Molecular Population Genetics of the Distal Portion of the *X* Chromosome in Drosophila: Evidence for Genetic Hitchhiking of the *yellow-achaete* Region

David J. Begun and Charles F. Aquadro

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

Manuscript received March 11, 1991
Accepted for publication September 3, 1991

ABSTRACT

We have estimated DNA sequence variation and differentiation within and between *Drosophila melanogaster* and its sibling species, *Drosophila simulans*, using six-cutter restriction site variation at *yellow-achaete* (*y-ac*), *phosphogluconate dehydrogenase (Pgd)*, and *period* (*per*). These three gene regions are of varying distance from the telomere of the *X* chromosome and range from very low to moderate rates of recombination in *D. melanogaster*. According to Tajima's test of neutrality, the *Pgd* region has been influenced by balancing selection in *D. melanogaster*. This is consistent with previous data suggesting the allozyme polymorphism at this locus is visible to selection. The Hudson, Kreitman, Aguadé test of neutrality reveals a significant departure from neutrality for the *y-ac* region compared to the *per* or *rosy* regions in *D. simulans*. There is also a significant departure for the *y-ac* region compared to the *Adh* 5′ flanking region in *D. melanogaster*. In both species the departure appears to be due to reduced variation at *y-ac* compared to that expected from divergence between *D. simulans* and *D. melanogaster*. We conclude that recent hitchhiking associated with the selective fixation of one or more advantageous mutants in the *y-ac* region is the best explanation for reduced variation at *y-ac*.

THE challenge of molecular population genetics is to explain how the interaction of genetic and population phenomena dictates levels and distributions of DNA sequence variation in natural populations. For example, selection on a single nucleotide site can greatly affect levels of neutral variation at tightly linked sites. This is commonly referred to as the hitchhiking effect (MAYNARD SMITH and HAIGH 1974; KAPLAN, HUDSON and LANGLEY 1989). Selective fixation of a variant will result in a reduction of linked neutral variation while balancing selection will, over time, result in increased neutral variation in the vicinity of the selected nucleotide. The magnitude of these effects depends on the selection intensity and the recombination rate in the region.

Comparison of genetic and polytene band maps near telomeres and centromeres in *Drosophila melanogaster* indicates that recombination rates in these regions are considerably reduced compared to other regions (ASHBURNER 1989) (Figure 1). Recent molecular walks through regions on the *X* chromosome with fine structure genetics reveal rates of recombination to be on the order of 2×10^{-3} cM/kilobase for the *white* locus in *D. melanogaster* (B. H. JUDD, cited in AGUADÉ, MIYASHITA and LANGLEY 1989). In contrast, BEECH and LEIGH BROWN (1989) estimate, from the genetic data of DUBININ, SOKOLOV and TINIAKOV (1937), a recombination rate of 1.2×10^{-4} cM/kilobase from the *achaete* and *scute* regions located near the telomere of the *X* chromosome in *D. melanogaster*. This seventeen-fold reduction in recombination rate in the *achaete-scute* region supports the interpretation that the flattening of the curve near the telomere of the *X* chromosome in Figure 1 results from reduced recombination rates extending from the *per* region through the *y-ac* region. Though there are few genetic data for *D. simulans*, the close evolutionary relationship to *D. melanogaster*, the homosequential *X* chromosome polytene banding pattern, and the similar order of known genes (Figure 2) suggests that *D. simulans* may also have reduced recombination in the telomere region. A predicted consequence of reduced recombination is an elevated level of linkage disequilibrium, as has been seen for the *achaete-scute* region in *D. melanogaster* (MACPHERSON, WEIR and LEIGH BROWN 1990). That prediction has also been borne out in our data for *D. simulans*, as detailed below.

Several investigators have reported reduced levels of intraspecific variation near the telomere and centromere of the *X* chromosome in Drosophila (AGUADÉ, MIYASHITA and LANGLEY 1989; STEPHAN and LANGLEY 1989; LANGLEY 1990; MIYASHITA 1990). One explanation for this pattern is that selective fixation and associated genetic hitchhiking reduce levels of neutral variation in these regions of low recombination. The data at hand for *D. melanogaster*, however, are inconclusive. First, the generally low level of variation present throughout the *D. melanogaster* genome makes it difficult to statistically demonstrate that variation is significantly lower near the telomere. For

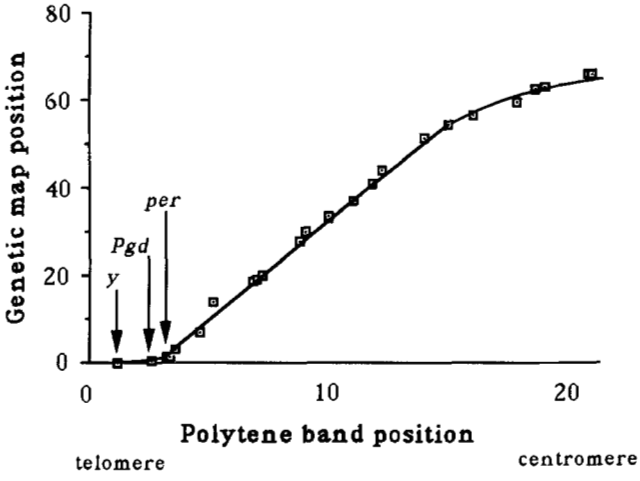

FIGURE 1.—Plot of cytological *vs.* genetic location of several *X*-linked genes in *D. melanogaster*. Data are from LINDSLEY and ZIMM (1985, 1990).

example, studies of the *yellow-achaete-scute* region have revealed less polymorphism in some populations than had been observed in other gene regions (AGUADÉ, MIYASHITA and LANGLEY 1989; BEECH and LEIGH BROWN 1989; EANES, LABATE and AJIOKA 1989), but only in the largest study was the lower level of variation statistically significant (AGUADÉ, MIYASHITA and LANGLEY 1989). However, this study did not take into account the fact that *X*-linked gene regions are expected to have lower levels of neutral DNA variation than autosomal gene regions because the effective population size of the *X* chromosome is smaller than that of the autosomes. Second, even if one unambiguously demonstrates that a gene region has reduced polymorphism, an equally good explanation for reduced variation can not be ruled out from polymorphism data alone, namely, that it results from a reduced mutation rate and/or increased functional constraint (AGUADÉ, MIYASHITA and LANGLEY 1989; LANGLEY 1990). However, the neutral theory of molecular evolution predicts that intraspecific variation is positively correlated with interspecific divergence (KIMURA 1983). Therefore, if reduced polymorphism in a gene region results from a low neutral mutation rate, then one expects to find reduced interspecific divergence compared to other gene regions. Alternatively, a selective sweep will reduce levels of linked neutral variation but will not affect divergence (BIRKY and WALSH 1988; KAPLAN, HUDSON and LANGLEY 1989).

A potentially powerful approach to determine if there is reduced variation near telomeres and/or centromeres in Drosophila and if so, whether this reduction results from hitchhiking, is to determine levels of polymorphism and divergence over several gene regions in the sibling species, *D. melanogaster* and *D. simulans*. This strategy has two advantages. *Drosophila simulans* appears to have several fold greater levels of polymorphism than *D. melanogaster* (AQUADRO 1990). Since selective sweeps reduce levels of linked, neutral variation, these sweeps may be easier to detect in *D. simulans*. Furthermore, data from several gene regions from a single population provide a larger context for asking questions about the relationship between recombination rates and levels of DNA sequence variation.

We have tested the hitchhiking effect hypothesis near the telomere of the *X* chromosome by measuring intra- and interspecific DNA sequence variation in three gene regions of varying distance from the telomere in *D. melanogaster* and *D. simulans*. *Yellow-achaete* (*y-ac*), *phosphogluconate dehydrogenase* (*Pgd*), and *period* (*per*) are located at polytene chromosome bands 1B, 2D, and 3B, and genetic map positions 0.0, 0.6, and 1.4, respectively, in *D. melanogaster* (LINDSLEY and ZIMM 1990). Since the polytene bands and available genetic markers for the *X* chromosomes of *D. melanogaster* and *D. simulans* are homosequential, we assume the physical locations of these genes are approximately the same in *D. simulans*. Our results demonstrate a significant reduction in DNA variation at *y-ac* in both *D. simulans* and *D. melanogaster* despite a typical amount of sequence divergence between the species. These data rule out reduced mutation rates and/or increased functional constraint as the explanation for reduced variation and support the hypothesis that the observed patterns result from one or more recent hitchhiking events in the region in each species.

MATERIALS AND METHODS

Samples: *D. melanogaster* and *D. simulans* isofemale lines were established from collections made during Fall 1989 by M. TURELLI and S. BRYANT in Arvin, California, and at the Zzyzx Desert Research Station near Soda Lake, California, respectively. A total of 35 *D. melanogaster* (18 from Arvin, 17 from Soda Lake) and 36 *D. simulans* (19 from Arvin, 17 from Soda Lake) *X* chromosomes were isolated by means of attached-*X* stocks provided by R. MACINTYRE (*D. melanogaster*) and J. COYNE (*D. simulans*).

Restriction map analysis: Genomic DNA was prepared from males of each attached-*X* line by a modification of the mini-prep protocol of BENDER, SPIERER and HOGNESS (1983). Six restriction endonucleases were used: *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Sal*I and *Xho*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 buffer of 0.8 mM Tris, 0.4 mM acetic acid and 0.04 mM EDTA, pH 8.0. Mapping of sites was accomplished by double digests. DNA fragments were transferred to Zetabind nylon membranes (WESTNEAT *et al.* 1988). Phage or plasmids containing *D. melanogaster* genomic DNA were radiolabeled with [^{32}P]dCTP by random priming (FEINBERG and VOGELSTEIN 1983), separated from unincorporated nucleotides on Sephadex G-50 spin columns (MANIATIS, FRITSCH and SAMBROOK 1982), and hybridized to membranes at 60° in 7% sodium dodecyl sulfate, 525 mM $NaPO_4$, 1 mM EDTA, 1% bovine serum albumin (w/v). Membranes were washed twice for 5–10 min at room temperature, and twice for 30 min at 55° in 1% sodium dodecyl sulfate, 40 mM $NaPO_4$, 1 mM EDTA.

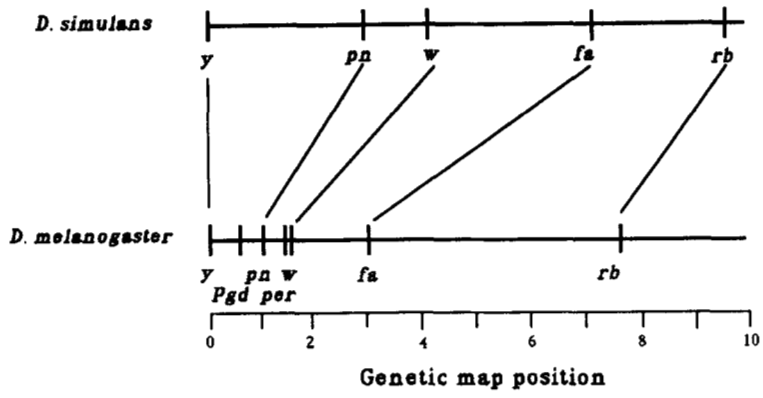

FIGURE 2.—Genetic maps of the telomeric region in *D. melanogaster* and *D. simulans*. *D. melanogaster* data are from LINDSLEY and ZIMM (1985, 1990); *D. simulans* data are from STURTEVANT *et al.* (1929). Genetic map positions for *Pgd* and *per* are only available in *D. melanogaster*, but are indicated for reference.

Starch gel electrophoresis: Homogenates of males from each attached-*X* line were electrophoresed through 12% (w/v) starch gels. Electrode and gel buffers were 40 mM citrate and 4 mM citrate (pH adjusted to 6.1 with *N*-(3-amino-propyl)morpholine), respectively (MAY 1991). Gel slices were stained for 6-phosphogluconate dehydrogenase according to standard procedures (HARRIS and HOPKINSON 1976).

Statistical analyses: Variant frequencies at polymorphic sites did not differ between populations for either species (Fisher's exact test). Therefore, all analyses were done on pooled data. The presence of nonrandom association among DNA variants was assessed by the calculation of D', the ratio of the linkage disequilibrium coefficient to its theoretical maximum (LEWONTIN 1964). Significance of association was determined by Fisher's exact test.

Nucleotide variability was estimated in two ways. One approach was to estimate the number of nucleotide differences (π) which is equivalent to the heterozygosity per nucleotide. Estimates of π and its standard error were obtained by the method of NEI and LI (1979). A second approach was to estimate θ (and its standard error) from the number of polymorphic sites following HUDSON (1982). The parameter θ is defined as $4N\mu$, where N is the effective population size and μ is the neutral mutation rate. For *X*-linked genes, the HUDSON (1982) method estimates $3N\mu$ rather than $4N\mu$ due to differences in effective population size for *X*-linked *vs.* autosomal genes (assuming equal effective population sizes for males and females). The value of $4N\mu$ estimated for an autosomal region should be equal to π under the assumptions of equilibrium and selective neutrality. For *X*-linked genes, the expectation of π is $3N\mu$ instead of $4N\mu$. HUDSON's estimator was also calculated by a modification of HUDSON (1982) in which the proportion of polymorphic sites (p) was divided by $\sum_{i=1}^{n-1} 1/i$ instead of $\ln(n)$; this provides a better estimate of $4N\mu$ (or $3N\mu$ in the case of *X*-linked gene regions) because $\ln(n)$ is an approximation of $\sum_{i=1}^{n-1} 1/i$ (WATTERSON 1975; R. HUDSON personal communication). Standard errors should be interpreted cautiously since they are calculated with neutral, equilibrium models, the assumptions of which may be violated by our data (see DISCUSSION). Sequence divergence and the associated standard errors were estimated according to NEI and JIN (1989) with the neighbor-joining method. Only sites in probed regions were used to estimate nucleotide polymorphism and divergence.

We used restriction site data to solve the appropriate simultaneous equations for the HUDSON, KREITMAN and AGUADÉ (1987) test. Following HUDSON (1982) the effective number of nucleotides surveyed is $(2m-k)j$ where j is the number of base pairs in the recognition site of the restriction enzyme, m is the total number of sites observed, and k is the number of polymorphic sites observed. The number of polymorphic nucleotides is considered equivalent to the number of polymorphic restriction sites. For the purposes of calculating divergence for the HKA test we estimated the total effective number of nucleotides observed between species as the mean of the effective number of nucleotides observed within species. The average pairwise number of restriction site differences was used as an estimate of D in HKA tests. The tests were done separately on *D. melanogaster* (ignoring intraspecific polymorphism in *D. simulans*) and on *D. simulans* (ignoring intraspecific polymorphism in *D. melanogaster*) in order to distinguish evolutionary processes occurring in individual species.

When testing two *X*-linked genes the HKA test can be applied as for two autosomal genes. However, when comparing autosomal and *X*-linked genes, the following equations, modified from HUDSON, KREITMAN and AGUADÉ (1987), must be used (R. HUDSON, personal communication). Let locus 1 be autosomal with neutral mutation rate μ_1 and locus 2 be *X*-linked with neutral mutation rate μ_2. Also let $\theta_1 = 4N\mu_1$ and $\theta_2 = 4N\mu_2$. Then,

$$E(S_2) = C(n)\theta_2 \tfrac{3}{4}$$

$$E(D_2) = \theta_2(T + \tfrac{3}{4}).$$

The modified versions of Equations 5 or 6 of HUDSON, KREITMAN and AGUADÉ (1987) are:

$$S_1 + S_2 = C(n)(\theta_1 + \theta_2 \tfrac{3}{4})$$

$$D_1 + D_2 = \theta_1(T + 1) + \theta_2(T + \tfrac{3}{4})$$

$$D_1 + S_1 = C(n)\theta_1 + \theta_1(T + 1).$$

The variances are:

$$\mathrm{Var}(S_2) = E(S_1) + \left[\left(\theta_2 \frac{3}{4}\right)^2 \sum_{j=1}^{n-1} \frac{1}{j^2}\right]$$

$$\mathrm{Var}(D_2) = E(D_2) + \left(\theta_2 \frac{3}{4}\right)^2.$$

The expectations and variances of S_1 and D_1 are as given in HUDSON, KREITMAN and AGUADÉ (1987). Finding the solution to the above system of three equations requires solving a quadratic. A computer program to solve the equations and estimate the chi-square statistic is available from the authors or R. R. HUDSON. TAJIMA's (1989a) test of neutrality was carried out on restriction site data as described.

RESULTS

Results from each of the three gene regions are presented separately below. Summary restriction

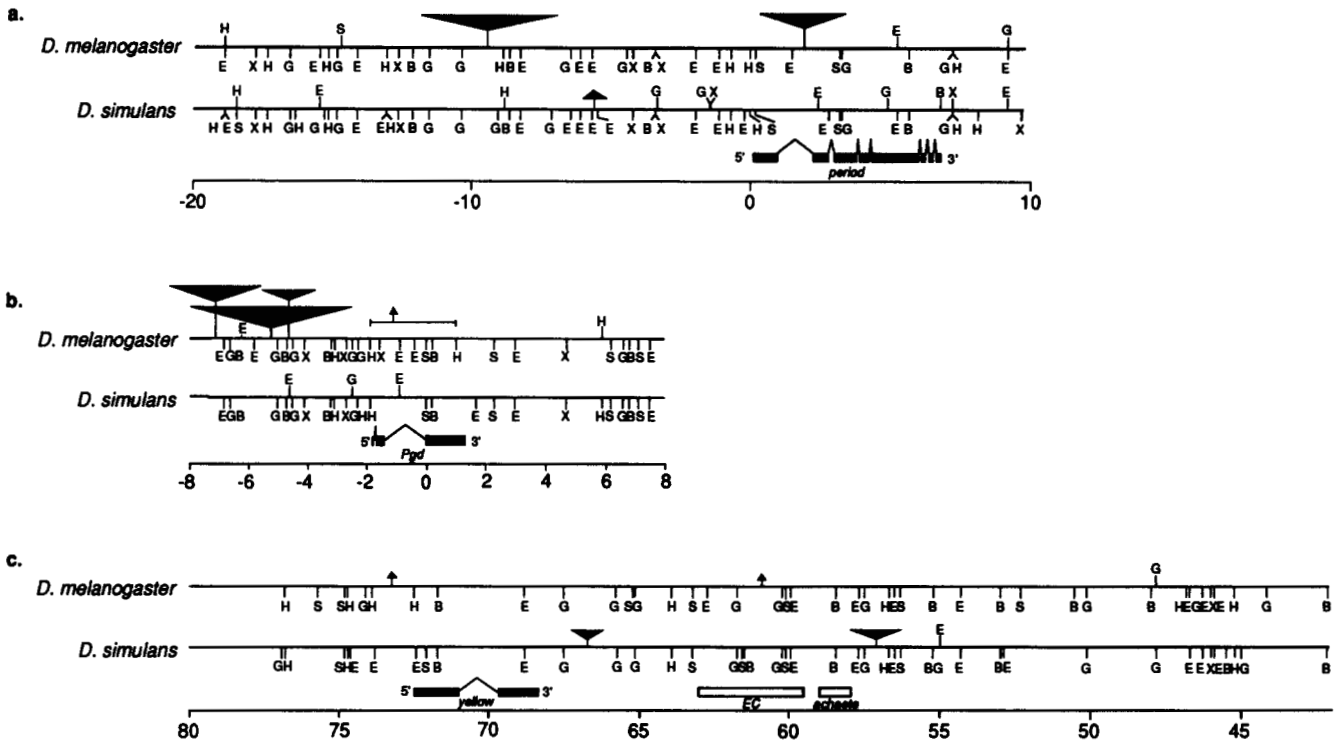

FIGURE 3.—Summary restriction site maps of the *per* (a), *Pgd* (b), and *y-ac* (c) gene regions in *D. melanogaster* and *D. simulans*. The scale in kilobases is given below the map of each region. B = *Bam*HI, G = *Bgl*II, E = *Eco* RI, H = *Hin*dIII, S = *Sal*I, and X = *Xho*I. Polymorphic restriction sites are shown above the maps, fixed sites below the maps. Insertions and deletions are indicated by triangles pointing toward or away from the map, respectively, and are shown approximately to scale. Locations of insertions/deletions are known only to the fragment shown. Probes used in the *y-ac* region were λsc112 and λsc101 (CAMPUZANO *et al.* 1985) provided by G. KARPEN and y 8.5 provided by P. GEYER. Coordinates in the *y-ac* region follow CAMPUZANO *et al.* (1985). The *Pgd* region was probed with λACP12A (GUTIERREZ *et al.* 1989) provided by M. SCOTT and J. LUCCHESI. The *per* region was probed with clones ZW106 and ZW108 (BARGIELLO and YOUNG 1984) provided by M. YOUNG. Transcript organization of gene regions in *D. melanogaster* is indicated below the maps. Exons are indicated by black boxes.

maps for *per*, *Pgd* and *y-ac* are shown in Figure 3. Frequencies of variants at polymorphic sites in *D. melanogaster* and *D. simulans* are shown in Tables 1 and 2, respectively. Estimates of heterozygosity per nucleotide and interspecific divergence for each region are in Table 3. Linkage disequilibrium estimates for *D. melanogaster* and *D. simulans* are shown in Tables 4 and 5, respectively. Results of the TAJIMA tests of neutrality are shown in Table 6. Table 7 shows the data used to perform HKA tests, while Tables 8 and 9 show the resulting test statistics in *D. simulans* and *D. melanogaster*, respectively.

***yellow-achaete* gene region:** In *D melanogasters*, no large insertions were found in the region. Two different deletion variants less than 100 bp were found (Figure 3, Table 1). Estimates of $3N\mu$ (± SE) and π (± SE) are 0.0005 ± 0.0005 and 0.0008 ± 0.0008, respectively. D' was significantly different from zero for the one possible pairwise comparison.

In *D. simulans*, two unique insertions of approximately 1.0 and 1.6 kb were found. Inserts of this size in *D. melanogaster* are thought to be transposable elements (AQUADRO *et al.* 1986; BEECH and LEIGH BROWN 1989). Estimates of $3N\mu$ and π are 0.0005 ± 0.0005 and 0.0001 ± 0.0003. Linkage disequilibrium could not be assessed because all variants were unique. Interspecific divergence over this region was estimated as 0.054 ± 0.013.

***phosphogluconate dehydrogenase* gene region:** Three unique insertions, probably transposable elements, of approximately 1.7, 3.0 and 5.3 kb were found in the 5′ flanking region of *D. melanogaster*. A small deletion occurring at a frequency of 0.257 (Table 1) was found in the transcribed region. A four-cutter restriction site survey of this region indicates that this variant is within the large intron (D. BEGUN and C. AQUADRO, unpublished data). Levels of per nucleotide heterozygosity estimated by $3N\mu$ and π are 0.0014 ± 0.0011 and 0.0030 ± 0.0022, respectively.

Three different protein electrophoretic variants were found (Table 1) with an expected protein heterozygosity of 0.438. The Fast and Slow electromorphs are common throughout North America and occurred at frequencies typical of previously surveyed western U.S. populations (SINGH and RHOMBERG 1987). Two lines had electrophoretic mobility slightly greater than the Fast electromorph (referred to as very fast or VF in Table 1). This electromorph, found once in each population, has been found at low frequency throughout North America (W. F. EANES, personal communication).

There was extensive nonrandom association across

TABLE 1

Genetic variation in the *per*, *Pgd*, and *y-ac* gene regions in *D. melanogaster*

	per							*Pgd*						*y-ac*		
Line	*Hin*dIII −18.8	*Sal*I −14.7	Ins −9.8	Ins 2.0	*Eco*RI 5.2	*Bgl*II 9.2	Ins −7.0	*Eco*RI −6.2	Ins −5.0	Ins −4.5	Del	Allo	*Hin*dIII 5.9	Del 64	Del 54	*Bgl*II 48.0
A1	+	+	−	−	−	−	−	−	−	−	−	F	−	−	−	−
A2	−	−	−	−	−	−	−	−	−	−	−	F	−	−	−	−
A3	−	+	−	−	−	−	−	+	−	−	−	F	+	−	−	−
A4	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	−
A6	−	+	−	−	−	−	−	+	−	−	−	F	+	+	−	+
A7	−	+	−	−	−	−	−	+	−	−	+	F	+	−	−	−
A8	−	+	−	−	+	+	−	+	−	−	−	F	+	−	−	−
A9	−	−	−	−	−	−	−	−	−	−	−	S	−	+	−	+
A11	−	+	−	−	−	−	−	−	+	−	−	F	−	−	−	−
A12	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	−
A13	−	+	−	−	−	−	−	+	−	+	−	VF	+	−	+	−
A15	−	−	−	−	−	−	−	+	−	−	−	F	+	−	−	−
A16	−	−	−	−	−	−	−	−	−	−	−	F	−	−	−	−
A17	−	+	−	−	−	−	−	−	−	−	−	F	−	+	−	+
A18	−	+	−	−	−	−	−	+	−	−	+	S	+	−	−	−
A19	−	+	−	−	−	−	−	+	−	−	+	S	+	−	−	−
A20	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	+
A21	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	+
SL2	−	+	−	−	−	−	−	+	−	−	−	F	+	−	−	−
SL3	−	+	−	−	−	−	−	+	−	−	+	S	+	−	−	−
SL5	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	−
SL6	−	+	−	−	−	−	−	+	−	−	+	F	+	−	−	−
SL7	−	−	−	−	−	+	−	−	−	−	−	F	−	−	−	−
SL8	−	+	−	−	−	−	−	+	−	−	+	S	+	−	−	−
SL9	−	+	−	+	−	−	+	−	−	−	−	F	−	−	−	+
SL10	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	−
SL11	−	+	−	−	−	−	−	+	−	−	+	S	+	−	−	−
SL12	−	+	−	−	−	−	−	−	−	−	−	F	+	−	−	+
SL13	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	+
SL14	−	+	+	−	−	−	−	−	−	−	−	F	−	−	−	+
SL15	−	+	−	−	+	−	−	−	−	−	−	VF	−	−	−	−
SL16	−	+	−	−	+	+	−	+	−	−	+	S	+	−	−	−
SL17	−	+	−	−	−	−	−	+	−	−	+	S	+	−	−	−
SL18	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	−
SL19	−	+	−	−	−	−	−	−	−	−	−	F	−	−	−	−
Freq.	0.03	0.11	0.03	0.03	0.09	0.09	0.03	0.43	0.03	0.03	0.26		0.46	0.09	0.03	0.26

Lines collected in Arvin, California, are indicated by "A"; lines collected near Soda Lake, California, are indicated by "SL." Restriction site coordinates are from Figure 3. Presence of a variant is indicated by "+" and absence by "−." Site frequency is the frequency of the less common variant. Positions of insertions (Ins) and deletions (Del) are approximate. The Fast (F), Slow (S), and Very Fast (VF) electromorphs occurred at frequencies 0.710, 0.233 and 0.057, respectively.

the region; all pairwise comparisons resulted in a significant D' (Table 4).

In *D. simulans* no sequence length variation was detected. Four polymorphic restriction sites were found (Figure 3, Table 2), one of which (*Hin*dIII +12.3) is located outside the probed region and not shown in Figure 3. Estimates of $3N\mu$ and π are 0.0023 ± 0.0015 and 0.0011 ± 0.0012, respectively. Consistent with previous studies (*e.g.*, CHOUDHARY and SINGH 1987), no allozyme variation was found.

All three pairwise comparisons among non-unique variants showed significant linkage disequilibrium. However, the low frequency of variants makes it difficult to generalize as to the strength of linkage disequilibrium in the region.

Sequence divergence between species over the *Pgd* region was estimated as 0.029 ± 0.011.

***period* gene region:** Two unique insertions of approximately 4.7 and 3.0 kb were found in *D. melanogaster*; the 4.7-kb insertion was located in 5′ flanking sequence while the 3.0-kb insert was located in the transcribed region, probably in the first intron. The size and restriction map of the 3.0-kb insert (data not shown) suggest it is a *hobo* element (ASHBURNER 1989). Estimates of $3N\mu$ and π over the 30-kb region were 0.0022 ± 0.0016 and 0.0014 ± 0.0012, respectively. One of three pairwise comparisons resulted in a significant D' (Table 4).

Extensive size variation was observed within the 200 bp fragment defined by *Eco*RI −5.7 and *Eco*RI −5.5 in *D. simulans*. Variants are designated as sequence length increases relative to an arbitrary baseline (see Table 2 legend). However, we can not infer the number of events required for transitions from

TABLE 2

Genetic variation in the *per*, *Pgd*, and *y-ac* gene regions in *D. simulans*

	per												*Pgd*				*y-ac*		
Line	*Hin*dIII −18.4	*Eco*RI −15.4	*Hin*dIII −8.8	Ins −5.6	*Bgl*II −3.3	*Bgl*II −1.4	*Xho*I −1.4	*Eco*RI 2.8	*Bgl*II 4.9	*Bam*HI 6.8	*Xho*I 7.2	*Eco*RI 9.2	*Eco*RI −4.6	*Bgl*II −2.5	*Eco*RI −0.9	*Hin*dIII 12.3	Ins 67	Ins 57	*Eco*RI 55.0
A1	−	+	−	0.4	+	−	−	−	−	−	+	−	+	−	+	+	−	−	+
A3	−	+	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
A4	−	−	−	−	+	−	−	−	−	−	+	−	+	−	+	+	−	−	+
A5	−	+	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
A6	+	+	−	−	+	−	−	−	+	+	−	+	+	−	+	+	−	−	+
A8	−	+	−	0.8	+	+	−	−	+	−	−	+	+	−	+	+	−	−	+
A12	−	−	−	0.8	+	+	−	−	+	−	−	+	+	−	+	+	−	−	+
A17	−	−	−	1.6	−	−	+	+	+	+	−	+	+	−	+	+	−	−	+
A21	+	+	−	−	+	−	−	−	−	−	+	−	+	−	+	+	−	−	+
A22	−	+	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
A23	+	+	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
A24	−	+	−	0.4	+	−	−	−	−	−	+	−	+	−	+	+	−	−	+
A25	−	+	−	−	+	−	−	−	−	−	+	−	+	−	+	+	−	−	−
A28	−	−	−	0.8	+	+	−	−	+	−	−	+	+	+	−	−	−	−	+
A29	−	−	−	1.8	−	−	+	−	−	−	+	−	+	+	−	−	−	−	+
A31	−	−	−	0.8	+	+	−	−	+	−	−	+	+	−	+	+	−	−	+
A33	−	+	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
A37	−	−	−	1.6	−	−	+	+	+	+	−	+	+	+	+	+	−	−	+
A43	−	+	−	−	+	−	−	−	−	−	+	−	+	−	+	+	−	−	+
SL51	−	−	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
SL52	−	+	−	0.6	−	−	−	+	+	+	−	+	+	−	+	+	−	+	+
SL54	−	−	−	−	−	−	−	+	+	+	−	+	+	−	+	+	−	−	+
SL56	−	+	−	0.4	+	+	−	−	+	−	−	−	+	−	+	+	−	−	+
SL58	−	+	+	NS	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
SL61	−	−	−	0.8	+	+	−	−	+	−	−	+	+	−	+	+	−	−	+
SL64	+	+	−	−	+	−	−	−	−	−	+	−	+	−	+	+	+	−	+
SL65	−	+	−	0.4	+	−	−	−	+	−	−	+	+	−	+	+	−	−	+
SL67	+	+	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
SL68	−	−	−	0.8	+	+	−	−	−	−	−	+	+	−	+	+	−	−	+
SL71	−	+	−	0.4	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
SL75	+	+	−	0.2	+	−	−	−	−	−	+	−	+	−	+	+	−	−	+
SL79	+	+	−	−	−	−	−	−	−	−	+	−	+	−	+	+	−	−	+
SL86	−	−	−	1.7	−	−	+	+	+	+	−	+	−	−	+	+	−	−	+
SL87	−	−	−	1.7	+	−	+	+	+	+	−	+	+	+	+	−	−	−	+
SL88	−	−	−	0.8	+	+	−	−	+	+	−	+	+	−	+	+	−	−	+
SL100	−	+	−	0.8	+	−	−	−	+	+	−	−	+	−	+	+	−	−	+
Freq.	0.19	0.39	0.03		0.19	0.22	0.14	0.17	0.31	0.50	0.28	0.42	0.03	0.11	0.06	0.08	0.03	0.03	0.03

Lines collected in Arvin, California, are indicated by "A"; lines collected near Soda Lake, California, are indicated by "SL." Restriction site coordinates are from Figure 3, except for *Hin*dIII(12.3), which is outside the probed region and is not shown in Figure 3. Presence of a variant is indicated by "+" and absence by "−." Site frequency is the frequency of the less common variant. Positions of insertions (Ins) are approximate. Ins −5.6 in the *per* region was scored as an increase in kilobases relative to an arbitrary baseline variant indicated by "−"; NS = not scored.

one size class to another or determine if a given change was a sequence length increase or decrease. All variants at this site except the +1.7-kb insert appear to be multiples of 200 bp. Therefore, we hypothesize that the size variation results from tandem repeats of an approximately 200-bp unit. Estimates of $3N\mu$ and π over the 30-kb region were 0.0046 ± 0.0020 and 0.0070 ± 0.0038, respectively. Seventeen of 45 pairwise comparisons (38%) resulted in estimates of linkage disequilibrium (D') significantly different from zero (Table 5), while only 5%, or 2 comparisons, are expected to be significant by chance.

Average divergence between *D. melanogaster* and *D. simulans* over the region was estimated as 0.050 ± 0.010.

Statistical tests of neutrality: TAJIMA's (1989a) test of neutrality is based on the observation that one can estimate the neutral parameter $4N\mu$ ($3N\mu$ for *X*-linked genes), where N and μ are the effective population size and neutral mutation rate, respectively, by the average pairwise differences in a sample or from the number of segregating sites. The test estimates the probability of deviation from the neutral expectation that the two quantities are equal. In effect, TAJIMA's test determines if the observed variant frequency distribution agrees with the neutral expectation. TAJIMA's test was applied to restriction site data at each gene region in both species. In only one case, *Pgd* in *D. melanogaster*, was the neutral model rejected (Table 6). This result is unlikely to be a consequence of nonequilibrium conditions associated with a recent population expansion in this species (DAVID and CAPY

TABLE 3

Nucleotide heterozygosity and divergence (±SE) in the *per*, *Pgd*, and *y-ac* gene regions in *D. melanogaster* and *D. simulans*

Region	Species	$3N\mu$[a]	π	Divergence[b]
y-ac	*D. melanogaster*	0.0005 (±0.0005) 0.0005 (±0.0005)	0.0008 (±0.0008)	0.054 (±0.013)
	D. simulans	0.0005 (±0.0005) 0.0005 (±0.0005)	0.0001 (±0.0003)	
Pgd	*D. melanogaster*	0.0014 (±0.0011) 0.0016 (±0.0013)	0.0030 (±0.0022)	0.029 (±0.011)
	D. simulans	0.0023 (±0.0015) 0.0026 (±0.0018)	0.0011 (±0.0012)	
per	*D. melanogaster*	0.0022 (±0.0016) 0.0026 (±0.0016)	0.0014 (±0.0012)	0.050 (±0.010)
	D. simulans	0.0046 (±0.0020) 0.0052 (±0.0024)	0.0070 (±0.0038)	

[a] Upper estimate calculated according to HUDSON (1982) except ln(n) replaced by $\sum_{i=1}^{n-1} 1/i$ where n is the number of sequences sampled. Lower estimate calculated according to HUDSON (1982); presented for comparison with previously published estimates of other gene regions. Standard errors of $3N\mu$ assume no recombination (HUDSON, 1982).

[b] Average pairwise divergence between *D. melanogaster* and *D. simulans*, uncorrected for within species variation.

TABLE 4

Linkage disequilibrium (D') in the *per*, *Pgd*, and *y-ac* gene regions in *D. melanogaster*

	per			*Pgd*				*y-ac*	
	Sal I −14.7	*Eco*RI 5.2	*Bgl*II 9.2	*Eco*RI −6.2	Del	Allo	*Hin*dIII 5.9	Del 64	*Bgl*II 48.0
per									
*Sal*I −14.7	—								
*Eco*RI 5.2	1.00	—							
*Bgl*II 9.2	−0.22	0.64*	—						
Pgd									
*Eco*RI −6.2	0.53	0.42	0.42	—					
Del	1.00	0.10	0.10	1.00**	—				
Allo	−0.01	0.34	0.12	0.78**	0.83**	—			
*Hin*dIII 5.9	0.56	0.39	0.39	1.00**	1.00**	0.77*	—		
y-ac									
Del 64	0.22	−1.00	−1.00	−0.74*	−1.00	−0.54	−0.51	—	
*Bgl*II 48.0	−0.22	−1.00	−1.00	−0.22	−1.00	0.12	−0.27	1.00**	—

Restriction map coordinates are from Figure 3. * $P < 0.05$; ** $P < 0.01$.

1989); such a phenomenon would have influenced other gene regions in a similar manner, and would result in an excess of rare variants (MARUYAMA and FUERST 1984) and a negative test statistic (TAJIMA 1989b). The positive test statistic indicates that variants occur at higher than expected frequencies, as would be the case under balancing selection. Population subdivision with recent admixture could also lead to this pattern, but again, would be expected to affect all gene regions, not *Pgd* alone.

Under a strictly neutral, infinite sites model of molecular evolution, intraspecific polymorphism and interspecific divergence are both functions of the neutral mutation rate in the region (KIMURA 1983). The neutral expectation that all gene regions have the equivalent ratio of polymorphism to divergence is the basis for the HKA test of neutrality (HUDSON, KREITMAN and AGUADÉ 1987).

Comparisons among the three *X*-linked gene regions (*y-ac*, *Pgd*, and *per*) in *D. simulans* using the HKA test revealed a significant departure from neutrality for *y-ac vs. per* (Table 8). Comparison of these three regions individually to the autosomal *rosy* region (data from AQUADRO, LADO and NOON 1988) also revealed a significant departure from neutrality for the *y-ac vs. rosy* comparison.

Based on their observation of unusually low levels of restriction site variation in the *y-ac-sc* region for several populations of *D. melanogaster*, AGUADÉ, MIYASHITA and LANGLEY (1989) hypothesized that the region may be influenced by the hitchhiking effect (KAPLAN, HUDSON and LANGLEY 1989). Our divergence data from the *y-ac* region between *D. melanogaster* and *D. simulans* allowed us to use the HKA model to test the hypothesis that levels of variation in this region in *D. melanogaster* are compatible with neutrality.

Comparison of our *y-ac* data to that from *Pgd*, *per*

TABLE 5

Linkage disequilibrium (D') in the *per* and *Pgd* gene regions in *D. simulans*

	per										*Pgd*		
	*Hin*dIII −18.4	*Eco*RI −15.4	*Bgl*II −3.3	*Bgl*II −1.4	*Xho*I −1.4	*Eco*RI 2.8	*Bgl*II 4.9	*Bam*HI 6.8	*Xho*I 7.2	*Eco*RI 9.2	*Bgl*II −2.5	*Eco*RI −0.9	*Hin*dIII 12.3
per													
*Hin*dIII −18.4	—												
*Eco*RI −15.4	1.00**	—											
*Bgl*II −3.3	0.27	0.53	—										
*Bgl*II −1.4	−1.00	0.59*	1.00	—									
*Xho*I −1.4	−1.00	−1.00**	−0.75**	−1.00	—								
*Eco*RI 2.8	−1.00	−0.73*	−0.79**	−1.00	0.76**	—							
*Bgl*II 4.9	−0.38	−0.30	−0.06	0.59	0.35	1.00	—						
*Bam*HI 6.8	−0.14	0.00	−0.43	−0.75*	0.60	1.00**	1.00**	—					
*Xho*I 7.2	0.41	0.49	−0.01	−1.00	−0.28	−1.00	−1.00**	−1.00**	—				
*Eco*RI 9.2	−0.66	−0.63**	−0.51	0.79**	0.66	1.00**	0.78*	0.07	−1.00**	—			
Pgd													
*Bgl*II −2.5	−1.00	−1.00*	−0.38	0.04	0.71*	0.40	0.18	0.00	−0.10	0.57	—		
*Eco*RI −0.9	1.00	1.00	0.38	−0.36	−0.42	1.00	0.28	1.00	−0.31	−0.14	−1.00*	—	
*Hin*dIII 12.3	1.00	1.00	0.17	−0.14	−0.61*	−0.20	0.04	0.33	−0.08	−0.43	−1.00**	1.00**	—

Restriction map coordinates are from Figure 3. * $P < 0.05$; ** $P < 0.01$. There was insufficient variation in the *y-ac* region to examine linkage disequilibrium in *D. simulans*.

TABLE 6

Tajima test of neutrality in the *Pgd* and *per* gene regions in *D. melanogaster* and *D. simulans*

Pgd	*D. melanogaster*	2.135*
	D. simulans	−1.119
per	*D. melanogaster*	−0.974
	D. simulans	0.993

* $P < 0.05$. TAJIMA's test statistic, D, was calculated using restriction site variation. There was insufficient variation for the test to be applied to the *y-ac* region in either species.

(this paper), *rosy* (AQUADRO, LADO and NOON 1988) and the 5′ flanking region of *Adh* (KREITMAN and HUDSON 1991) revealed a significant departure from neutrality only for the 5′ *Adh* comparison (Table 9). We also tested neutrality in the *y-ac* region with polymorphism data from AGUADÉ, MIYASHITA and LANGLEY (1989). For this test we used our estimated pairwise difference between *D. melanogaster* and *D. simulans*. We then determined the total number of sites scored and the number of polymorphic sites scored for each of the three populations sampled by AGUADÉ, MIYASHITA and LANGLEY (1989) over the same genomic region for which we had estimated divergence. Populations were analyzed separately because there was evidence for between-population heterogeneity (AGUADÉ, MIYASHITA and LANGLEY 1989). As was true for our *y-ac* data from California, comparisons of *y-ac* for Texas and Japan yielded significant departures from neutrality (Table 9). However, the *y-ac* data from North Carolina did not show a significant departure from neutrality. Comparison of the 5′ flanking *Adh* region to *per*, *Pgd*, and *rosy* also did not reveal significant departures from neutrality (Table 9).

DISCUSSION

Transposable element frequencies and recombination: The observation that large insertions (presumed to be transposable elements) are usually found at low individual frequency in *D. melanogaster* and *D. simulans* (AQUADRO 1990; LANGLEY 1990) is taken as evidence that such variation is deleterious (see also GOLDING, AQUADRO and LANGLEY 1986). Furthermore, the observation that the incidence of large insertion variation does not vary significantly between autosomes and the *X* chromosome suggests that the transposable elements themselves are usually not deleterious (unless they insert into coding DNA; MONTGOMERY, CHARLESWORTH and LANGLEY 1987). The alternative hypothesis that such inserts are deleterious because homologous inserts located in different regions increase the rate of nonhomologous recombination events (LANGLEY *et al.* 1988) predicts that the inserts should accumulate in regions of reduced crossing over.

Our restriction site data from *D. melanogaster* do not support this prediction. We found no increase in the frequency of inserts in *y-ac*. In fact, *y-ac* was the only one of the three gene regions surveyed in which we found no large inserts. This result is consistent with three previous studies of *y-ac* in *D. melanogaster* which revealed no increase in frequency of large insertions (AGUADÉ, MIYASHITA and LANGLEY 1989; BEECH and LEIGH BROWN 1989; EANES, LABATE and AJIOKA 1989).

On the other hand, our data from *D. simulans* are consistent with the unequal cross over model. The density of large insertions in *y-ac* (inserts/kb/chromosome surveyed) is 0.0016 while the average density

TABLE 7

Nucleotide data from *per*, *Pgd*, *y-ac*, *rosy*, and *Adh* gene regions used in the HKA test of neutrality in *D. melanogaster* and *D. simulans*

Region	Species	Within species			Between species		
		m	*k*	Effective number of nucleotides surveyed	*D*	Effective number of nucleotides surveyed	*N*
y-ac	*D. melanogaster* (CA)	43	1	510	23.7	510	35
	D. melanogaster (NC)[a]	56	2	660			20
	D. melanogaster (TX)[a]	56	1	666			27
	D. melanogaster (JP)[a]	56	1	666			17
	D. simulans	43	1	510			36
Pgd	*D. melanogaster*	30	2	348	8.9	333	35
	D. simulans	28	3	318			36
per	*D. melanogaster*	39	4	444	21.2	513	35
	D. simulans	54	11	582			36
rosy	*D. melanogaster*[b]	41	7	450	19.2	477	60
	D. simulans[b]	56	28	504			30
Adh 5′ flanking							
	D. melanogaster[c]	—	30	1243	77[d]	1243	11

Only sites spanned by probes are included. Effective number of nucleotides surveyed by restriction map studies within species estimated following HUDSON (1982), as $(2m\text{-}k)j$, where m is the total number of restriction sites observed (fixed and polymorphic), k is the number of polymorphic restriction sites or nucleotides, and j is the number of base pairs in the restriction enzyme site ($j = 6$ for all studies here). Effective number of nucleotide sites surveyed by restriction map studies between species is the mean of the effective number of nucleotides in *D. melanogaster* and *D. simulans*. D is the average pairwise difference of sites between species. N is the number of sequences sampled in each species.

[a] Data from AGUADÉ, MIYASHITA and LANGLEY (1989).
[b] Data from AQUADRO, LADO and NOON (1988).
[c] DNA sequence data from KREITMAN and HUDSON (1991).
[d] Number of differences between two randomly chosen DNA sequences.

TABLE 8

Test statistics from HKA tests of neutrality in *D. simulans*

Region	*y-ac*	*Pgd*	*per*	*rosy*[a]
y-ac	—	3.42	4.71*	9.23**
Pgd		—	0.09	1.80
per			—	1.64
rosy[a]				—

* $P < 0.05$, ** $P < 0.005$.
[a] Data from AQUADRO, LADO and NOON (1988).

TABLE 9

Test statistics from HKA tests of neutrality in *D. melanogaster*

Region	*per*	*Pgd*	*rosy*[a]	*Adh* 5′ flanking[b]
y-ac (CA)	1.97	3.05	2.55	4.31*
y-ac (NC)[c]	1.11	0.87	1.64	3.28
y-ac (TX)[c]	2.57	2.24	3.21	5.01*
y-ac (JP)[c]	2.10	1.84	2.65	4.12*
per		<0.01	0.05	0.33
Pgd			0.04	0.55
rosy[a]				0.54

* $P < 0.05$. For the *y-ac* data: CA refers to this study (California); NC, TX, and JP refer to data from samples from North Carolina, Texas, and Japan.
[a] Data from AQUADRO, LADO and NOON (1988).
[b] Data from KREITMAN and HUDSON (1991).
[c] Data from AGUADÉ, MIYASHITA and LANGLEY (1989).

over several previously surveyed gene regions in this species was 0.0006 (AQUADRO 1990); furthermore we observed no inserts in the *Pgd* and *per* regions in our sample. An extensive survey of additional kilobases in the *y-ac* region of *D. simulans* should be pursued to determine if there are significantly more insertions in this region.

Selection at *Pgd* in *D. melanogaster*? It has been suggested previously that the Fast/Slow allozyme polymorphism at *Pgd* in *D. melanogaster* is a target of selection (BIJLSMA and VAN DELDEN 1977; BIJLSMA and KERVER 1980; BIJLSMA and VAN DER MEULEN-BRUIJNS 1979; CAVENER and CLEGG 1981a,b; OAKESHOTT *et al.* 1983). Our restriction site data are consistent with the notion that some form of balancing selection is influencing patterns of DNA variation at *Pgd*. Since the polymorphic sites in the *Pgd* region are in nearly complete linkage disequilibrium with the amino acid polymorphism underlying the allozyme difference, balancing selection on the Fast/Slow polymorphism leading to intermediate frequencies of Fast and Slow electromorphs will also cause polymor-

phic restriction sites to appear at intermediate frequency. Alternatively, the same pattern could result from balancing selection outside *Pgd* at a site in strong nonrandom association with the Fast/Slow polymorphism. Under balancing selection one expects to observe an excess of segregating sites in a region around the selected site because regions affected by balancing selection remain in a population for a longer period of time than expected under neutrality, thereby accumulating excess polymorphism. The extent of the excess polymorphism depends on the mutation rate, effective population size, and recombination rate (HUDSON and KAPLAN 1988). That the HKA test did not reveal an excess of segregating sites in the *Pgd* region (Table 9) may simply be a consequence of the low resolution afforded by the six-cutter method or may reflect a recent origin for the putative balanced polymorphism. The unusual pattern seen at *Pgd* could also result from a selective substitution in progress at *Pgd* or at a linked site. We are currently sequencing a representative sample of *Pgd* alleles from a natural population in order to understand the history of this region in *D. melanogaster*.

Hitchhiking of the *y-ac* region in *D. simulans*: The HKA test reveals that patterns of polymorphism and divergence at *y-ac* compared to *per* and *rosy* in *D. simulans* are incompatible with a neutral model. There is no evidence to suggest dramatic differences in functional constraint or mutation rate between the *y-ac* region and previously surveyed regions. The 5.4% divergence at *y-ac* is similar to estimates from other regions, which range from about 3% for *hsp70* and *Pgd* to 5% for *per* and *rosy* (LEIGH BROWN and ISH-HOROWICZ 1981; LANGLEY, MONTGOMERY and QUATTLEBAUM 1982; AQUADRO, LADO and NOON 1988; data from this study). In contrast, levels of variation at *y-ac* are the lowest observed to date for any gene region in any population of *D. simulans* ($\hat{\pi}$ = 0.0001 for *y-ac*, *vs.* 0.0011 for *Pgd*, 0.007 for *per*, all on the *X* chromosome, 0.015 for *Adh* on the second chromosome, and 0.019 for *rosy* on the third chromosome; AQUADRO, LADO and NOON 1988; AQUADRO 1990; and present data). Thus, the departure from neutrality appears to result from reduced polymorphism at *y-ac* in *D. simulans*, consistent with a selective sweep in this region in this species. The fact that the two significant departures from neutrality involved *y-ac* provide additional support that reduced variation at the *y-ac* region is the cause of the departure.

Not only is variation at *y-ac* reduced relative to *per* and *rosy* in *D. simulans*, but *Pgd* and *per* have considerably lower estimated heterozygosity per nucleotide than the autosomal genes surveyed in this species, even when corrected for the expected difference between autosomal and *X*-linked genes (AQUADRO, LADO and NOON 1988; AQUADRO 1990). It is possible that the lower estimated levels of variation for these *X*-linked gene regions compared to *rosy* and *Adh* result from more effective selection against restriction site variation on the *X* chromosome or from selective sweeps across these regions of the *X* chromosome. We hasten to add, however, that analysis of the current data do not reveal a statistically significant departure from neutrality for the *Pgd* and *per* regions. Additional data will be required to gain a better understanding of the forces determining levels of variation in this part of the *X* chromosome in *D. simulans*.

The *y-ac* gene region in *D. melanogaster*: The relatively low level of variation typical of *D. melanogaster* makes it difficult to demonstrate a statistically significant reduction in variation. Our estimates of heterozygosity per nucleotide at *y-ac* are considerably lower than those from *Pgd* and *per* sampled from the same population, and lower than those from most other gene regions surveyed in *D. melanogaster* (AQUADRO 1990; LANGLEY 1990). Significant departures from neutrality were revealed by HKA tests comparing *y-ac* to 5′ flanking *Adh* (Table 9). As stated earlier, the neutral mutation rate in the *y-ac* region does not appear to be lower than that for other regions based on a comparison of sequence divergence in the *y-ac* region with that observed in other gene regions. Thus, a reasonable interpretation of the HKA tests is that levels of neutral variation in this region in *D. melanogaster*, as well as in *D. simulans*, have been reduced by the hitchhiking effect. However, there are several reasons why this conclusion must be considered cautiously.

The *Adh* 5′ flanking *vs. y-ac* HKA test was significant for some populations but not for others. This reflects differences in the level of variation detected at *y-ac* in different population samples (our data; AGUADÉ, MIYASHITA and LANGLEY 1989). There are several possible explanations for this result. First, a hitchhiking event (or events) may have occurred relatively recently such that the advantageous mutant (or mutants) has spread to some populations and not to others. Another explanation is that variation is lower in California, Texas, and Japan compared to North Carolina not because of different histories of the *y-ac* region *per se*, but rather because of different population histories; local effective population sizes in California, Texas, and Japan may be smaller than that of North Carolina. However, these explanations seem unlikely, at least for the California compared to the North Carolina population. Four-cutter restriction site data from the *Adh* region (KREITMAN and AGUADÉ 1986) and six-cutter data from the *per* region (this study and D. STERN, W. NOON, E. KINDAHL and C. AQUADRO, unpublished results) reveal that at the DNA sequence level, there is neither detectable genetic differentiation nor significantly different amounts of variation between *D. melanogaster* populations from North Carolina and California.

One must also be concerned that the 5′ flanking region of *Adh* may be elevated in variability due to its proximity to regions in the *Adh* structural gene under balancing selection (KREITMAN and HUDSON 1991). The lack of significant departures from neutrality for *rosy*, *Pgd*, and *per* comparisons to 5′ *Adh* may result from the fact that there is relatively little power to detect departures from neutrality with the small number of nucleotides surveyed in most six-cutter studies. Another concern regarding sampling is that the 11 *Adh* sequences, a worldwide sample, were compared to samples from single geographic locations. Finally the disparate HKA results for *y-ac* from different populations may result from sampling error; one polymorphic site was observed in three populations (CA, TX, JP), while two polymorphic sites were observed in the fourth (NC).

Therefore, while there is some evidence for hitchhiking in the *y-ac* region in *D. melanogaster*, it remains important to examine many more nucleotides in population samples from several geographic regions. It is also important that data from the *y-ac* region and data from other gene regions compared to *y-ac* derive from the same population. It would be useful to apply TAJIMA's test to the *y-ac* region and compare these results to the those from the HKA test. Unfortunately, there are too few polymorphic nucleotides in the presently available data to use the test in a meaningful way.

Comparison of levels of variation between species: The level of variation we observed across the *per* region indicates that *D. simulans* is two to six times more variable (depending on the measure used) than *D. melanogaster*. The magnitude of this difference is consistent with that observed in previous studies of the *rosy* and *Adh* gene regions (AQUADRO, LADO and NOON 1988; AQUADRO 1990). Our data for *Pgd* and *y-ac* indicate that π is larger in *D. melanogaster*. However, in light of our evidence that selection has raised levels of variation at *Pgd* in *D. melanogaster* and decreased levels of variation at *y-ac* in *D. simulans*, we feel it is premature to discard the earlier generalization that *D. simulans* has considerably higher average levels of neutral DNA variation than *D. melanogaster*. Additional data from unlinked gene regions will be required before the issue is settled.

CONCLUSIONS

We have found evidence that the distributions of DNA variation in the *y-ac* region of *D. simulans* and the *y-ac* and *Pgd* regions of *D. melanogaster* have been influenced by selection. The *Pgd* region of *D. melanogaster* shows a departure from neutrality in the direction of balancing selection, consistent with previous studies on the allozyme polymorphism at this locus. The *y-ac* region in *D. simulans* has severely reduced nucleotide polymorphism but typical levels of sequence divergence compared to other gene regions in this species. Comparison of *y-ac* to *per* and *rosy* in *D. simulans* by the HKA test reveal significant departures from neutrality which we interpret as evidence of a selective sweep at *y-ac*. There was also a significant reduction in polymorphism at *y-ac* in some *D. melanogaster* populations compared to the *Adh* 5′ flanking region. The precise mechanism(s) or target(s) of selection affecting the *y-ac* region is unknown but may include selection for a favorable mutation, meiotic drive, and/or biased gene conversion (*e.g.*, HILLIS *et al.* 1991). The low rate of recombination and low level of variation across this region will make it difficult to determine precisely where and how selection has acted.

We thank S. BRYANT, J. COYNE, R. MACINTYRE, and M. TURELLI for fly stocks. P. GEYER, G. KARPEN, J. LUCCHESI, M. SCOTT and M. YOUNG kindly provided clones and/or unpublished data. The allozyme analysis was conducted at the Cornell Laboratory for Ecological and Evolutionary Genetics under the direction of B. MAY. We thank I. MONTELEONE for making fly media. We thank M. KREITMAN, an anonymous reviewer, and members of the Aquadro lab for comments. We especially thank R. HUDSON for many useful comments on earlier versions of this manuscript, for many useful discussions regarding the HKA test and for modifying the original HKA test to allow comparison of *X*-linked and autosomal genes. This work was supported by a grant from the National Institutes of Health (GM 36431) to C. F. A. and a National Institutes of Health predoctoral traineeship to D. J. B..

LITERATURE CITED

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

AQUADRO, C. F., 1990 Contrasting levels of DNA sequence variation in Drosophila species revealed by "six-cutter" restriction map surveys, pp. 179–198 in *Molecular Evolution, UCLA Symposium on Molecular and Cell Biology*, edited by M. CLEGG and S. O'BRIEN. Alan R. Liss, New York.

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

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.

ASHBURNER, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

BARGIELLO, T. A., and M. W. YOUNG, 1984 Molecular genetics of a biological clock in Drosophila. Proc. Natl. Acad. Sci. USA **81:** 2142–2146.

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

BENDER, W. P., 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.

BIJLSMA, R., and J. W. M. KERVER, 1980 The effect of DDT on the polymorphism at the *G6pd* and *Pgd* loci in *Drosophila melanogaster*. Genetics **103:** 447–464.

BIJLSMA, R., and W. VAN DELDEN, 1977 Polymorphism at the *G6PD* and *6PGD* loci in *Drosophila melanogaster*. I. Evidence for selection in experimental populations. Genet. Res. **30:** 221–236.

BIJLSMA, R., *and* C. VAN DER MEULEN-BRUIJNS, 1979 Polymorphism at the *G6pd* and *6Pgd* loci in *Drosophila mela-*

nogaster. III. Developmental and biochemical aspects. Biochem. Genet. **17:** 1131–1144.

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

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:** 327–338.

CAVENER, D. R., and M. T. CLEGG, 1981a Temporal stability of allozyme frequencies in a natural population of *Drosophila melanogaster*. Genetics **98:** 613–623.

CAVENER, D. R., and M. T. CLEGG, 1981b Evidence for biochemical and physiological differences between enzyme genotypes in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **78:** 4444–4447.

CHOUDHARY, M., and R. S. SINGH, 1987 A comprehensive study of genic variation in natural population of *Drosophila melanogaster*. III. Variation in genetic structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila simulans*. Genetics **117:** 697–710.

DAVID, J. R., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

DUBININ, N. P., N. N. SOKOLOV and G. G. TINIAKOV, 1937 Crossing-over between the genes "yellow" and "scute." Drosophila Inform. Ser. **8:** 76.

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

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6.

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.

GUTIERREZ, A. G., A. C. CHRISTENSEN, J. E. MANNING and J. C. LUCCHESI, 1989 Cloning and dosage compensation of the 6-*phosphogluconate dehydrogenase gene* (*Pgd*$^+$) of *Drosophila melanogaster*. Dev. Gen. **10:** 155–161.

HARRIS, H., and D. A. HOPKINSON, 1976 *Handbook of Allozyme Electrophoresis in Human Genetics*. North Holland, Amsterdam.

HILLIS, D. M., C. MORITZ, C. A. PORTER and R. J. BAKER, 1991 Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science **251:** 308–310.

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

HUDSON, R. R., and N. L. KAPLAN, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

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

KAPLAN, N. L., 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, 1986 Genetic uniformity in two populations of *Drosophila melanogaster* revealed by filter hybridization of four nucleotide recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA. **83:** 3562–3566.

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

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. Baikufum, Japan.

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.

LANGLEY, C. H., E. MONTGOMERY, R. R. HUDSON, N. L. KAPLAN and B. CHARLESWORTH, 1988 On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52:** 223–235.

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

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.

LINDSLEY, D., and G. ZIMM, 1985 The genome of *Drosophila melanogaster*. Drosophila Inform. Serv. **62.**

LINDSLEY, D., and G. ZIMM, 1990 The genome of *Drosophila melanogaster*. Drosophila Inform. Ser. **68.**

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.

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

MARUYAMA, T., and P. A. FUERST, 1984 Population bottlenecks and nonequilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. Genetics **108:** 745–763.

MAY, B., 1991 Gel electrophoresis of allozymes, in *The Molecular Analysis of Populations: A Practical Approach*, edited by A. R. HOELZEL. IRL Books, Oxford University Press (in press).

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

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

MONTGOMERY, E., B. CHARLESWORTH and C. H. LANGLEY, 1987 A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. Genet. Res. **49:** 31–41.

NEI, M., and L. JIN, 1989 Variances of the average numbers of nucleotide substitutions within and between populations. Mol. Biol. Evol. **6:** 290–300.

NEI, M., and W.-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

OAKESHOTT, J. G., G. K. CHAMBERS, J. B. GIBSON, W. F. EANES and D. A. WILLCOCKS, 1983 Geographic variation in *G6pd* and *Pgd* allele frequencies in *Drosophila melanogaster*. Heredity **50:** 67–82.

SINGH, R. S., and L. R. RHOMBERG, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. Genetics **115:** 313–322.

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.

STURTEVANT, A. H., C. B. BRIDGES, T. H. MORGAN, L. V. MORGAN and J. C. LI, 1929 Contributions to the genetics of *Drosophila simulans* and *Drosophila melanogaster*. Carnegie Inst. Wash. **399:** 1–62.

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

TAJIMA, F., 1989b The effect of change in population size on DNA polymorphism. Genetics **123:** 597–601.

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

WESTNEAT, D. F., W. A. NOON, H. K. REEVE and C. F. AQUADRO, 1988 Improved hybridization conditions for DNA "fingerprints" probed with M13. Nucleic Acids Res. **16:** 4161.

Communicating editor: R. R. HUDSON