Reduced Variation in the *yellow-achaete-scute* Region in Natural Populations of *Drosophila melanogaster*

Montserrat Aguadé,* Naohiko Miyashita[†,1] and Charles H. Langley[†]

**Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain; and [†]Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709*

Manuscript received August 9, 1988
Accepted for publication March 13, 1989

ABSTRACT

Restriction map variation in 64 *X* chromosome lines extracted from three different populations of *Drosophila melanogaster* was investigated with seven six-nucleotide-recognizing restriction enzymes for a 106-kb region encompassing the *yellow* gene and the *achaete-scute* complex that is located in the region of reduced crossing over close to the telomere. Nine restriction site polymorphisms (out of 176 sites scored) and 19 length polymorphisms (15 insertions and 4 deletions) were detected. The estimated level of heterozygosity per nucleotide, $H = 0.0003$, is much lower than that reported for autosomal and sex-linked loci located in regions with normal levels of crossing over. The overall frequency of polymorphic restriction sites is reduced. Six out of nine restriction site polymorphisms are unique and the other three have frequencies less than 0.17. Some large insertions have reached relatively high frequencies, 0.08 to 0.17. Consistent with the theoretically predicted negative relationship between crossing over and the magnitude of linkage disequilibrium, an increase in the relative number of nonrandom associations was observed in the *y-ac-sc* region.

THE infinite-site model of the neutral theory of molecular evolution (KIMURA 1971) provides a basis for analyzing DNA sequence variation in natural populations. The analytical development of this model has been based either on no recombination (WATTERSON 1975) or on free recombination (EWENS 1979) while properties of the model under intermediate levels (not zero) recombination have been studied by simulation (HUDSON 1983). The expectation of the number of segregating sites does not differ under the different models, but the variance of the number of segregating sites with restricted recombination will be larger than under free recombination. Restricted recombination will also cause stronger linkage disequilibria among segregating sites. The hitchhiking effects of directional selection at a few sites on levels of standing variation and linkage disequilibrium are also sensitive to the level of crossing over (MAYNARD SMITH and HAIGH 1974; HUDSON and KAPLAN 1985, 1988). Thus we might expect to see considerable differences in the quantity and quality of molecular genetic variation in natural populations among regions of the genome showing large differences in the crossing over per unit length of DNA.

To compare molecular variation in natural populations in regions with large differences in crossing over requires detailed comparison of the genetic and physical maps and a battery of cloned and characterized genes from appropriate regions. The contraction of the genetic map near telomeres and centromeres is well documented in several *Drosophila* species (LINDSLEY and SANDLER 1977; MORIWAKI and TOBARI 1975) and has been considered good evidence of regions of restricted recombination. Both telomeres and centromeres show specific effects in suppressing recombination over long stretches of euchromatin. The *X* chromosome of *Drosophila melanogaster* provides the best opportunity to investigate the effects of variation in crossing over on restriction map polymorphism. Regions close to the tip and the euchromatic base of the *X* chromosome have been cloned and characterized in *D. melanogaster*, providing an opportunity to study molecular variation in regions of restricted recombination.

The region analyzed in the present study includes both the *yellow* locus and the *achaete-scute* complex (AS-C), which are located at the tip of the *X* chromosome of *D. melanogaster* (all at genetic position 1-0.0 and at cytological positions 1A5-8, 1A5-8 and 1B3, respectively; LINDSLEY and GRELL 1972). The *yellow* gene is involved in controlling the pattern of pigmentation of the adult and larval cuticle of Drosophila. Cloning and sequencing (CAMPUZANO *et al.* 1985; BIESSMANN 1985; PARKHURST and CORCÉS 1986; GEYER, SPANA and CORCÉS 1986) indicate that it consists of a single transcription unit with two exons separated by a long intron. The presumed *yellow* protein is 541 amino acids long and is thought to be a

[1] Present address: Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan.

membrane or secreted protein which plays a structural and not a catalytic role in pigmentation. Germ-line transformation has shown that a genomic fragment containing the coding region plus 2.8 kb of 5′ flanking sequences is sufficient to restore wild-type pigmentation (CHIA *et al.* 1987).

The *achaete-scute* complex mediates the differentiation of the bristles and hairs that cover the cuticle of the adult fly. Genetic analysis has subdivided the complex into four regions: *achaete*, *scute α*, *lethal of scute* (*l'sc*), *scute β* (MULLER 1955; GARCÍA-BELLIDO 1979) and *scute γ* or *asense* (DAMBLY-CHAUDIÈRE and GHYSEN 1987; GHYSEN and DAMBLY-CHAUDIÈRE 1988). The whole AS-C has been recently cloned (CAMPUZANO *et al.* 1985) and its transcription pattern analyzed. At least eight different transcription units have been identified in the complex (CAMPUZANO *et al.* 1985; VILLARES and CABRERA 1987), which are separated by long stretches of apparently untranscribed DNA. It is within these noncoding regions where most lesions associated with the nonlethal *ac* or *sc* phenotypes have been localized. A subset of the AS-C genes has been sequenced (T3, T4 and T5 by VILLARES and CABRERA 1987; VILLARES, GONZALES and MODOLELL 1987); analysis of these sequences and of their homology to other transcription units within the complex has suggested that the AS-C encodes several homologous polypeptides with redundant functions (T1a, T3, T4, and T5). Both sequence information and knowledge about the temporal and spatial expression of some of these genes (T3, T4, and T5, ROMANI, CAMPUZANO and MODOLELL 1987; CABRERA, MARTINEZ-ARIAS and BATE 1987) suggest their specific involvement in the process of neuroblast segregation. It should be noted that not all transcription units within the AS-C are involved in chaetae formation or neurogenesis. In fact T2, located in the *scute β* region codes for an aspartic proteinase that is expressed in the larval gut and most likely unrelated to *scute* or *achaete* phenotypes (VILLARES, GONZALEZ and MODOLELL 1987).

In this paper we report the analysis of naturally occurring variation in a 106-kb segment encompassing both the *yellow* locus and the *achaete-scute* complex at the tip of the *X* chromosome of *D. melanogaster*. The amount of crossing over in this region is known to be quite low: 1×10^{-6} cross overs per kilobase. This estimate is based on the 5 recombinants in the ac^3-sc^1 interval observed by DUBININ, SOKOLOV and TINIAKOV (1937) in 75578 progeny and the 57 kb between these markers determined by CAMPUZANO *et al.* (1985) (see Figure 1). The observed levels of crossing over in the regions between *zeste* and *white* and within the *white* locus (more proximal to the centromere) are approximately 20-fold higher: 2×10^{-5} crossovers per kilobase (B. JUDD, personal communication). Restriction map variation in a sample of 64 independent isogenic *X* chromosome lines from three different populations of *D. melanogaster* is reported. The results of this survey are compared to a similar study (BEECH and LEIGH BROWN 1989) and surveys of other regions of the genome of Drosophila.

MATERIALS AND METHODS

Fly stocks: Sixty-four isogenic *X* chromosome lines extracted from three different populations (20 from North Carolina, 27 from Texas and 17 from Fukuoka, Japan) have been used in the present study. Lines were constructed as described in MIYASHITA *et al.* (1986).

Restriction map analysis: Seven hexanucleotide-recognizing restriction enzymes were used: *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Pvu*II, *Sal*I and *Sac*I. Genomic DNA was CsCl purified according to BINGHAM, LEVIS and RUBIN (1981). Seven single digestions and two double digestions (*Eco*RI and *Bam*HI, *Hin*dIII and *Sac*I) were performed. Digestion, electrophoresis, blotting and probing was carried out as previously described (MIYASHITA and LANGLEY 1988). The probes used in the present analysis were the λ phage clones of CAMPUZANO *et al.* (1985): λsc112, λsc101, λsc94, λsc64, λsc22, λsc17, λsc14, λsc31 and λsc53, which cover a 106kb region which includes the *yellow* (λsc112), *achaete* (λsc112 and λsc101) and *scute* (rest of λ clones) genes.

RESULTS

A total of 9 restriction site polymorphisms (out of 176 sites scored) and 19 length polymorphisms (15 insertions and 4 deletions) have been detected in the 64 *X* chromosome lines examined. Figure 1 shows the location of polymorphic sites and of insertions/deletions, based on the coordinate system of CAMPUZANO *et al.* (1985). Location of monomorphic sites is not shown, but can be found in CAMPUZANO *et al.* (1985) for five of the enzymes used (*Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I and *Sal*I) and in the legend to Figure 1 for *Pvu*II and *Sac*I. All length variants and six out of nine restriction site variants are located in noncoding regions; the other three restriction site polymorphisms are located in coding regions (*Pst*I site at 19.6 in T3, *Bam*HI site at −0.4 in T2 and *Eco*RI site at −26.0 in T1a).

Six out of 9 restriction site polymorphisms are represented only once in the sample, five being unique to North Carolina and one to Fukuoka. The other three restriction site polymorphisms are multiply represented and present either in North Carolina and Texas or in North Carolina and Fukuoka.

The observed insertions vary in length from 200 bp to 6.5 kb. The deletions found within the 106-kb range in size from 100 to 400 bp. Thirteen out of 15 insertions (whether unique or multiply represented) are present in only one of the three geographical areas. The remaining two insertions (ins11 and ins15) are present in both American samples studied, and in Texas and Fukuoka, respectively. In the case of deletions, only one (del2) is present in both American

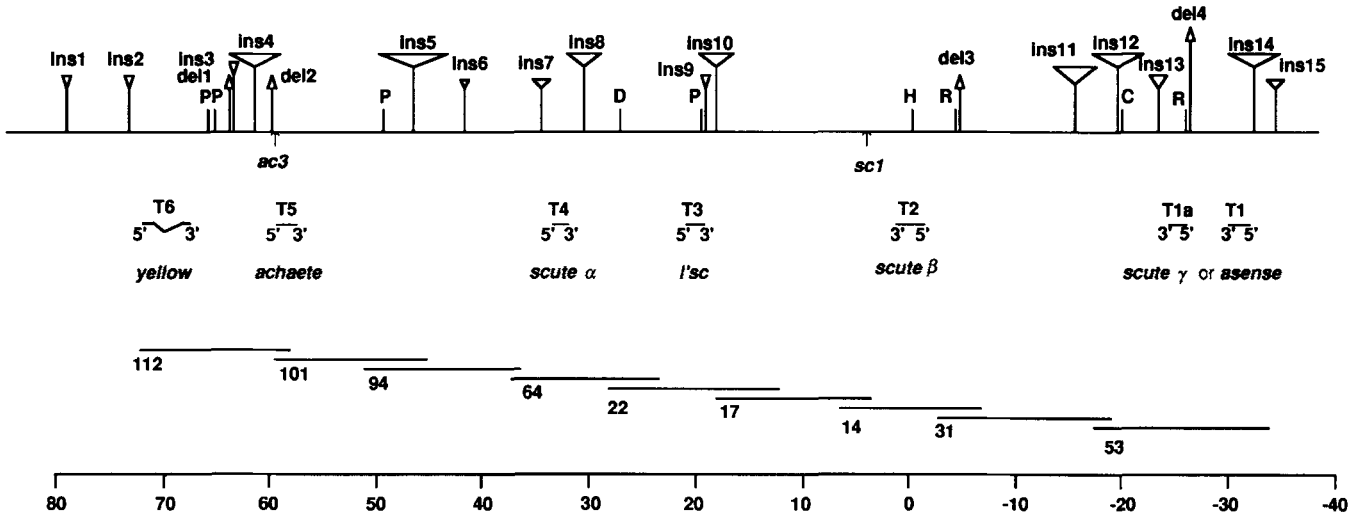

FIGURE 1.—Restriction site and length polymorphisms in the *y-ac-sc* region. Coordinates are given according to Campuzano *et al.* (1985). H, *Bam*HI; R, *Eco*RI; D, *Hin*DIII; P, *Pst*I; C, *Sac*I; ins1, 1 kb; ins2, 0.6 kb; ins3, 0.2 kb; ins4, 5 kb; ins5, 6.5 kb; ins6, 0.4 kb; ins7, 1.2 kb; ins8, 3.2 kb; ins9, 0.6 kb; ins10, 3 kb; ins11, 3.5 kb; ins12, 5 kb; ins13, 1.4 kb; ins14, 4.6 kb; ins15, 1.7 kb; del1, 0.4 kb; del2, 0.1 kb; del3, 0.15 kb; del4, 0.2 kb. The overlapping phage clones λsc112, λsc101, λsc94, λsc64, λsc17, λsc14, λsc31 and λsc53 used as probes are represented as 112, 101, 94, 64, 22, 17, 14, 31 and 53, respectively. Location of *Pvu*II sites: 75.6, 71.6, 68.8, 67.9, 65.2, 59.2, 55.4, 45.7, 42.85, 35.8, 35.4, 30.2, 26.1, 20.9, 19.2, 11.3, 7.0, 5.1, −3.4, −6.95, −11.0, −13.7, −15.5, −19.3, −25.1, −26.35, −37.35. There are two additional sites between sites 55.4 and 45.6 which could not be unambigously located in the present study, but which give rise to three fragments 0.95 kb, 2.2 kb and 6.5 kb long. Location of *Sac*I sites: 78.4, 65.4, 64.7, 61.8, 55.3, 50.3, 44.1, 28.4, −6.4, −7.4, −12.2, −14.3, −17.4, −17.9, −20.4, −22.6, −54.6.

FIGURE 2.—Frequency spectrum of restriction site and length polymorphisms. Polymorphisms are labeled by the notation in Table 1.

samples. The other three deletions are restricted to one geographical area.

Figure 2 shows the frequency spectrum of restriction sites, insertions and deletions. Previously reported surveys for both autosomal and sex-linked loci (LANGLEY, MONTGOMERY and QUATTLEBAUM 1982, LEIGH-BROWN 1983, AQUADRO *et al.* 1986; LANGLEY and AQUADRO 1987; AQUADRO, LADO and NOON 1988; MIYASHITA and LANGLEY 1988; AGUADÉ, MIYASHITA and LANGLEY 1988) indicated that large insertions were individually rare (usually unique) while the frequencies of restriction site polymorphisms and some small insertions/deletions are more continuously distributed from high to low frequencies. In contrast, in the *yellow-achaete-scute* region the overall frequency of restriction site polymorphism is re-

TABLE 1

Frequency distribution of haplotypes

Haplotype No.	ins1 78.8 1	ins2 72.8 2	*Pst*I 65.7 3	*Pst*I 65.1 45	del1 64 5	ins3 63.6 6	ins4 61.2 7	del2 60.4 8	*Pst*I 49.4 9	ins5 46.4 10	ins6 41.6 11	ins7 34.4 12	ins8 30.4 13	*Hin*dIII 27.0 14	*Pst*I 19.6 15
1	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
2	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
3	−	−	−	+	−	−	−	+	+	−	−	−	−	−	+
4	−	−	+	+	−	−	−	−	+	−	−	+	−	−	+
5	−	−	+	+	−	−	−	−	+	−	−	−	−	−	−
6	−	−	−	+	−	−	−	+	+	−	−	−	−	−	+
7	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−
8	−	−	+	+	−	−	−	−	+	−	−	−	+	−	+
9	−	+	+	+	−	−	−	−	+	−	−	−	−	−	+
10	−	−	+	+	−	+	−	−	+	−	−	−	−	−	+
11	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
12	−	−	+	−	+	−	−	+	+	−	−	−	−	−	+
13	−	−	−	+	−	−	−	−	+	−	−	−	−	+	+
14	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
15	−	−	+	+	−	−	+	−	+	−	−	−	−	−	+
16	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
17	−	−	−	+	−	−	−	−	+	+	−	−	−	−	−
18	−	−	−	+	−	−	−	−	+	−	−	−	−	−	−
19	+	−	+	+	−	−	−	−	+	−	−	−	−	−	+
20	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
21	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
22	−	−	+	+	−	−	−	−	−	−	−	−	−	−	+
23	−	−	+	+	−	−	−	−	+	−	+	−	−	−	+
24	−	−	+	+	−	−	−	−	+	−	−	−	−	−	+
Freq.															
Total	.02	.02	.16	.02	.02	.02	.02	.09	.02	.02	.02	.05	.03	.02	.09
NC	.00	.05	.25	.05	.05	.05	.00	.25	.00	.00	.00	.00	.00	.05	.10
TX	.04	.00	.18	.00	.00	.00	.04	.04	.00	.04	.00	.11	.00	.00	.15
JPN	.00	.00	.00	.00	.00	.00	.00	.00	.06	.00	.06	.00	.12	.00	.00

duced. Six out of nine polymorphisms are unique and the other three have frequencies less than 0.17. On the other hand, some large insertions, *i.e.*, ins11, ins15 and ins10, have reached relatively high frequencies (0.17, 0.09, and 0.08 respectively).

Table 1 shows the 24 different haplotypes observed and their frequency distribution. Only the most common haplotype (#1) is present in the three populations and only one other haplotype (#3) is present in both North American samples. The distribution of haplotypes differs in the three populations ($P < 0.0004$, after 50,000 trials using a Monte-Carlo 24 × 3 contingency table test; a modification of the procedure in LEWONTIN and FELSENSTEIN 1965; W. ENGELS, personal communication), mainly due to haplotypes present only in one geographical area. Most of the haplotypes unique to each population are related by a single mutational event to the most common haplotype (#1) present in the three populations. Two haplotypes (#2 and 6) contain multiple insertions (ins9, ins10 and ins11 for haplotype #2, and ins11 and ins12 for haplotype #6) within a relatively short region.

Variation attributable to nucleotide substitution can be estimated from restriction site polymorphisms by different methods. Table 2 gives the estimates for each population as well as for the pooled data obtained by three different methods (NEI and LI 1979, NEI and TAJIMA 1981; ENGELS 1981; HUDSON 1982). Estimates obtained according to NEI and TAJIMA (1981) and to ENGELS (1981) are in all cases very close. The method of HUDSON, which provides an estimate of $\theta = 4N_e\mu$ (N_e is the effective population size and μ is the mutation rate to selective neutral nucleotide substitutions) under the assumption of stationarity and selective neutrality is a function of the number of segregating sites and not their frequencies; it can thus differ from the other two estimators when the number of rare sites is high (as in the case of North Carolina). The fact that the estimates of θ exceed those of H or π may indicate some deviation from the assumption of selective neutrality.

Variation for insertion/deletion in the *yellow-achaete-scute* region has been quantified as the number of insertions/deletions larger than 100 bp, relative to the most comon haplotype, per kilobase surveyed: 0.0028, 0.0035 and 0.0022 in North Carolina, Texas anf Fukuoka, respectively, and 0.0027 for the pooled data.

ins9 19.0 16	ins10 18.4 17	*Bam*HI −0.4 18	*Eco*RI −4.4 19	del3 −4.8 20	ins11 −15.6 21	ins12 −19.6 22	*Sac*I −20.0 23	ins13 −23.6 24	*Eco*RI −26.0 25	del4 −26.4 26	ins14 −32.4 27	ins15 −34.4 28	Total	NC	TX	JPN
−	−	−	−	−	−	−	−	−	+	−	−	−	29	8	11	10
+	+	−	−	−	+	−	−	−	+	−	−	+	5	0	5	0
−	−	−	−	−	+	−	−	−	+	−	−	−	3	2	1	0
−	−	−	−	−	−	−	−	−	+	−	−	−	3	0	3	0
−	−	−	−	−	−	−	−	−	+	−	−	−	2	2	0	0
−	−	−	−	−	+	+	−	−	+	−	−	−	2	2	0	0
−	−	−	−	−	−	−	−	−	+	−	−	−	2	0	2	0
−	−	−	−	−	−	−	−	−	+	−	−	−	2	0	0	2
−	−	−	−	−	−	−	−	−	+	−	−	−	1	1	0	0
−	−	−	−	−	−	−	−	−	+	−	−	−	1	1	0	0
−	−	+	+	−	−	−	−	−	+	−	−	−	1	1	0	0
−	−	−	−	−	+	−	−	−	+	−	−	−	1	1	0	0
−	−	−	−	−	−	−	+	−	+	+	−	−	1	1	0	0
−	−	−	−	−	−	−	−	−	−	−	−	−	1	1	0	0
−	−	−	−	−	−	−	−	−	+	−	−	−	1	0	1	0
−	−	−	−	−	−	−	−	−	+	−	+	−	1	0	1	0
−	−	−	−	+	−	−	−	−	+	−	−	−	1	0	1	0
−	−	−	−	+	−	−	−	−	+	−	−	−	1	0	1	0
−	−	−	−	−	−	−	−	−	+	−	−	−	1	0	1	0
−	−	−	−	−	−	−	−	−	+	−	−	+	1	0	0	1
−	−	+	−	−	−	−	−	−	+	−	−	−	1	0	0	1
−	−	−	−	−	−	−	−	−	+	−	−	−	1	0	0	1
−	−	−	−	−	−	−	−	−	+	−	−	−	1	0	0	1
−	−	−	−	−	−	−	−	+	+	−	−	−	1	0	0	1
.08	.08	.03	.02	.03	.17	.03	.02	.02	.02	.02	.02	.09				
.00	.00	.05	.05	.00	.30	.10	.05	.00	.05	.05	.00	.00				
.18	.18	.00	.00	.07	.22	.00	.00	.00	.00	.00	.04	.18				
.00	.00	.06	.00	.00	.00	.00	.00	.06	.00	.00	.00	.06				

"+" and "−" indicate the presence and absence of restriction site and insertion/deletion variation. NC, North Carolina; TX, Texas; JPN, Japan. Location of restriction site and insertion/deletion polymorphisms is given according to the coordinates in CAMPUZANO *et al.* (1985). In the lower part of the table, the frequency of the less common variant is given for each polymorphism, for the pooled data and for each population.

Linkage disequilibrium between both restriction site and length polymorphisms has been analyzed for each population separately, since the frequencies of polymorphic sites as well as two-locus haplotypes do (in many cases) differ among populations (analysis not shown). Because only polymorphisms in which the rarer variant was not unique have been considered, no linkage disequilibrium analysis is possible in the Japanese sample. As shown in Table 3, 3 out of 10 comparisons in the North Carolina sample and 9 out of 28 in the Texas sample demonstrate significant departure from random association.

Genetic analysis has shown reduced levels of crossing over in the *yellow-achaete-scute* region (DUBININ, SOKOLOV and TINIAKOV 1937). The present results are consistent with low levels of crossing over. If no back mutation and no recurrent mutation occur, the presence of all four gametic types for a pair of sites indicates at least one cross over in the history of the sample (HUDSON and KAPLAN 1985). When this "four-gamete" test is used, only one out of 10 comparisons in North Carolina (between polymorphisms #3 and 8) and 1 out of 28 in Texas (between polymorphisms #3 and 21) show all four gametic classes. Because some variants are present in more than one population, the pooled data have also been considered. In this case, only 5 out of 66 comparisons show all four gametic classes (between polymorphisms 3 and 8, 3 and 15, 3 and 21, 8 and 21, 21 and 28). This low number of four-gamete pairs is clearly consistent with some restriction in crossing over in the evolution of the *y-ac-sc* region.

DISCUSSION

If variation at the DNA level were maintained by random genetic drift and purifying selection against at least partly recessive variants, genes on the *X* chromosome should show lower levels of variation than those on the autosomes because of the hemizygosity of the heterogametic sex. Recent surveys of variation at the *white* (LANGLEY and AQUADRO 1987; MIYASHITA and LANGLEY 1988), *Notch* (SCHAEFFER, AQUADRO

and LANGLEY 1988), *zeste-tko* (AGUADÉ, MIYASHITA and LANGLEY 1989) and *forked* and *vermilion* (C. H. LANGLEY and N. MIYASHITA, unpublished results) have shown no evidence for such a reduction. These results suggest that simple purifying selection does not play a major role in determining the standing level of population genetic variation at the DNA level in natural populations of *D. melanogaster*. Thus the same level of molecular variation was expected at the tip of the *X* chromosome. However, in the present study we detected a lower level of variation and also a reduction in the overall frequency distribution (Figure 2).

The unweighted average estimate of θ for various autosomal regions [0.006 for *Adh* (LANGLEY, MONTGOMERY and QUATTLEBAUM 1982, AQUADRO *et al.* 1986), 0.002 for 87A heat shock locus (LEIGH-BROWN 1983), 0.008 for *Amy* (LANGLEY *et al.* 1988b) and 0.003 for *rosy* (AQUADRO, LADO and NOON 1988)] and *X*-linked regions [0.013 for *white* (MIYASHITA and LANGLEY 1988), 0.004 for *zeste-tko* (AGUADÉ, MIYASHITA and LANGLEY 1989), 0.005 for *Notch* (SCHAEFFER, AQUARDO and LANGLEY 1988), 0.002 for *forked* (LANGLEY and MIYASHITA unpublished results) and 0.003 for *vermilion* (C. H. LANGLEY and N. MIYASHITA, unpublished results)] located in chromosomal segments where there is no evidence for restricted recombination is 0.005. While no well-documented statistical procedure had been published that can be used to compare the estimate of θ with that reported here for the *yellow-achaete-scute* region, 0.001 (9 out of 176 sites polymorphic in 64 lines), the following approach may be instructive (R. HUDSON, personal communication). Under the infinite sites model without selection or recombination, the expectation and variance of the number of segregating sites (given $\theta = 0.005$, sample size = 64 and the number of nucleotide sites examined $\approx 2 \times 176 \times 6 = 2112$) can be calculated (WATTERSON 1975) to be 51 and 241. Assuming that the (observed − expected)^2/variance $\approx \chi^2$ we can test the significance of the 9 or fewer segregating sites: $(9 - 51)^2/241 = 7.3$ with one degree of freedom; $P < 0.01$. This suggests that the reduced restriction site variation observed in the *y-ac-sc* region is inconsistent with selective neutrality with a common θ for all regions. BEECH and LEIGH-BROWN (1989) surveyed the *y-ac-sc* region in 49 *X* chromosomes utilizing four restrictions enzymes. They did not report any significant reduction in restriction site variation. This difference in the conclusions of this report and that of BEECH and LEIGH-BROWN (1989) can be attributed to the differences in the total number of restriction sites (69 *vs.* 176) and chromosomes surveyed (49 *vs.* 64) in the two studies.

Most surveys (utilizing cytological or restriction mapping) have shown that transposable elements (or large insertions) are individually quite rare in natural populations of *D. melanogaster* (LANGLEY, MONTGOMERY and QUATTLEBAUM 1982; LEIGH-BROWN 1983; MONTGOMERY and LANGLEY 1983; AQUADRO *et al.* 1986). This low frequency has been considered an indication of their deleterious effect even when inserted in noncoding regions. Under the simple assumption of natural selection acting against the additive and partially recessive effects of individual insertions, one would expect a lower level of variation for large insertion/deletions in regions of the *X* chromosome than of the autosomes. MONTGOMERY, CHARLESWORTH and LANGLEY (1987) suggested that for some elements this may not be the mechanism containing copy number.

TABLE 2

Heterozygosity per nucleotide

Estimate	Pooled	NC	TX	JPN
π	0.00033	0.00057	0.00028	0.00011
H	0.00033	0.00057	0.00028	0.00011
θ	0.00104	0.00128	0.00029	0.00033

LANGLEY *et al.* (1988a) presented a model of unequal recombination among transposable elements to explain the dynamics and observed distribution of transposable elements on *X* chromosomes and autosomes in natural populations. They also examined the density of one family of transposable elements, *roo*, in those regions of the chromosomes where normal (equal) crossing over is reduced (centromeric and telomeric). Assuming unequal recombination is also reduced in those regions they predicted increased densities of *roo*. Note also that the Muller's rachet hypothesis (FELSENSTEIN 1974; HAIGH 1978) might predict an increased density of deleterious insertions in regions of reduced crossing over (LANGLEY *et al.* 1988b). While LANGLEY *et al.* (1988b) did find an increased density of *roo* near the base of the *X* chromosome, there was no statistically significant evidence of an increase at the tip (near *y-ac-sc*). More recently, CHARLESWORTH and LAPID (1989) have extended this analysis to many more families of transposable elements and report a consistently high density in the centromere-proximal regions. The level of insertion/deletion variation (measured as the number of insertions/deletions larger than 100 bp per kilobase per chromosome surveyed) in the *yellow-achaete-scute* region, 0.0027, is comparable to that observed in other autosomal regions [0.0139 and 0.0140 for *Adh* (LANGLEY, MONTGOMERY and QUATTLEBAUM 1982; AQUADRO *et al.* 1986), 0.0055 for *hsp70* (LEIGH-BROWN 1983), 0.0086 for *Amy* (LANGLEY *et al.* 1988b), 0.0017 for *rosy* (AQUADRO, LADO and NOON 1988)] and X-linked regions [(0.0042 for *w* (MIYASHITA and LANGLEY 1988), 0.0023 for *z-tko* (AGUADÉ, MIYASHITA and LANGLEY 1989), 0.0018 for *Notch*

TABLE 3

Linkage disequilibrium expressed as R

North Carolina							
8	**15**	**21**	**22**				
0.733*	−0.192	0.733*	0.577	**3**			
	−0.192	1.000*	0.577	**8**			
		−0.192	−0.111	**15**			
			0.577	**21**			
Texas							
12	**15**	**16**	**17**	**19**	**21**	**28**	
−0.169	0.875*	−0.227	−0.227	0.593*	−0.025	−0.227	**3**
	−0.147	−0.169	−0.169	−0.100	−0.189	−0.169	**12**
		−0.199	−0.199	0.678*	−0.223	−0.191	**15**
			1.000*	−0.135	0.892*	1.000*	**16**
				−0.135	0.892*	1.000*	**17**
					−0.151	−0.135	**19**
						0.892*	**21**

[a] $P < 0.05$ by Fisher's exact test of independence. Boldtype numbers represent polymorphic sites as detailed in Table 1.

(SCHAEFFER, AQUADRO and LANGLEY 1988), 0.0029 for *forked* (C. H. LANGLEY and N. MIYASHITA, unpublished results) and 0.0020 for *vermilion* (C. H. LANGLEY and N. MIYASHITA, unpublished results)] with no restricted recombination. It is also similar to that reported by BEECH and LEIGH-BROWN 1988). Thus the restriction map surveys of the *y-ac-sc* region show no evidence for an increase in the density of large insertions and is consistent with the *in situ* survey of CHARLESWORTH and LAPID. The assumption of strict proportionality of the unequal recombination to normal crossing over may be incorrect. A quantitative test of the model in LANGLEY *et al.* (1988a) will require more detailed knowledge of the rates of unequal recombination in different regions of the chromosome.

Under the assumption of reduced recombination, one would expect higher levels of linkage disequilibrium. Despite the facts that the region under study spans over 106 kb, that the number of polymorphisms detected is small and that most of them are at low frequencies, 3 out of 10 comparisons in North Carolina and 9 out of 28 in Texas show significant linkage disequilibria (see Table 2). Linkage disequilibrium has also been analyzed in other regions of this same set of *X* chromosomes. In the *white* locus region significant linkage disequilibria were rarely observed; on average, only some pairs of polymorphic sites that are tightly linked (<2 kb apart) within the *white* transcription unit were not in linkage equilibrium (MIYASHITA and LANGLEY 1988). The survey of linkage disequilibrium in the *zeste-tko* region yielded little linkage disequilibrium (AGUADÉ, MIYASHITA and LANGLEY 1989). The *forked* and *vermilion* regions also exhibited some nonrandom association among restriction map polymorphisms (C. H. LANGLEY and N. MIYASHITA, unpublished results). It is difficult to quantify and test the differences in linkage disequilibrium in various regions of the genome. As a nonparametric test we compare the proportion of statistically significant ($P < 0.05$) pairwise comparisons in this study (12/38) to those observed in the same chromosomes using the same restriction enzymes at other loci: *white*, 32/496; *z-tko*, 9/74; *v*, 7/14; *f*, 0/1. Since the marginal frequencies of the polymorphic sites are lower in *y-ac-sc* and the distances between polymorphic sites greater (on average) this may be a conservative test. Fisher's exact test applied to the number of significant comparisons in the *y-ac-sc* region versus that from the other loci (12/38 *vs.* 48/585; $P < 0.001$) supports the conclusion that the proportion of nonrandom associations in the *y-ac-sc* region is higher than in those other regions. All these regions differ from the *yellow-achaete-scute* region in their higher level of crossing over per physical unit of length. The 10- to 20-fold reduction in crossing over per physical unit of length (10^{-6} *vs.* 2×10^{-5} per kb) in the *yellow-achaete-scute* region might well be sufficient to cause an increase in the relative number of nonrandom associations over a large region.

There are three main differences in the quantity and quality of naturally occurring molecular variation in this region at the tip of the *X* chromosome as compared to other regions where there is no evidence for restricted levels of crossing over: (1) a lower average level of nucleotide sequence variation; (2) an overall reduction in the expected frequency of heterozygotes at polymorphic restriction sites, accompanied by a higher proportion of unique restriction polymorphisms; and (3) a higher proportion of nonrandom associations spanning a large region.

According to molecular studies on the distribution of transcriptional units in the 106-kb region studied, most of this region is noncoding. Nevertheless, the

observation that most lesions associated with the nonlethal *ac* or *sc* phenotypes have been localized within those noncoding regions may indicate that at least some of these regions may play a role in gene regulation. Both observation (1) and (2) might be partly attributable to higher levels of functional constraint in this region of the genome.

The reduced level of crossing over per kilobase in the *yellow-achaete-scute* region may also account for the distinct properties observed in this region. MAYNARD SMITH and HAIGH (1974) showed that occasional selected substitutions of favorable mutations reduces through a "hitchhiking effect" the expected heterozygosity at closely linked neutral polymorphisms. Further analysis (N. L. KAPLAN, R. HUDSON and C. H. LANGLEY, unpublished results) supports the conclusion that even rarely occurring favorable mutations will significantly reduce standing molecular variation in large populations, especially in regions where crossing over is reduced. The increased proportion of nonrandom associations in the *y-ac-sc* region is also consistent with the reduction in crossing over. Finally, the "hitchhiking effect" may account for the large disparity between the estimates of θ (0.0010) and π (0.0003) in the *y-ac-sc* results. The low frequencies of virtually all the polymorphisms led to the big differences in the estimates of θ and π, since θ is a function of numbers of polymorphic restriction sites, while π is also a function of their frequencies. If neutral variation is continually recovering from being forced to monomorphism by the hitchhiking effect, then most polymorphisms will be in the early phase, *i.e.*, rare or unique in small samples. More surveys of regions with varying rates of crossing over may well lead to a quantitative understanding of the forces that shape standing variation at the DNA sequence level.

We would like to thank J. MODOLELL and S. CAMPUZANO for providing the phage clones for the *yellow-achaete-scute* region. We also thank M. MATSUDA, W. STEPHAN, C. AQUADRO and R. BEECH for helpful comments and criticism.

LITERATURE CITED

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

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.

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

BEECH, R. N., and A. J. LEIGH-BROWN, 1988 Insertion-deletion variation at the *yellow, achaete-scute* region in two natural populations of *Drosophila melanogaster*. Genet. Res. (in press).

BIESSMANN, H., 1985 Molecular analysis of the *yellow* gene (*y*) region of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **82:** 7369–7373.

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

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.

CABRERA, C. V., A. MARTINEZ-ARIAS and M. BATE, 1987 The expression of three members of the *achaete-scute* gene complex correlates with neuroblast segregation in Drosophila. Cell **50:** 425–433.

CHARLESWORTH, B., and A. LAPID, 1989 A study of ten transposable elements on X chromosomes from a population of *Drosophila melanogaster*. Genet. Res. (in press.)

CHIA, W., M. MARTIN, G. HOWES and B. MENG, 1987 On the specificity of *yellow* gene expression in Drosophila. Tenth European Drosophila Research Conference, Abstract 88.

DAMBLY-CHAUDIÈRE, C., and A. GHYSEN, 1987 Independent subpatterns of sense organs require independent genes of the *achaete-scute* complex in *Drosophila* larvae. Genes Dev. **1:** 297–306.

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

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

EWENS, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, New York.

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

GARCIA-BELLIDO, A., 1979 Genetic analysis of the *achaete-scute* system of *Drosophila melanogaster*. Genetics **88:** 469–486.

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

GHYSEN, A., and C. DAMBLY-CHAUDIÈRE, 1988 From DNA to form: the *achaete-scute* complex. Genes Dev. **2:** 495–501.

HAIGH, J., 1978 The accumulation of deleterious genes in a population—Muller's ratchet. Theor. Popul. Biol. **14:** 251–267.

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

HUDSON, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

HUDSON, R. R., and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

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

KIMURA, M., 1971 Theoretical foundations of population genetics at the molecular level. Theor. Popul. Biol. **2:** 174–208.

LANGLEY, C. H., and C. F. AQUADRO, 1987 Restriction map variation in natural populations of *Drosophila melanogaster*. *White*-locus region. Mol. Biol. Evol. **151:** 651–663.

LANGLEY, C. H., E. MONTGOMERY and W. QUATTLEBAUM, 1982 Restriction map variation at 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, 1988a On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52:** 223–235.

LANGLEY, C. H., A. E. SHRIMPTON, T. YAMAZAKI, N. MIYASHITA, Y. MATSUO and C. F. AQUADRO, 1988b Naturally occurring variation in the restriction map of the *Amy* region of *Drosophila melanogaster*. Genetics **119:** 619–629.

LEIGH BROWN, A. J., 1983 Variation at the 87A heat shock locus

in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80:** 5350–5354.

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

LINDSLEY, D. L., and E. H. GRELL, 1972 *Genetic Variations of Drosophila melanogaster*. Carnegie Inst. Washington Publ. 627.

LINDSLEY, D. L., and L. SANDLER, 1977 The genetic analysis of meiosis in female *Drosophila melanogaster*. Philos. Trans. Soc. Lond. (Biol.) **277:** 295–312.

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

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

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

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.

MONTGOMERY, E., and C. H. LANGLEY, 1983 Transposable elements in mendelian populations. II. Distribution of three *copia*-like elements in a natural population of *Drosophila melanogaster*. Genetics **104:** 473–483.

MORIWAKI, D., and Y. N. TOBARI, 1975 *Drosophila ananassae*, pp. 523–535 in *Handbook of Genetics*, Vol. 3, edited by R. C. KING. Plenum, New York.

MULLER, H. J., 1955 On the relation between chromosome changes and gene mutations. Brookhaven Symp. **8:** 126–147.

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.

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

PARKHURST, M. S., and V. G. CORCÉS, 1986 Interactions among the gypsy transposable element and the *yellow* and the *suppressor of Hairy-wing* loci in *Drosophila melanogaster*. Mol. Cell. Biol. **1:** 47–53.

ROMANI, S., S. CAMPUZANO and J. MODOLELL, 1987 The *achaete-scute* complex is expressed in neurogenic regions of Drosophila embryos. EMBO J. **6:** 2085–2092.

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.

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

VILLARES, R., F. GONZALEZ and J. MODOLELL, 1987 Conserved domains in the protein coding regions of the transcripts of the *achaete-scute* complex of *D. melanogaster*. Tenth European Drosophila Research Conference, Abstract 260.

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

Communicating editor: M. TURELLI