Insertion–deletion variation at the *yellow-achaete-scute* region in two natural populations of *Drosophila melanogaster*

ROBIN N. BEECH* AND ANDREW J. LEIGH BROWN†

Department of Genetics, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh, EH9 3JN, Scotland

(*Received 6 May 1988 and in revised form 17 August 1988*)

Summary

We have surveyed the region of the X chromosome of *Drosophila melanogaster* which encodes the *yellow*, *achaete* and *scute* genes for restriction map variation. Two natural populations, one from North Carolina, U.S.A. and the other from southern Spain were screened for variation at about 70 restriction sites and for variation due to DNA insertion or deletion events in 120 kilobases of DNA. Mean heterozygosity per nucleotide was estimated to be 0·0024 and 15 large insertions were found in the 49 chromosomes screened. Extensive disequilibrium between polymorphic sites was found across much of the region in the North Carolina population. The frequency of large insertions, which usually correspond to transposable genetic elements, is significantly lower than has been observed in autosomal regions of the genome. This is predicted for X-linked loci by certain models of transposable element evolution, where copy number is restricted by virtue of the recessive deleterious effects of the insertions. Our results appear to support such models. The deficiency of insertions may in this case be enhanced by hitch-hiking effects arising from the high level of disequilibrium.

1. Introduction

Many of the spontaneous classical mutations in *Drosophila melanogaster* are caused by the insertion of a transposable element (Bender *et al.* 1983; Levis *et al.* 1984; Pirrotta & Brockl, 1984; Campuzano *et al.* 1985, 1986; Parkhurst & Corces, 1986). In terms of genetical analysis such mutations have been invaluable to geneticists but the importance of transposable elements relative to other sources of mutation in natural populations is still unclear. Variation due to the presence of transposable elements close to known genes appears to be a common phenomenon, not only in *Drosophila* (Langley, Montgomery & Quattlebaum, 1982; Leigh Brown, 1983; Aquadro *et al.* 1986) and yeast (Errede *et al.* 1985) but also in the DNA of prokaryotes (Kleckner, 1981). However, only one case has been described where a transposable element insertion appeared to have an effect on enzyme activity (Aquadro *et al.* 1986).

Several population genetic models of the transposition process have been developed. These assume that (1) natural selection or (2) copy number dependent regulation of transposition is responsible for maintaining an evolutionarily stable copy number distribution (Charlesworth & Charlesworth, 1983; Langley, Brookfield & Kaplan, 1983; Kaplan & Brookfield, 1983; Golding, Aquadro & Langley, 1986; Montgomery, Charlesworth & Langley, 1987). If the presence of transposable elements was generally deleterious then selection against individuals with higher numbers of elements could prevent the number rising indefinitely under the influence of replicative transposition (Charlesworth & Charlesworth, 1983). However, the effects of each individual insertion must be more than simply additive before the models can explain the observed stable copy number distributions (Brookfield, 1982; Charlesworth & Charlesworth, 1983; Charlesworth, 1985).

Most of the spontaneous mutations in *Drosophila melanogaster* are at least partly recessive (Simmons & Crow, 1977). Newly arisen P-element insertions have also been found to induce recessive deleterious fitness effects (Mackay, 1986; Fitzpatrick & Sved, 1986). It might be expected therefore, that if the number of elements was limited by natural selection, fewer elements would persist at X chromosomal sites since the action of selection on recessive alleles is more rapid if they are X-linked than if they are autosomal (Haldane, 1927; Morton, 1971). A comparison of the

---

* Present address: Department of Zoology, Biological Sciences Centre, University of Alberta, Edmonton, Alberta, T6G 2E9 Canada.
† Corresponding author.

number of transposable element insertion sites, detected by using *in situ* hybridization to *Drosophila* salivary gland chromosomes (Montgomery *et al.* 1987), indicated that the number of insertions on *X* chromosomes compared to autosomes was not significantly different in two out of three cases. There was thus no evidence of a more rapid loss of elements from the *X* due to recessive deleterious effects. With the *in situ* technique only certain specific elements for which molecular probes are available may be examined and the position of any insertion can only be determined to the nearest chromosomal band, which may contain as many as 160 kilobases (kb) of DNA (Bossy, Hall & Spierer, 1984). However, by looking at variation in the restriction map of previously cloned loci, insertion–deletion events involving any DNA sequence may be detected and mapped much more precisely. The autosomal 87A7 heat-shock and *Adh* loci have both been examined in this way (Leigh Brown, 1983; Aquadro *et al.* 1986; Beech, 1987) as have the *X*-linked *white* and *notch* genes (Langley & Aquadro, 1987; Schaeffer *et al.* 1987) and the degree of variation due to large insertion events has proved to be substantial. Cloning and subsequent analysis of several of these insertions has shown that they are indeed members of transposable element families (Leigh Brown, 1983; Aquadro *et al.* 1986).

In order to determine the fine scale distribution of insertions on *X* chromosomes of *Drosophila melanogaster*, 120 kb of DNA surrounding the *yellow, achaete-scute* loci has been examined. This region lies at the telomeric end of the *X* chromosome (1B1-2 to 1B4-5, Carramolino *et al.* 1982; Campuzano *et al.* 1985). The overall length of the region cloned allows a much larger contiguous fraction of the genome to be analysed than has been achieved before. This region is also interesting from an evolutionary point of view since the distribution of macrochaetae is a major phenotypic character difference between *Drosophila* species. The patterns displayed by different species have been interpreted as suppression of specific macrochaetae from a grid-like pattern in a manner similar to the effects of mutations within the scute region (Garcia-Bellido, 1981).

The achaete-scute complex (AS-C) of *Drosophila melanogaster* has been extensively studied genetically (e.g. Muller, 1955; Garcia-Bellido, 1979) and has been subdivided into four regions, scute-alpha (*sc-α*), scute-beta (sc-*β*), lethal of scute (*l'sc*) and achaete (*ac*). Mutations which fall within the first two classes affect the distribution of the macrochaetae, those in the last, the microchaetae. The *l'sc* mutations result in degeneration of the central nervous system (Jimenez & Campos-Ortega, 1979; White, 1980). The *yellow* locus was known to be very closely linked to the *achaete* and *scute* loci (Dubinin, Sokolov & Tiniakov, 1937). Recently, molecular genetic studies have shown that several transcripts are produced from the achaete-scute complex and the *yellow* locus (Campuzano *et al.*

1985; Villares & Cabrera, 1987; Chia *et al.* 1986). Several of these transcripts share sequence similarity with each other and they appear to be involved in the predisposition of ectodermal cells to differentiate into neuroblasts, which is the first step in the production of the sensory macrochaetae (Villares & Cabrera, 1987).

We have examined 120 kb of DNA in the AS-C region in 27 *X* chromosomes isolated from a North Carolina population and 22 from a Spanish population. We have found extensive disequilibrium between restriction site markers across this region and significantly fewer transposable element insertions than in other regions which have been surveyed. We consider these results support models invoking recessive deleterious effects associated with transposable element insertion (Charlesworth & Charlesworth, 1983; Charlesworth, 1985) rather than the more recent selection models invoking dominant effects of such insertions (Montgomery *et al.* 1987).

2. Materials and methods

(i) *Population samples*

A series of wild-caught *X* chromosomes from two population samples were perpetuated as attached X-lines as described previously (Leigh Brown & Moss, 1987). The first sample, supplied by Dr A. E. Shrimpton, was collected at a fruit market at Raleigh, North Carolina and the second, supplied by Dr J. S. Jones, was collected at Zahara de las Atunas near Cape Trafalgar, Southern Spain. DNA was prepared using males from each of 27 *NC* lines and 22 *FV* lines (derived from the North Carolina and Zahara populations respectively). For *X*-linked loci, only sequences derived from the single wild *X* chromosomes will be present in a DNA sample produced in this way.

(ii) *Molecular analysis*

Preparation of genomic DNA: DNA was prepared from flies which had been stored at −70 °C since collection. Between 200 and 500 flies were homogenized in 5–10 ml per fly of 10 mM-Tris-Cl (pH 7·5), 100 mM-NaCl, 10 mM-EDTA, 15 mM spermine and 15 mM spermidine, using a manual glass homogenizer (Dounce, Kontes Scientific Glassware). To the homogenate an equal volume of 200 mM-Tris-Cl (pH 9·0), 30 mM-EDTA 2% SDS and 1 mg per ml pronase-E (Sigma) was added and mixed gently before incubating at 37 °C for 1 h. After two extractions with phenol, one with phenol-chloroform and one with chloroform (Maniatis, Fritsch & Sambrook, 1983) the DNA was precipitated by adding 1/10 vol. of 7 M ammonium acetate and 2 vol. of ethanol then dried completely under vacuum. The pellet was resuspended in 0·2–0·5 ml TE and stored at 4 °C.

Southern analysis of DNA fragments: DNA was digested to completion with one of four restriction

enzymes (Maniatis *et al.* 1983) and the resulting fragments were separated by horizontal agarose gel electrophoresis, transferred to nitrocellulose, Gene Screen Plus (New England Nuclear) or Hybond (Amersham International) and hybridized according to the method of Southern (1975) as modified by Wahl, Stern & Stark (1979) and Johnson *et al.* (1984). Radioactive DNA probes were prepared by either the method of Rigby *et al.* (1977) or Feinberg & Vogelstein (1984). Hybridization was at 65 °C and afterwards the filters were washed at this temperature in 0·1 × SSC, 0·5 % SDS.

Cloned sequences used as probes: A series of 10 lambda clones encompassing 110 kb of genomic sequence at the AS-C region were provided by J. Modellel (Caramolino *et al.* 1982; Campuzano *et al.* 1985). From some of these, subclones into pUC8 (Vieira & Messing, 1982) were constructed. These plasmids were denoted by pASC, followed by the number of the phage from which they were derived, a single letter representing the enzyme producing the subcloned fragment and a number denoting the fragment number from the centromeric end of the lambda clone. For example pASC94R4 is a plasmid clone containing the fourth *Eco* RI fragment from the centromeric end of λsc94. R represents *Eco* RI; P, *Pst* I and B, *Bam* H1. The clones produced were pASC53R1, 31P4, 17B1, 64R3, 94R1, 94R2, 94R3, 94R4, 101R5, 101R7 and 133R1. The position of the probes used is shown in Fig. 1.

In order to characterize further the large insertions which were detected in digests of genomic DNA, a series of genomic libraries were constructed from seven lines. The procedures used were essentially as described earlier (Leigh Brown, 1983) except that lambda EMBL3 was the cloning vector (Frischauf *et al.* 1983). The libraries were screened with the closest available subclone to the insertion.
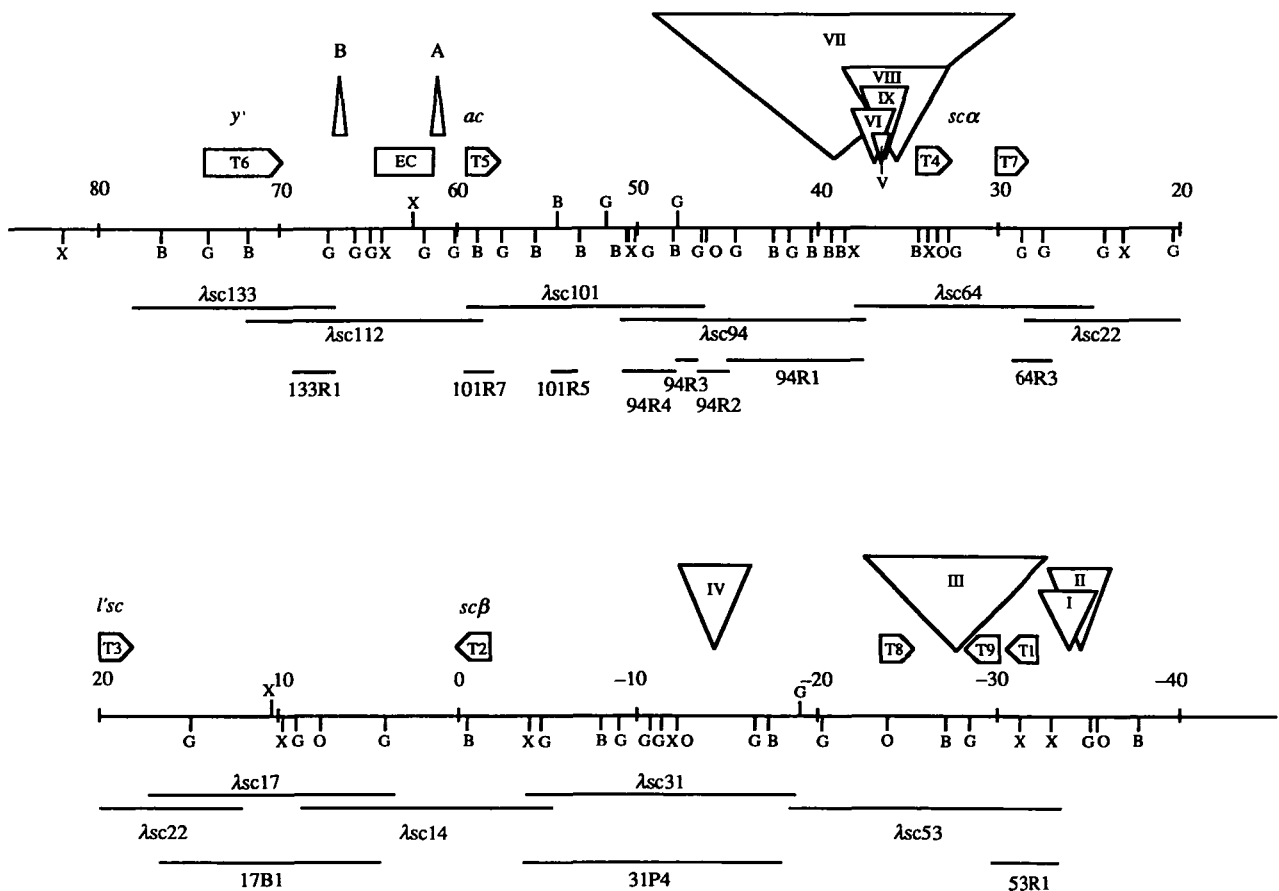

Fig. 1. Summary restriction map of the variation found at the *yellow achaete-scute* region in *Drosophila melanogaster*. Map co-ordinates (in kb) are those given by Campuzano *et al.* (1985). The series of λsc clones used to survey the region are indicated below the map and below them are shown the derived plasmid clones. (For these pASC has been omitted from the name for ease of presentation; see materials and methods). The approximate positions of the genetically defined functions in the region (Campuzano *et al.* 1985) are indicated above (*y, ac, scα, l'sc, scβ*). Known transcripts are represented by open boxes labelled T1–T9 and arrowed to show the direction of transcription (Campuzano *et al.* 1986; Villares & Cabrera, 1987). EC represents a cluster of embryonic transcripts (Chia *et al.* 1986). Restriction sites are shown below the map with polymorphic ones above. G represents *Bgl* II; B, *Bam* H1; X, *Xba* I and O, *Xho* 1. Large insertions, numbered I–IX have been shown approximately to scale above the map. A and B represent two putative insertion–deletion events of 30 and 40 bp respectively.

3. Results

(i) *Restriction map variation in the yellow-achaete-scute region*

Fig. 1 shows a molecular map of the region examined, on which are summarized the results of this survey. In the *NC* chromosomes a total of 69 restriction sites were surveyed, 67 in the *FV* chromosomes. The difference was due to two sites which were present only once in the *NC* sample. Of these sites, six were polymorphic in the *NC* sample and four in the *FV* sample (Table 1). The proportion of polymorphic nucleotides in the *NC* and *FV* populations was estimated to be 0.0074 ± 0.003 and 0.0051 ± 0.0025 and the heterozygosity per nucleotide was 0.0024 ± 0.01 and 0.0018 ± 0.008 respectively (Ewens, Spielman & Harris, 1981; Hudson, 1982). The heterozygosity per nucleotide estimated by the method of Nei & Tajima (1983) was 0.0033 for the North Carolina population and 0.0011 for the Spanish population. Polymorphic restriction sites were found throughout the region but only the *Xba* I site at position 62 was within a region known to be transcribed (Chia *et al.* 1986).

Two variants have been interpreted as small insertions, or deletions, of 30 and 40 bp and have been shown as A and B respectively, in Fig. 1. While only fragments produced by the enzyme *Bgl* II were unusual in this region, the other three enzymes used all produced large fragments covering this part of the sequence so that insertion–deletion events smaller than 50 bp would not have been detected. There remains the possibility that both these events are due to polymorphic *Bgl* II sites not present on the map of Campuzano *et al.* (1985). Further work will be required to resolve this. These two events lie very close to the boundaries of a group of embryonic transcripts (Chia *et al.* 1986). 'A' would also appear to be within 500 bp 5' to the transcript T5.

Any variation due to the insertion, or deletion, of DNA larger than 50–100 bp should have been detected throughout the region between the *Xba* I site at position 81 and the *Bam* H1 site at position -37, a distance of about 120 kb. A total of about 15 such insertions was observed. Insertion IV was apparently present six times in the *NC* sample and one of these chromosomes (*NC*22) also contained insertion II at a distance of approximately 20 kb. Two chromosomes appeared to have insertion III, both from the *NC* population. It is not clear at present whether these sets of insertions IV and III were in fact identical by descent or represent multiple insertion events. Variability due to the presence of large insertions may be represented as the mean number of differences between two randomly chosen chromosomes (Leigh Brown, 1983). The values for the *NC* and *FV* samples, corrected for the length of sequence examined were 6.89×10^{-3} per kilobase and 2.27×10^{-3} per kilobase. These values are approximately a thousand-fold lower than the estimated heterozygosity for nucleotide substitutions. On this basis, if the variation was distributed uniformly over the DNA, there would be a 95% chance of observing at least one nucleotide difference between two DNA segments 400 bp long and one large insertion difference for DNA segments 435 kb long. However, the distribution of insertions appears to be highly non-uniform. Insertion IV lies about 10 kb 5' to transcript T8 and insertion VII lies about 6 kb 5' to transcript T4. Insertions V, VI, VIII and XI all lie within about 3 kb 5' to T4 (a maximum of 5 kb). Insertions I and II lie within 3 kb of T1 and insertion II lies, at most 500 bp 3' to T9 but could reside within the coding region. Thus most of the sites of large insertions cluster into two groups, either about 3 kb 5' to T4 or within 3 kb of T1/T9. The distribution of insertions over the entire AS-C region was tested for departure from Poisson expectations by subdividing the region into twenty-four 0.5 kb segments. The observed distribution is indeed significantly different from expected ($\chi^2_3 = 8.4$; $P < 0.05$).

Recombinant bacteriophage containing at least part of three of the insertions described above were purified from genomic lambda libraries. A clone containing insertion II was isolated from DNA of line NC22 using pASC53R1 as a probe. This clone was shown to contain middle repetitive DNA sequences by hybridizing the radiolabelled clone to filters of genomic *Bgl* II digests. As different band patterns were observed in different lines we conclude that this insertion is a transposable genetic element but an exact identification has not been attempted.

The other two insertions which have been cloned are Insertions III and VII using pASC53R1 and pASC94R4 respectively as probes. In neither case did hybridisation of genomic filters give the classic pattern expected of dispersed middle repetitive sequences. Insertion 3 hybridized to three bands in genomic *Bgl* II digests which were identical in different lines. Insertion VII also gave an invariant pattern with an unusual feature in that one of the bands (a 7.3 kb *Bam* H1 fragment) was about 10 times as intense as the others. This could indicate a repetitive element normally present as part of a tandem repeat that has inserted into the AS-C. It therefore appears that the low frequency of insertions reported here is, if anything, an overestimate of the number of transposable element insertions (see below).

(ii) *Gametic disequilibrium*

We have tested for gametic disequilibrium between 5 sites where the frequency of the least common allele in the *NC* population was about 0.3 or more. These were restriction sites *Bgl* II at 48 kb on the map, *Xba* I at 11, *Bgl* II at -19, Insertion IV and deletion A. Although both alleles at all 5 sites were also present among the Zahara lines there are no significant pairwise disequilibria in the data set for this population

Table 1. *Haplotypes of the North Carolina (NC) and Zahara (FV) chromosomes examined for each of the polymorphic events revealed*

Site Position	Xba I 62.5	Insertion A	B	Bam HI 54	Bgl II 52	Bgl II 48	Xba I 11	Bgl II −19	Insertion
Line									
NC1	n	n	−	−	−	−	+	+	
NC2	n	n	−	−	−	−	+	−	
NC3	n	−	−	+	−	−	−	+	
NC4	n	n	−	−	−	+	−	−	IV
NC5	n	−	−	−	−	−	−	−	
NC6	n	+	−	−	−	+	−	−	IV
NC7	n	−	−	−	−	−	+	+	
NC8	n	+	−	−	−	+	−	−	IV
NC9	n	n	−	−	−	−	−	−	
NC10	n	n	−	−	−	−	+	+	III
NC11	n	n	−	−	−	−	+	+	
NC12	n	n	−	−	−	+	−	−	V
NC13	n	n	−	−	−	−	n	+	
NC14	n	n	−	−	−	−	+	+	
NC15	+	+	−	−	−	+	−	−	IV
NC16	−	−	−	−	−	+	−	−	
NC17	+	−	−	−	−	−	−	+	
NC18	+	−	−	−	−	−	+	+	III
NC19	+	+	−	−	−	+	−	−	IV
NC20	+	+	−	−	+	+	−	−	I
NC21	+	−	−	−	−	−	−	+	
NC22	+	−	−	−	−	−	−	−	II, IV
NC23	+	−	−	−	−	−	+	+	
NC24	+	n	n	−	n	n	−	n	
NC25	−	−	−	−	−	−	−	−	VI
NC26	n	−	−	−	−	+	n	−	
NC27	−	n	−	−	−	+	−	−	
FV1	n	+	−	n	−	+	−	+	
FV2	+	−	−	n	−	−	−	+	VII
FV3	+	−	−	−	−	−	−	+	
FV4	+	−	−	−	−	−	+	+	
FV5	+	−	−	−	−	−	−	+	
FV6	+	−	−	−	−	−	−	+	
FV7	+	−	−	−	−	−	+	+	
FV8	+	−	−	−	−	−	−	+	
FV9	+	−	−	−	−	−	−	+	
FV10	+	−	+	−	−	−	−	+	
FV11	+	−	−	−	−	−	−	+	
FV12	+	−	−	−	−	−	−	+	
FV13	+	−	−	n	−	−	−	+	VIII
FV14	+	−	−	n	−	−	−	+	IX
FV15	+	−	−	n	−	−	−	+	
FV16	+	−	−	n	−	−	−	+	
FV17	+	−	−	n	−	−	−	+	
FV18	+	−	+	n	−	−	−	+	
FV19	+	−	−	n	−	−	−	+	
FV20	+	−	−	n	−	−	+	+	
FV21	+	−	−	n	−	−	−	+	
FV22	−	−	−	n	−	+	−	−	

I–IX represent the large insertions found in the survey and A and B, the putative small insertion–deletion events.
+ Signifies the presence of a restriction site, or a small deletion and −, its absence. n indicates that the chromosome was not scored for that site.

sample. The lower level of variation and smaller sample size will both prejudice the sensitivity of such tests (Brown, 1975). Significant disequilibria were found for 6 out of the 10 pairwise comparisons in the NC population sample. The values of r^2 (Hill, 1974) and $D' = D/D_{max}$ (Lewontin, 1964) for these com-parisons are presented in Table 2. Absolute dis-equilibrium ($D' = \pm 1\cdot0$) was observed in seven cases but is not necessarily significant by Fisher's exact test. The distance in kilobases between each pair of sites is given in parentheses. Remarkably high values of r^2 (and D') are found even at distances of over 50 kb.

Table 2. *Gametic disequilibrium in the achaete-scute complex*

	Insertion A	*Bgl* II 48	*Xba* I 11	*Bgl* II −19	Insertion IV
Insertion A	—	0·58	0·13	0·27	0·50
Bgl II (48)	1·0 (13 kb)	—	0·30	0·46	0·26
Xba I (11)	−1·0* (50 kb)	−1·0* (37 kb)	—	0·43	0·15
Bgl II (−19)	−1·0	−1·0	0·79*	—	0·22
Insertion IV	0·71* (76 kb)	0·73* (63 kb)	−1·0 (26 kb)	−1·0 (4 kb)	—

Values of r^2 (above diagonal Hill, 1974) and D' (below diagonal Lewontin, 1964) are shown for all comparisons where $P = q > 0\cdot3$. *$P < 0\cdot5$ (Fisher's exact test).

Table 3. *Summary of results of restriction map surveys at five gene regions in* Drosophila melanogaster *showing the number of large insertions found, the number of chromosomes screened and the length of DNA surveyed for each locus. The expected numbers of insertions were calculated as explained in the text*

Locus	3rd chromosome 87A heat-shock[a]	2nd chromosome adh[b]	X chromosome white[c]	X chromosome notch[d]	X chromosome AS-C[e]
Kb screened	25	13	45	65	120
Populations	2	4	1 (world-wide)	1 (world-wide)	2
No. chromosomes	29 + 32	23 + 6 13 + 6 (+ 1)	38	37	22 + 27
Total	61	49	38	37	49
No. insertions	4 + 6	7 + 0 3 + 1	14	3	3 + 12
Total	10	11	14	3	15
Insertions/kb/ chromosome ($\times 10^{-3}$) (mean = 7·16)	6·5	17	8·2	1·25	2·6
Expected no. of insertions	10·9	4·6**	12·2	17·2***	42·1***

Expected number calculated from the mean of the four estimates of the frequency of insertions per kilobase given in the line above.

Significant differences between expected and observed number of insertions indicated by asterisks: **$P < 0\cdot01$; ***$P < 0\cdot001$.

Data from: [a] Leigh Brown, 1983; Beech, 1987. [b] Aquadro *et al.* 1986. [c] Langley & Aquadro, 1987. [d] Schaeffer *et al.* 1988. [e] This paper.

(iii) *Frequency of insertions on the* X *chromosome and autosomes*

Table 3 gives a summary of the data on large insertions (over 0·5 kb in length) found at the *Adh* (Aquadro *et al.* 1986), 87A heat-shock (Leigh Brown, 1983, Beech, 1987), *white* (Langley & Aquadro, 1987), *notch* (Schaeffer *et al.* 1987) and AS-C regions. As discussed above, insertions above this size in general comprise transposable genetic elements, while below that size they are likely to be insertions or deletions of single copy DNA (Leigh Brown, 1983; Aquadro *et al.* 1986). The expected number of insertions was calculated as the mean number of insertions over all loci, weighted by the length of DNA screened. The significance of the difference between observed and expected values was determined using the χ^2 test. For *Adh* the number of insertions was significantly higher than this weighted mean ($P < 0\cdot01$), for the AS-C and *notch* loci very significantly lower ($P < 0\cdot001$).

4. Discussion

120 kb of DNA around the *yellow, achaete-scute* region in *Drosophila melanogaster* has been examined with a view to determining the distribution of insertional variation at the restriction map level on the *X* chromosome. Two different types of variation between individuals were found in the restriction maps: changes involving single nucleotide substitutions and insertion–deletion events. For the first class, the proportion of polymorphic nucleotides (0·007 and 0·005) and the heterozygosity estimates (0·0024 and 0·0018) were slightly lower for the two populations examined here than the corresponding values for the *Adh* (0·014 and 0·006, Langley *et al.* 1982; 0·016 and 0·006, Kreitman, 1983; 0·027 and 0·007, Aquadro *et al.* 1986; and heat-shock (0·007 and 0·0024, Leigh Brown, 1983) loci, but not significantly so. There seems to be no difference between *X* chromosomal and autosomal loci with respect to nucleotide variability. This is consistent with the results for electrophoretic variation of proteins encoded by *X* chromosomal or autosomal loci in man, *Drosophila* and the red kangaroo (Cooper *et al.* 1979), but not with null enzyme alleles in *Drosophila* (Voelker *et al.* 1980; Langley *et al.* 1981). Standard population genetics theory predicts that recessive deleterious mutations on the *X* chromosome should be removed rapidly from the population compared to similar mutations at autosomal loci (Haldane, 1927; Morton, 1971). On this criterion, like electrophoretic variation, much variation due to nucleotide substitutions does not appear to have a deleterious effect.

Variation due to large insertions in the AS-C appears to behave differently to nucleotide variation. The numbers of insertions found here have been compared to the numbers which have been found in similar surveys at *Adh* (Aquadro *et al.* 1986), heat-shock (Leigh Brown, 1983; Beech, 1987), *notch* (Schaeffer *et al.* 1988) and *white* loci (Langley & Aquadro, 1987). In the comparison the expected number of insertions was calculated for each locus assuming that the likelihood of finding an insertion is uniform over the DNA sequence. The AS-C region had significantly fewer insertions than any other locus. This is inconsistent with the data of Montgomery *et al.* (1987) who found no significant difference in the number of elements 297, and B104 between chromosomes using the techniques of *in situ* hybridization, although they did find significantly fewer insertions of element 412 on the *X* chromosome.

Recently, restriction map surveys of two other *X*-linked loci have been published. Forty-five kilobases around the *white* gene at 3C1-3 was screened in 38 strains from natural populations of diverse origin (Langley & Aquadro, 1987). About 65 kb around the *notch* gene has been screened in 33 lines from similar sources (Schaeffer *et al.* 1987). In the *white* region, the frequency of insertions per kilobase screened was quite similar to that found around the 87A heat-shock region, and much higher than that found in the AS-C. However, in the *notch* region, a lower frequency of insertions was observed, similar to the results presented here (Table 3). Clearly, although the data presented on *achaete-scute* would support a model in which transposable element numbers were constrained by associated recessive deleterious effects, we need an explanation for the difference between the AS-C and the *white* locus. As the regions which have been screened at *white* and the AS-C are not small, the difference is unlikely to be a statistical artefact. We therefore seek a satisfactory explanation on the basis of what is known to distinguish these two regions. Two features may be relevant: the number and distribution of transcribed sequences in the AS-C and the frequency of recombination.

If insertions were distributed non-uniformly along the DNA sequence, then smaller restriction map surveys such as those of the *Adh* and heat-shock loci might give an over estimate of the number of elements per nucleotide. This would seem to be the case, rather than the reverse, since all the insertions in the AS-C region were found very close to transcripts and the DNA examined at the two autosomal loci was in the vicinity of known transcripts. The difference in number between loci might be explained in terms of this non-uniformity. The preferential insertion of transposable elements into regions of DNA which are more accessible to enzymes which interact with DNA might be a possible cause of the non-uniform distribution. At the 87A heat-shock locus, the 1·2 kb spacer region between the heat-shock coding regions has been shown to be unusually sensitive to the enzyme topoisomerase II (Udvardy *et al.* 1985; Rowe, Wang & Liu, 1986). This enzyme catalyses DNA strand cleavage and religation *in vivo* (Gellert, 1981). In a survey of the 87A heat-shock locus in *Drosophila* from a North Carolina population, all four of the large insertions found were located in the intergenic spacer (Leigh Brown, 1983). A second survey at the same locus in a French population found three out of six large insertions occurred within the spacer (Beech, 1987). It is possible that transposition into the regions around coding sequences could occur more frequently than elsewhere because the chromatin is more likely to adopt a favourable configuration. The implication from this discussion would be that such configurations only arise in small subsets of the AS-C region but are more widely distributed around the *white* gene.

The second unusual feature of the AS-C which is reflected in our data is the relationship between genetic and molecular distance. The recombination fraction between the *achaete* and *scute* loci has been estimated to be about $6·6 \times 10^{-3}$ cM (Dubinin *et al.* 1937), indicating a mapping function of about $1·2 \times 10^{-4}$ cM/kb. In contrast the mapping function at the *white* locus is about 2×10^{-3} cM/kb (Langley & Aquadro, 1987). Even this low figure for the AS-C is

an overestimate since stocks bearing recombination suppressors in other parts of the genome were used. These differences in recombination frequency are reflected in the incidence of linkage disequilibrium at the two loci. Significant disequilibria are found in the AS-C over intervals greater than 50 kb (Table 2) in the *NC* population while in *white* only two out of 49 comparisons reached significance, not more than would be expected as Type I errors.

The existence of significant levels of disequilibria across such distances which span several genes will have a significant impact on the evolution of this region of the chromosome. The possibility of survival of any individual insertion will depend not only on its own effect on fitness but also on fitness effects due to insertions and other mutations anywhere in the region spanning several hundred kilobases. However, should such effects occur in one direction consistently, they should also have a detectable effect on nucleotide heterozygosity. We have not detected any drastic difference between the ASC and other regions of the genome in this respect (see above) so the extent of these interactions remains hypothetical.

Our interpretation supports the earlier selection models whereby transposable elements were assumed to have recessive deleterious effects (Charlesworth & Charlesworth, 1983; Charlesworth, 1985) rather than more recent ones where dominant effects are emphasized (Montgomery *et al.* 1987). There are two predictions which can be made from this hypothesis. First, the region of the chromosome distal to the AS-C, where recombination is also restricted, should show a low average frequency of insertions. Secondly, as the effects are expected to be recessive, this should not be the case near, for example, autosomal telomeric regions. The accumulation of data on appropriate autosomal regions should permit an assessment of our conjecture.

References

Aquadro, C. H., Deese, S. F., Bland, M. M., Langley, C. H. & Laurie-Ahlberg, C. C. (1986). Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster. Genetics* **114**, 1165–1190.

Beech, R. N. (1987). Insertion–deletion variation in the DNA of three natural populations of *Drosophila melanogaster.* Unpublished Ph.D. Thesis, University of Edinburgh.

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B. & Hogness, D. S. (1983). Molecular genetics of the bithorax complex in *Drosophila melanogaster. Science* **221**, 23–29.

Bossy, B., Hall, L. M. C. & Spierer, P. (1984). Genetic activity along 315 kb of the *Drosophila* chromosome. *EMBO Journal* **3**, 2537–2541.

Brookfield, J. F. Y. (1982). Interspersed, repetitive DNA sequences are unlikely to be parasitic. *Journal of Theoretical Biology* **94**, 281–299.

Brown, A. H. D. (1975). Sample sizes required to detect linkage disequilibrium between two or three loci. *Theoretical Population Biology* **8**, 184–201.

Cabrera, C. V., Martinez-Arias, A., & Bate, M. (1987). The expression of three members of the *achaete-scute* gene complex correlates with neuroblast segregation in *Drosophila. Cell* **50**, 425–433.

Campuzano, S., Carramolino, L., Cabrera, C. V., Ruiz-Gomez, M., Villares, R., Boronat, A. & Modellel, J. (1985). Molecular genetics of the *achaete-scute* gene complex of *Drosophila melanogaster. Cell* **40**, 327–338.

Campuzano, S., Balcells, L., Villares, R., Carramolino, L., Garcia-Alonso, L. & Modellel, J. (1986). Excess function *hairy-wing* mutants caused by *gypsy* and *copia* insertions within structural genes of the *achaete-scute* locus of *Drosophila. Cell* **44**, 303–312.

Carramolino, L., Ruiz-Gomez, M., Guerrero, M. C., Campuzano, S. & Modellel, J. (1982). DNA map of mutations at the *scute* locus of *Drosophila melanogaster. EMBO Journal* **1**, 1185–1191.

Charlesworth, B. (1985). The population genetics of transposable elements. In *Population Genetics and Molecular Evolution* (ed. T. Ohta and K. Aoki), pp. 213–232. New York: Springer Verlag.

Charlesworth, B. & Charlesworth, D. (1983). The population dynamics of transposable elements. *Genetical Research* **42**, 1–27.

Chia, W., Howes, G., Martin, M., Meng, Y., Moses, K. & Tsubota, S. (1986). Molecular analysis of the yellow locus of *Drosophila melanogaster. EMBO Journal* **5**, 3597–3605.

Cooper, D. W., Johnston, P. G., Vandeberg, J. L., Maynes, G. M. & Chew, G. K. (1979). A comparison of genetic variability at X-linked and autosomal loci in kangaroos, man and *Drosophila. Genetical Research* **33**, 243–252.

Dubinin, N. P., Sokolov, N. N. & Tiniakov, G. G. (1937). Crossing over between the genes *yellow, achaete* and *scute. Drosophila Information Service* **8**, 76.

Economou-Pachnis, A., Lohse, M. A., Furano, A. V. & Tsichlis, P. N. (1985). Insertion of long interspersed repeated elements in the *Igh* (immunoglobulin heavy chain) and *Mlvi-2* (molony leukaemia virus integration 2) loci of rats. *Proceedings of the National Academy of Sciences, USA.* **82**, 2857–2861.

Errede, B., Cardillo, T. S., Sherman, F., Dubois, E., Deschamps, J. & Waine, J. M. (1980). Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. *Cell* **22**, 427–436.

Ewens, W. J., Spielman, R. S. & Harris, H. (1981). Estimation of genetic variation at the DNA level from restriction endonuclease data. *Proceedings of the National Academy of Sciences, USA.* **78**, 3748–3750.

Feinberg, A. P. & Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. *Analytical Biochemistry* **132**, 6–13.

Fitzpatrick, B. J. & Sved, J. A. (1986). High levels of fitness modifiers induced by hybrid dysgenesis in *Drosophila melanogaster. Genetical Research* **48**, 89–94.

Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983). Lambda replacement vectors carrying poly linker sequences. *Journal of Molecular Biology* **170**, 827–842.

Garcia-Bellido, A. (1979). Genetic analysis of the *achaete-scute* system of *Drosophila melanogaster. Genetics* **91**, 491–520.

Garcia-Bellido, A. (1981). From the gene to the pattern: chaeta differentiation. In *Cellular Controls in Differ-*

entiation (ed. C. W. Lloyd and D. E. Rees), pp. 281–304. New York: Academic Press.

Gellert, M. (1981). DNA topoisomerases. *Annual Review of Biochemistry* **50**, 879–910.

Golding, G. B., Aquadro, C. F. & Langley, C. H. (1986). Sequence evolution within populations under multiple types of mutation. *Proceedings of the National Academy of Sciences, USA* **83**, 427–431.

Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**, 836–844.

Hill, W. G. (1974). Estimation of linkage disequilibrium in finite populations. *Heredity* **33**, 229–239.

Hobbs, H. H., Lehrman, M. A., Yamamoto, T. & Russell, D. W. (1985). Polymorphism and evolution of *alu* sequences in the human low density lipoprotein receptor gene. *Proceedings of the National Academy of Sciences, USA* **82**, 7651–7655.

Hudson, R. R. (1982). Estimating genetic variability with restriction endonucleases. *Genetics* **100**, 711–719.

Jimenez, F. & Campos-Ortega, J. A. (1979). A region of the *Drosophila* genome necessary for CNS development *Nature* **282**, 310–312.

Johnson, D. A., Gautsch, J. W. Sportsman, R. J. & Elder, J. H. (1984). Improved technique utilizing non fat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. *Genetical Analysis Techniques* **1**, 3–8.

Kaplan, N. & Brookfield, J. F. Y. (1983). Transposable elements in Mendelian populations. III. Statistical results. *Genetics* **104**, 485–495.

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

Kleckner, N. (1981). Transposable elements in prokaryotes. *Annual Review of Genetics* **15**, 341–404.

Langley, C. H. & Aquadro, C. F. (1987). Restriction map variation in natural populations of *Drosophila melanogaster*: *white* locus region. *Molecular Biology and Evolution* **4**, 651–663.

Langley, C. H., Brookfield, J. F. Y. & Kaplan, N. (1983). Transposable elements in Mendelian populations. I. A theory. *Genetics* **104**, 457–471.

Langley, C. H., Montgomery, E. & Quattlebaum, W. E. (1982). Restriction map variation in the *adh* region of *Drosophila*. *Proceedings of the National Academy of Sciences, USA* **79**, 5631–5635.

Langley, C. H., Voelker, R. A., Leigh Brown, A. J., Ohnishi, S., Dickson, B. & Montgomery, E. (1981). Null allele frequencies at allozyme loci in natural populations of *Drosophila melanogaster*. *Genetics* **99**, 151–156.

Leigh Brown, A. J. (1983). Variation at the 87A heat shock locus in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA* **80**, 5350–5354.

Leigh Brown, A. J. & Moss, J. E. (1987). Transposition of the I element and *copia* in a natural population of *Drosophila melanogaster*. *Genetical Research* **49**, 121–128.

Levis, R., O'Hare, K. & Rubin, G. M. (1984). Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila*. *Cell* **38**, 471–481.

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

Mackay, T. F. C. (1986). Transposable element induced fitness mutations in *Drosophila melanogaster*. *Genetical Research* **48**, 77–87.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1983). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Laboratory. New York: Cold Spring Harbour.

Montgomery, E. A., Charlesworth, B. & Langley, C. H. (1987). A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genetical Research* **49**, 31–41.

Morton, N. E. (1971). Population genetics and disease control. *Social Biology* **18**, 243–251.

Muller, H. J. (1955). On the relation between chromosome changes and gene mutations. *Brookhaven Symposia* **8**, 126–147.

Nei, M. & Tajima, F. (1983). Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. *Genetics* **105**, 207–217.

Parkhurst, S. M. & Corces, V. G. (1986). Interactions among the gypsy transposable element and the *yellow* and *suppressor of Hairy-wing* loci in *Drosophila melanogaster*. *Molecular and Cellular Biology* **6**, 47–53.

Pirrotta, V. & Brockl, C. (1984). Transcription of the Drosophila *white* locus and some of its mutants. *EMBO Journal* **3**, 563–568.

Rigby, P. J., Breckmann, M., Rhodes, C. & Berg, P. (1977). Labelling deoxyribonucleic acid *in vitro* by nick translation with DNA polymerase I. *Journal of Molecular Biology* **113**, 237–251.

Rowe, T. C., Wang, J. C. & Liu, L. F. (1986). *In vivo* localization of DNA topoisomerase II cleavage sites in *Drosophila melanogaster*. *Molecular Biology and Evolution* **5**, 30–40.

Schaeffer, S. W., Aquadro, C. F. & Langley, C. H. (1988). Restriction map variation in the *Notch* region of *Drosophila melanogaster*. *Molecular Biology and Evolution* **5**, 30–40.

Simmons, M. J. & Crow, J. F. (1977). Mutations affecting fitness in *Drosophila* populations. *Annual Review of Genetics* **11**, 49–78.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

Udvardy, A., Schedl, P., Saunder, M. & Hsieh, T.-S. (1985). Novel partitioning of DNA cleavage sites for *Drosophila* topoisomerase II. *Cell* **40**, 933–941.

Villares, R. & Cabrera, C. V. (1987). The *achaete-scute* gene complex of *Drosophila melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* **50**, 415–424.

Vieira, J. & Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with universal primers. *Gene* **19**, 259–268.

Voelker, R. A., Langley, C. H., Leigh Brown, A. J., Ohnishi, S., Dickson, B., Montgomery, E. & Smith, S. C. (1980). Enzyme null alleles in a natural population of *Drosophila melanogaster*: frequencies in a North Carolina population. *Proceedings of the National Academy of Sciences, USA* **77**, 1091–1095.

Wahl, G. M., Stern, M. & Stark, G. R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization using dextran sulphate. *Proceedings of the National Academy of Sciences, USA* **76**, 3683–3687.

White, K. (1980). Defective neural development in *Drosophila melanogaster* embryos deficient for the tip of the X chromosome. *Developmental Biology* **80**, 332–344.