNIER 1976). This explanation alone may not be satisfactory because *D. pseudoobscura*, which has a large number of polymorphic inversions, also appears to possess very low levels of insertion and deletion variation over 100 bp in length [0.002 per chromosome per kilobase for the 32-kb *Adh* region, none of which was over 200 bp in size (SCHAEFFER, AQUADRO and ANDERSON 1987)].

The absence of large sequence length polymorphism in both the *rosy* and *Adh* regions of *D. simulans* does not imply that transposable elements do not cause mutations in this species, for they clearly have been observed [*Notch* locus (KIDD and YOUNG (1986); *white* locus (INOUE and YAMAMOTO 1987)]. Our data do indicate that transposable elements are a much less prevalent source of variation in natural populations of *D. simulans* than *D. melanogaster*. There is clearly less opportunity in *D. simulans* for the occurence of rearrangements caused by recombination between homologous elements in nonhomologous chromosomal locations that have been demonstrated in *D. melanogaster* (GOLDBERG *et al.* 1983; DAVIS, SHEN and JUDD 1987). That insertion/deletion variation is also rare in *D. pseudoobscura* (SCHAEFFER, AQUADRO and ANDERSON 1987) suggests that the pattern of molecular variation observed in *D. melanogaster* may not be typical of other species of *Drosophila*. Whether transposable element copy number has recently increased in *D. melanogaster* relative to other *Drosophila* is unknown and warrants further study.

DNA *vs.* protein variation: Restriction site variation in the *rosy* region of *D. simulans* leads to an estimate of heterozygosity per nucleotide roughly six times higher than that observed in *D. melanogaster* for the homologous *rosy* region (0.019 *vs.* 0.003). The variation estimate for *rosy* in *D. melanogaster* is comparable to that seen in other regions of the genome for that species: 0.002 for the 87A heat shock locus, 0.006 for *Adh* and 0.007 for *Notch* (LEIGH BROWN 1983; AQUADRO *et al.* 1986; SCHAEFFER, AQUADRO and LANGLEY 1988). Only the *white* locus region in *D. melanogaster* appears to be nearly as variable (0.014) as observed for the *rosy* region in *D. simulans*. However, the *white* region estimate comes from lines collected worldwide, while the *D. simulans* are all from a single locality. In addition, a survey of the *Adh* region of *D. simulans* in these same lines indicates several fold higher levels of nucleotide variation relative to *D. melanogaster* (C. AQUADRO, K. SYKES and P. NELSON, unpublished data). Together, these results suggest that the standing level of DNA sequence variation

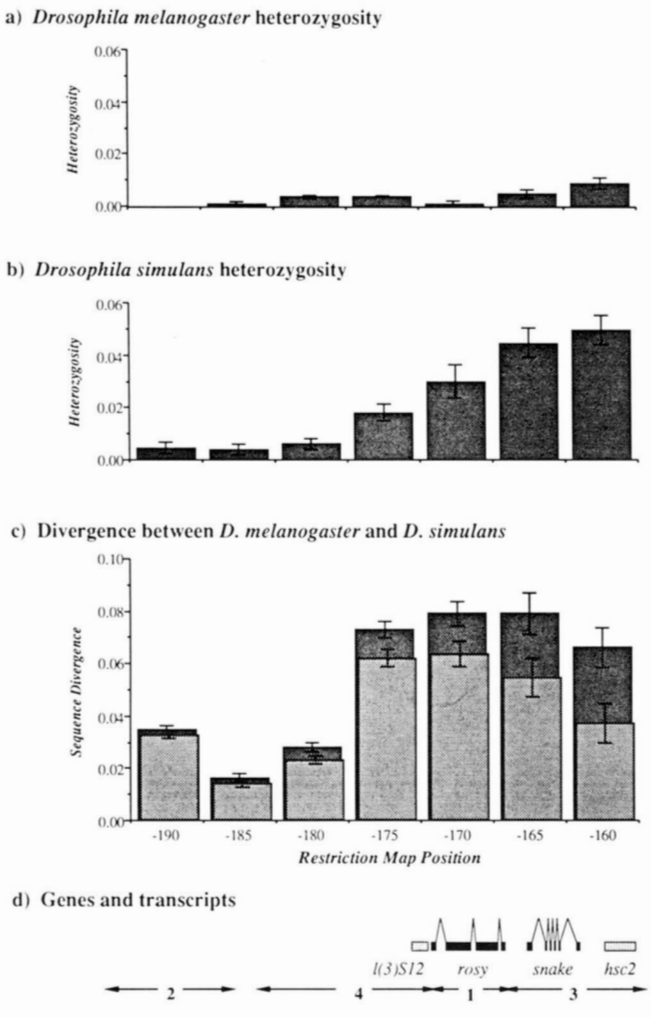

FIGURE 3.—Summary of intra- and interspecific DNA sequence variation estimated for the *rosy* locus region in *D. melanogaster* and *D. simulans*. Nucleotide heterozygosity and divergence are graphed for a 10-kb "window" slid along the 40-kb region 5 kb at a time: (a) heterozygosity per nucleotide (π) among 60 lines of *D. melanogaster*; (b) heterozygosity per nucleotide among the 30 lines of *D. simulans*; (c) average nucleotide divergence (NEI and TAJIMA 1981) between the *D. melanogaster* and the *D. simulans* (uncorrected and corrected for intraspecific variation indicated by dark and light gray, respectively); (d) genes and transcripts detectable from the 40-kb region. Specific transcripts are indicated in cases when such information is available (DELOTTO and SPIERER 1986; KEITH *et al.* 1987). The number of transcripts for the indicated segments are also shown (data from BOSSY, HALL and SPIERER 1984). Error bars indicate the 95% confidence interval (= 1.96 × SE) for each estimate of variation or differentiation (NEI and TAJIMA 1981).

due to nucleotide substitutions is perhaps sixfold higher in *D. simulans* than in *D. melanogaster*. The direction and magnitude of this difference is remarkable given the substantial body of literature on protein polymorphism, morphology and chromosomal variation that has indicated *D. melanogaster* to be the more polymorphic of the two species.

Over 100 enzymes and general proteins have been examined in numerous populations of both species, and *D. melanogaster* has consistently appeared to be more variable, *on average*, than *D. simulans* (Table 4). Several recent, extensive analyses have confirmed that the proportion of allozyme loci polymorphic is significantly higher (roughly twofold) in *D. melanogaster* than in *D. simulans* (*e.g.*, HYYTIA *et al.* 1985; WATADA, TOBARI and OHBA 1986; SINGH, CHOUDHARY and DAVID 1987; CHOUDHARY and SINGH 1987b). However, these workers have found that in several localities in Europe, Africa and Japan consideration of only those loci that are polymorphic shows *D. simulans* to have similar or perhaps slightly increased levels of protein heterozygosity and average number of electromorphs compared to *D. melanogaster* from the same localities (see in particular Tables 5 and 6 of CHOUDHARY and SINGH 1987b). Studies of morphological and quantitative characters and inversion polymorphisms appear to lend support to the hypothesis that the entire genome of *D. simulans* is less variable than that of *D. melanogaster* (HYYTIA *et al.* 1985; ASHBURNER and LEMEUNIER 1976; reviewed in CHOUDHARY and SINGH 1987b).

Allozyme polymorphism at *rosy* appears to fit this general pattern: in a study of 216 isofemale *D. simulans* lines from Europe and Africa, M. CHOUDHARY and R. SINGH (1987b and personal communication) found five XDH electromorphs and a per locus heterozygosity of 0.54. A previous study of 262 isofemale *D. melanogaster* lines in the same laboratory, using the same standard electrophoretic conditions, had found eight XDH allozymes and a heterozygosity of 0.52 (SINGH, HICKEY and DAVID 1982). Similarly, BUCHANAN and JOHNSON (1983) found seven XDH allozymes in 62 isochromosomal lines of *D. melanogaster* under standard conditions.

TABLE 4

Comparison of allozyme and DNA variation in *D. melanogaster* and *D. simulans*

	D. melanogaster	*D. simulans*
Allozymes[a]		
Proportion of loci polymorphic	0.420 ± 0.070	0.289 ± 0.025
Average number of alleles per locus	1.48 ± 0.11	1.62 ± 0.03
Heterozygosity per locus	0.102 ± 0.014	0.096 ± 0.012
Fixation index (F_{ST})	0.091 ± 0.130	0.025 ± 0.052
Genetic distance	0.031 ± 0.015	0.013 ± 0.006
DNA—*rosy* region[b]		
Heterozygosity per nucleotide	0.003 ± 0.0003	0.019 ± 0.0009
Large insertion/deletion variation (per line per kb)	0.003	0.000

Estimates are given ± 1 SE.

[a] Allozyme data from Table 5 in CHOUDHARY and SINGH (1987b) (mainland populations of *D. simulans*). Data represents 117 loci for 15 populations in *D. melanogaster* and 114 loci for four populations in *D. simulans*. Estimates are over all loci unless stated.

[b] DNA data from 40-kb *rosy* region, present study.

The high level of restriction site polymorphism we have observed in *D. simulans* is particularly striking given that both our *D. simulans* and *D. melanogaster* samples represent North American *Drosophila* populations. In previous studies of allozyme polymorphism, North American *D. simulans* populations have shown substantially lower levels of protein variation than North American *D. melanogaster* populations (O'BRIEN and MACINTYRE 1969; BERGER 1970; OHNISHI *et al.* 1982; even when only the polymorphic loci are considered). In light of our results, it will be of particular interest to compare the levels of DNA polymorphism in European and African populations of *D. simulans* and *D. melanogaster*.

One would expect that levels of variation in mitochondrial DNA (mtDNA) should be positively correlated with levels of nuclear gene variation across species in the absence of substantially different kinds or patterns selection acting on the two molecules. A consideration of published mtDNA data for both species suggests that considering each species as a whole, *D. simulans* is approximately threefold more variable than *D. melanogaster*, a difference similar in direction and magnitude to that observed for nucleotide variation in the *rosy* nuclear gene region. A large survey has recently been published analyzing mtDNA variation in *D. melanogaster* [92 isofemale lines from throughout the world analyzed with 10 restriction enzymes (HALE and SINGH 1987)]. To date, the only similar scale analysis of *D. simulans* in terms of large numbers of enzymes and wide geographic representation is that of BABA-AÏSSA and SOLIGNAC (1984) (see also SOLIGNAC and MONNEROT 1986) in which 13 lines, each from a separate locality from around the world, were analyzed by 12 restriction enzymes. Considering each species as a whole, nucleotide heterozygosity (π) for mtDNA is 0.002 for *D. melanogaster* and 0.006 for *D. simulans*. Clearly additional samples of particularly *D. simulans* are needed to confirm this pattern, but *D. simulans* appears to be at least as variable if not more variable for mtDNA than *D. melanogaster*, consistent with our *rosy* region data from the nuclear genome. We should note that SHAH and LANGLEY (1979) found less nucleotide polymorphism for mtDNA in five lines of *D. simulans* than in ten lines of *D. melanogaster*, each species sampled from a single locality. However, two of three variable sites in their *D. melanogaster* lines (their *Hin*dIII "b" and "c" sites) turn out be be due to length variation in the A + T rich region and not base substitution (we confirmed this by examining photographs of the original gels for the other three enzymes they used; photos kindly provided by C. H. LANGLEY). We feel the results of the more extensive surveys cited above provide our best available picture of mtDNA variation in these two species. We will return to the contrast between protein variation and DNA polymorphism after a consideration of intra- *vs.* interspecific variation in the *rosy* region.

Differences in heterozygosity and divergence along the sequence: Despite the disparity in restriction site polymorphism, the two species show a similar distribution of variation across the *rosy* region. Nucleotide heterozygosity in *D. simulans* in the region containing *hsc2* is as much as 12 times that in the adjacent region (see Table 3 and Figure 3). A similar pattern, though not as extreme, is also found in *D. melanogaster*. The "density" or abundance of transcriptional units across regions does not appear to be strongly associated with either divergence or polymorphism suggesting that the patterns are not simply due to a comparison of coding versus noncoding sequences (Figure 3).

Similarly juxtaposed regions of high *vs.* low polymorphism or divergence have been observed in the major histocompatibility complex among strains of mice and individual humans (STEINMETZ *et al.* 1984) and in the 68C glue gene complex among species of *Drosophila* (MARTIN and MEYEROWITZ 1986). KREITMAN and AGUADÉ (1986) and HUDSON, KREITMAN and AGUADÉ (1987) have argued that the significant departure from neutrality expectations that they observed for the *Adh* region (using polymorphism estimates for *D. melanogaster*, and comparing a single *D. melanogaster* sequence to a single sequence of either *D. simulans* or *D. sechellia*) was due to an "excess" of nucleotide polymorphism in the *Adh* gene. Such a pattern would be caused by balanced polymorphism at the locus (they argue in fact for at least two balanced polymorphisms at *Adh*).

Our data for the *rosy* region indicate significantly more heterozygosity in the *rosy*, *snake* and *hsc2* gene region compared to the adjacent 20 kb, particularly in *D. simulans*. However in our case, the highly polymorphic region also appears to diverge more rapidly between species. The application of the HUDSON, KREITMAN and AGUADÉ test to our data in fact shows no significant departure from selective neutrality predictions. Since equilibrium heterozygosity is approximately equal to $4N_e\mu$ under an infinite allele neutral model, our results suggests as much as an eightfold higher neutral mutation rate for the *rosy-snake-hsc2* region compared to region −195 to −175. (A twofold difference is suggested from the HUDSON, KREITMAN and AQUADÉ model considering both species together.) Assuming these differences are biologically significant, a neutralist interpretation would call for stronger purifying selection in the −195 to −175 region compared to the *rosy*, *snake* and *hsc2* region or an intrinsic difference in mutation rate. We should note that the association between polymorphism and divergence is not exact and needs to be investigated

further with additional data (of particular interest is the *hsc2* gene where divergence falls off yet heterozygosity is highest). In addition, the relatively small number of restriction sites compared in our analysis make the application of the HUDSON, KREITMAN and AGUADÉ test rather weak. The use of four-base restriction endonucleases and/or direct DNA sequence data would clearly be preferable (KREITMAN and AGUADÉ 1986; HUDSON, KREITMAN and AGUADÉ 1987) and are being pursued.

A model of different effective population sizes: Drawing largely on analyses of extensive allozyme data sets, a variety of hypotheses have been posed to explain the apparently lower level of within-population variation and geographic differentiation observed for *D. simulans* versus *D. melanogaster* (SINGH, CHOUDHARY and DAVID 1987; CHOUDHARY and SINGH 1987b). These include the (1) neutral-mutation hypothesis, (2) population bottleneck and recent colonization hypothesis, (3) niche-width hypothesis, and (4) selection hypothesis (different genetic "strategies" of adaptation). SINGH, CHOUDHARY and DAVID (1987) concluded that a population bottleneck in *D. simulans* (hypothesis 2) and variation in niche-width between the species (hypothesis 3) were the two most likely explanations for the lower level of variation and differentiation observed in *D. simulans*. CHOUDHARY and SINGH (1987b) revised this conclusion to favor variation in niche-width and/or "genetic 'strategies' of adaptation" based on the reduced difference in allozyme heterozygosities observed when substantially larger numbers of loci were examined.

We advance the hypothesis that a larger species effective population size in *D. simulans* relative to *D. melanogaster* is the major determinant of the contrasting patterns of nucleotide diversity, insertion/deletion variation and protein polymorphism in *D. simulans* and *D. melanogaster*. As detailed below, a difference in effective population size (N_e) would lead to a large increase in strictly neutral variation (perhaps many synonymous and noncoding sequences). It would, however, only slightly increase variation for characters under purifying selection (perhaps many allozymes) due to an increase in the efficacy of selection as genetic drift is reduced. Note that the effect of positive Darwinian selection, as well as purifying selection would be strengthened under this hypothesis and could account for the diverse patterns of variation and differentiation seen among allozyme loci for both species (reviewed in CHOUDHARY and SINGH 1987b).

Under the infinite allele, strictly neutral model, the expected heterozygosity is approximately equal to $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate to selectively neutral alternative alleles. The roughly six times higher level of nucleotide heterozygosity in *D. simulans vs. D. melanogaster* would indicate that either N_e or μ (or a combination thereof) was approximately six times larger in *D. simulans* (a difference of 4.7 was estimated in our application of the HUDSON, KREITMAN and AGUADÉ analysis). SCHAEFFER, AQUADRO and ANDERSON (1987) have recently found nucleotide variation in the *Adh* region of *Drosophila pseudoobscura* to also be four times higher than has been reported in *D. melanogaster*. This difference between *D. pseudoobscura* and *D. melanogaster* is similar in direction (but of a substantially larger magnitude) than the pattern of protein polymorphism, heterozygosity and number of alleles reported by CHOUDHARY and SINGH (1987a) which these authors interpret as being due to a difference in effective population size between the species. However, SINGH and RHOMBERG (1987) concluded that differences in the pattern of geographic differentiation between these two species has resulted from differences in migration and natural selection. CHOUDHARY and SINGH (1987b) report that N_em (the product of effective population size and the migration rate m; estimated from F_{ST}) is two to four times lower in *D. melanogaster* than in *D. simulans*. These authors interpreted these data to mean that migration rates (m) are lower among populations of the former species than the latter. Our estimates of restriction site variation lead us to believe that it is effective population size, and not migration rate or mutation rate, that is particularly low in *D. melanogaster* relative to both *D. simulans* and *D. pseudoobscura*. Such an interpretation would also reconcile the estimates of N_em with what are considered to be a higher likelihood of gene flow in *D. melanogaster* compared to the other two species given the close association with *D. melanogaster* and human habitation (CHOUDHARY and SINGH 1987b).

A substantial difference in effective population size could have a significant effect on the distribution of genetic variation if much of the selection acting on the genes and phenotypes examined is purifying selection. The effect on the level of heterozygosity of differences in effective population size depends on the distribution of selective effects on new mutants. OHTA (1976) modified the strictly neutral theory of molecular evolution to include slight deleterious selection for most variants. Under this model (KIMURA 1983) mutants increasingly behave as neutral variants when $2N_es < 1$, where s is the selective disadvantage of the mutant. A sixfold increase in N_e will mean that a smaller proportion of mutants will be effectively neutral (mutants with selection coefficients up to six times smaller will now be influenced more by selection than by genetic drift). In other words, the intensity of purifying selection will increase as effective population size increases. KIMURA (1979) has shown that despite the increase in selection intensity, heterozygosity is expected to increase, although the rate of increase

with a given increase in N_e varies with various parameters in his model and is less than would be expected under a strictly neutral model. The increase can be relatively small for even a tenfold increase in N_e, and would be expected to be larger for silent and noncoding nucleotide variation than for phenotypes such as proteins.

The application of this model to *D. simulans* and *D. melanogaster*, assuming a sixfold larger effective population size for the former species, leads to predictions generally consistent with the observed data if we assume that much of the nucleotide polymorphism we estimate is located in synonymous, noncoding and nonregulatory sites. We would expect a sizeable increase in restriction site variation and only a small increase in per locus heterozygosity for allozymes. This is the pattern observed. The marked decrease in proportion of allozyme loci polymorphic would have to be due, under this scenario, to a loss of low frequency alleles that at larger population size became sufficiently deleterious to be dramatically reduced in frequency. Again, the data are consistent with this prediction: the nearly twofold higher proportion of loci polymorphic in *D. melanogaster* relative to *D. simulans* "is mostly due to the fact that *D. melanogaster* harbors low frequency alleles at many loci which are monomorphic in *D. simulans*" [CHOUDHARY and SINGH (1987b), p. 706]. The reduction in variation of morphology, karyotype and dispersed middle-repetitive DNA (including copy number of transposable elements) suggests an analogous increase in purifying selection on these characters caused by differences in effective population size between the species. The apparently higher level of mitochondrial DNA variation noted previously for *D. simulans* compared to *D. melanogaster* is consistent with both the higher effective population size in the former species and with the interpretation that much of the mitochondrial DNA site variation observed is effectively neutral.

Our results indicate that protein polymorphism is not an accurate indicator of the underlying nucleotide variation. In addition, our findings suggest the testable hypothesis that the products of the *rosy* and other protein-coding loci are under different kinds and/or levels of selection in *D. simulans* compared to *D. melanogaster*. This hypothesis predicts that much of the nucleotide polymorphism in *D. simulans* must be occurring at synonymous, noncoding and/or nonregulatory sites. Direct DNA sequence analysis of population samples of *rosy* alleles from both species should allow us to test this hypothesis in the following manner. A simple increase in the mutation rate should raise the level of polymorphism in synonymous, nonsynonymous and noncoding sites in a roughly proportionate manner. In contrast, stronger purifying selection as a result of a larger effective population size in *D. simulans* should lead to an increase in the level of nucleotide variation primarily in synonymous and noncoding sites where selection is likely to be weak or absent, resulting in a decrease in the ratio of nonsynonymous to synonymous substitutions in *D. simulans* as compared to that seen in *D. melanogaster*. A similar decrease should be observed when comparing nonsynonymous to noncoding sites. It will also be important to examine the nature, level and distribution of DNA sequence variation in regions adjacent to the *rosy* region and at unlinked loci in these two species.

This work was supported by grants to C.F.A. from the National Institutes of Health (GM-36431) and the Cornell Biotechnology Program. We appreciate the assistance of LYNN SWIATKOWSKI, VIPIN BANSAL, THOMAS BOYCE and H. KERN REEVE, and the comments of RICHARD HUDSON, ROSS MACINTYRE, RICHARD HARRISON, CAROL YOON and two anonymous reviewers on earlier versions of this manuscript.

LITERATURE CITED

ASHBURNER, M., and F. LEMEUNIER, 1976 Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). I. Inversion polymorphism in *Drosophila melanogaster* and *Drosophila simulans*. Proc. R. Soc. Lond. (Biol.) **193:** 137–157.

AQUADRO, C. F., S. F. DEESE, M. M. BLAND, C. H. LANGLEY and C. C. LAURIE-AHLBERG, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics **114:** 1165–1190.

BABA-AÏSSA, F., and M. SOLIGNAC, 1984 La plupart des populations de *Drosophila simulans* ont probablement pour ancêtre une femelle unique dans un passé récent. C. R. Acad. Sci. (Paris) **299:** 289–292.

BENDER, W., P. SPIERER and D. S. HOGNESS, 1983 Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the Bithorax complex in *Drosophila melanogaster*. J. Mol. Biol. **168:** 17–33.

BERGER, E. M., 1970 A comparison of gene-enzyme variation between *Drosophila melanogaster* and *Drosophila simulans*. Genetics **66:** 677–683.

BINGHAM, P. M., M. G. KIDWELL and G. M. RUBIN, 1982 The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain specific transposon. Cell **29:** 995–1004.

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

BODMER, M., and M. ASHBURNER, 1984 Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. Nature **309:** 425–430.

BOSSY, B., L. M. C. HALL and P. SPIERER, 1984 Genetic activity along 315 kb of the *Drosophila* chromosome. EMBO J. **3:** 2537–2541.

BUCHANAN, B. A., and D. L. E. JOHNSON, 1983 Hidden electrophoretic variation at the xanthine dehydrogenase locus in a natural population of *Drosophila melanogaster*. Genetics **104:** 301–315.

CHOVNICK, A., W. GELBART and M. MCCARRON, 1977 Gene organization in higher organisms. pp. 445–468. In: *The Organization and Expression of the Eukaryotic Genome*, Edited by E. M. BRADBURY and K. JAVAHERIAN. Academic Press, London.

CHOUDHARY, M., and R. S. SINGH, 1987a Historical effective size and the level of genetic diversity in *Drosophila melanogaster* and *Drosophila pseudoobscura*. Biochem. Genet. **25:** 41–51.

CHOUDHARY, M., and R. S. SINGH, 1987b A comprehensive study

of genic variation in natural populations of *Drosophila melanogaster*. III. Variations in genetic structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila simulans*. Genetics **117:** 697–710.

COHN, V. H., M. A. THOMPSON and G. P. MOORE, 1984 Nucleotide sequence comparisons of the *Adh* gene in three Drosophilids. J. Mol. Evol. **20:** 31–37.

COTÉ, B., W. BENDER, D. CURTIS and A. CHOVNICK, 1986 Molecular mapping of the *rosy* locus in *Drosophila melanogaster*. Genetics **112:** 755–767.

CRAIG, E. A., T. D. INGOLIA and L. J. MANSEAU, 1983 Expression of *Drosophila* heat-shock cognate genes during heat shock and development. Dev. Biol. **99:** 418–426.

CROSS, S. R. H., and A. J. BIRLEY, 1986 Restriction endonuclease map variation in the *Adh* region in populations of *Drosophila melanogaster*. Biochem. Genet. **24:** 415–433.

DAVID, J. R., and J. TSACAS, 1981 Cosmopolitan, subcosmopolitan and widespread species: different strategies within the *Drosophila* family. C. R. Soc. Biogeog. **57:** 11–26.

DAVIS, P. S., M. W. SHEN and B. H. JUDD, 1987 Assymmetrical pairings of transposons in and proximal to the *white* locus of *Drosophila* account for four classes of regularly occurring exchange products. Proc. Natl. Acad. Sci. USA **84:** 174–178.

DELOTTO, R., and P. SPIERER, 1986 A gene required for the specification of dorsal-ventral pattern in *Drosophila* appears to encode a serine protease. Nature **323:** 688–692.

DOWSETT, A. P., and M. W. YOUNG, 1982 Differing levels of dispersed repetitive DNA among closely related species of *Drosophila*. Proc. Natl. Acad. Sci USA **79:** 4570–4574.

EDWARDS, T. C. R., E. P. M. CANDIDO and A. CHOVNICK, 1977 Xanthine dehydrogenase from *Drosophila melanogaster*: a comparison of the kinetic parameters of the pure enzyme from two wild-type isoalleles differing at a putative regulatory site. Mol. Gen. Genet. **154:** 1–6.

ENGELS, W. R., 1983 The P family of transposable elements in *Drosophila*. Annu. Rev. Gen. **17:** 315–344.

FEINBERG, A. P., and B. VOGELSTEIN, 1984 Addendum to a technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137:** 266–267.

GAUSZ, J., L. M. C. HALL, A. SPIERER and P. SPIERER, 1986 Molecular genetics of the *rosy-Ace* region of *Drosophila melanogaster*. Genetics **112:** 65–78.

GELBART, W. M., M. MCCARRON, J. PANDEY and A. CHOVNICK, 1974 Genetic limits of the xanthine dehydrogenase structural element within the *rosy* locus of *Drosophila melanogaster*. Genetics **78:** 869–886.

GINZBURG, L. R., P. M. BINGHAM and S. YOO, 1984 On the theory of speciation induced by transposable elements. Genetics **107:** 331–341.

GOLDBERG, M. L., J.-Y. SHEEN, W. J. GEHRING and M. M. GREEN, 1983 Unequal crossing over associated with asymmetrical synapsis between nomadic elements in the *Drosophila melanogaster* genome. Proc. Natl. Acad. Sci USA **80:** 5017–5021.

GOLDING, G. B., C. F. AQUADRO and C. H. LANGLEY, 1986 Sequence evolution within populations under multiple types of mutation. Proc. Natl. Acad. Sci USA **83:** 427–431.

GONZALEZ, A. M., V. M. CABRERA, J. M. LARRUGA and A. GULLÓN, 1982 Genetic distance in the sibling species *Drosophila melanogaster*, *Drosophila simulans*, and *Drosophila mauritania*. Evolution **30:** 517–522.

HALE, L. R., and R. SINGH, 1987 Mitochondrial DNA variation and genetic structure in populations of *Drosophila melanogaster*. Mol. Biol. Evol. **4:** 622–637.

HILLIKER, A. J., S. H. CLARK, A. CHOVNICK and W. M. GELBART, 1980 Cytogenetic analysis of the chromosomal region immediately adjacent to the rosy locus in *Drosophila melanogaster*. Genetics **95:** 95–110.

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

HYYTIA, P., P. CAPY, J. R. DAVID and R. S. SINGH, 1985 Enzymes and quantitative variation in European and African populations of *D. simulans*. Heredity **54:** 209–217.

INOUE, Y. H., and M.-T. YAMAMOTO, 1987 Insertional DNA and spontaneous mutation at the white locus in *Drosophila simulans*. Mol. Gen. Genet. **209:** 94–100.

KEITH, T. P., 1983 Frequency distribution of Esterase-5 alleles in two populations of *Drosophila pseudoobscura*. Genetics **105:** 135–155.

KEITH, T. P., M. A. RILEY, M. KREITMAN, R. C. LEWONTIN, D. CURTIS and G. CHAMBERS, 1987 Sequence of the structural gene for xanthine dehydrogenase (*rosy locus*) in *Drosophila melanogaster*. Genetics **116:** 67–73.

KIDD, S., and M. W. YOUNG, 1986 Transposon-dependent mutant phenotypes at the *Notch* locus of *Drosophila*. Nature **323:** 89–91.

KIMURA, M., 1979 Model of effectively neutral mutations in which selective constraint is incorporated. Proc. Natl. Acad. Sci USA **76:** 3440–3444.

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

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

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

LANGLEY, C. H., and C. F. AQUADRO, 1987 Restriction-map variation in natural populations of *Drosophila melanogaster*: *white*-locus region. Mol. Biol. Evol. **4:** 651–663.

LANGLEY, C. H., E. MONTGOMERY and W. F. QUATTLEBAUM, 1982 Restriction map variation in the *Adh* region of *Drosophila*. Proc. Natl. Acad. Sci USA **79:** 5631–5635.

LEE, C. S., D. CURTIS, M. MCCARRON, C. LOVE, M. GRAY, W. BENDER and A. CHOVNICK, 1987 Mutations affecting expression of the *rosy* locus in *Drosophila melanogaster*. Genetics **116:** 55–66.

LEIGH BROWN, A. J., 1983 Variation at the 87A heat shock locus in *Drosophila melanogaster*. Proc. Natl. Acad. Sci USA **80:** 5350–5354.

LEIGH BROWN, A. J., and D. ISH-HOROWICZ, 1981 Evolution of the 87A and 87C heat-shock loci in *Drosophila*. Nature **290:** 677–682.

MARTIN, C. H., and E. M. MEYEROWITZ, 1986 Characterization of the boundaries between adjacent rapidly and slowly evolving genomic regions in *Drosophila*. Proc. Natl. Acad. Sci USA **83:** 8654–8658.

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

O'BRIEN, S. J., and R. J. MACINTYRE, 1969 An analysis of gene-enzyme variability in natural populations of *Drosophila melanogaster* and *D. simulans*. Am. Nat. **103:** 97–112.

OHNISHI, S., A. J. LEIGH BROWN, R. A. VOELKER and C. H. LANGLEY, 1982 Estimation of genetic variability in natural populations of *Drosophila simulans* by two-dimensional and starch gel electrophoresis. Genetics **100:** 127–136.

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

OHTA, T., 1976 Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Popul. Biol. **10:** 254–275.

REED, K. C., and D. A. MANN, 1985 Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13:** 7207–7221.

ROSE, M. R., and W. F. DOOLITTLE, 1983 Molecular biological mechanisms of speciation. Science **220:** 157–162.

RUBIN, G. M. 1983 Dispersed repetitive DNA's in *Drosophila*. pp. 329–361. In: *Mobile Genetic Elements*, Edited by J. A. SHAPIRO. Academic Press, New York.

SCHAEFFER, S. W., C. F. AQUADRO and W. W. ANDERSON, 1987 Restriction map variation in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Mol. Biol. Evol. **4:** 252–263.

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

SHAH, D. M., and C. H. LANGLEY, 1979 Inter- and intraspecific variation in restriction maps of *Drosophila* mitochondrial DNAs. Nature **281:** 695–699.

SINGH, R. S., and L. R. RHOMBERG, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. Genetics **117:** 255–271.

SINGH, R. S., M. CHOUDHARY and J. R. DAVID, 1987 Contrasting patterns of geographic variation in the cosmopolitan sibling species *Drosophila melanogaster* and *Drosophila simulans*. Biochem. Gen. **25:** 27–40.

SINGH, R. S., D. A. HICKEY and J. DAVID, 1982 Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. Genetics **104:** 301–315.

SOLIGNAC, M., and M. MONNEROT, 1986 Race formation, speciation, and introgression within *Drosophila simulans*, *D. mauritiana*, and *D. sechellia* inferred from mitochondrial DNA analysis. Evolution **40:** 531–539.

STEINMETZ, M., M. MALISSEN, L. HOOD, A. ORN, R. A. MAKI, G. R. DASTOORNIKOO, D. STEPHAN, E. GIBB and R. ROMANIUK, 1984 Tracts of high or low sequence divergence in the mouse major histocompatibility complex. EMBO J. **3:** 2995–3003.

STURTEVANT, A. H., 1920 Genetic studies on *D. simulans*. I. Introduction. Hybrid with *D. melanogaster*. Genetics **5:** 488–500.

WATADA, M., Y. N. TOBARI and S. OHBA, 1986 Genetic differentiation in Japanese populations of *Drosophila simulans* and *D. melanogaster*. I. Allozyme polymorphisms. Jpn. J. Genet. **61:** 253–269.

ZWIEBEL, L. J., V. H. COHN, D. R. WRIGHT and G. P. MOORE, 1982 Evolution of single-copy DNA and the *Adh* gene in seven Drosophilids. J. Mol. Evol. **19:** 62–71.

Communicating editor: R. R. HUDSON