Genet. Res., Camb. (1992), **60**, *pp.* 115–130 *With 2 text-figures*

The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. II. Inferences on the nature of selection against elements

BRIAN CHARLESWORTH*, ANGELA LAPID AND DARLENE CANADA
Department of Ecology and Evolution, The University of Chicago, 1101 E. 57th St, Chicago IL 60637-1573, USA

(*Received 28 January 1992 and in revised form 16 June 1992*)

Summary

Data were collected on the distribution of nine families of transposable elements among a sample of autosomes isolated from a natural population of *Drosophila melanogaster*, by means of *in situ* hybridization of biotinylated probes to polytene chromosomes. There is no general tendency for elements to accumulate at the tips of chromosomes. Elements tend to be present in excess of random expectation in the euchromatin proximal to the centromeres of the major autosomes, and on chromosome four. There is considerable heterogeneity between different families in the extent of this excess. The overall abundance of element families is inversely related to the extent to which they accumulate proximally. The level of proximal accumulation for the major autosomes is similar to that on the fourth chromosome, but less than that for the X chromosome. There is an overall deficiency of elements in the mid-section of the X compared with the mid-sections of the major autosomes, with considerable heterogeneity between families. The magnitude of this deficiency is positively related to the extent to which elements accumulate proximally. No such deficiency is seen if the proximal regions of the X and autosomes are compared. There is a small and non-significant excess of elements in third chromosomes carrying inversions. There is some between-year heterogeneity in element abundance. The implications of these findings are discussed, and it is concluded that they generally support the hypothesis that transposable element abundance is regulated primarily by the deleterious fitness consequences of meiotic ectopic exchange between elements. If this is the case, such exchange must be very infrequent in the proximal euchromatin, and the elements detected in population surveys of this kind must be inserted into sites where they have negligible mutational effects on fitness.

1. Introduction

The accompanying paper showed that transposable elements (TEs) on autosomes from a population of *Drosophila melanogaster* are generally distributed randomly along the distal sections of the euchromatic portions of the chromosome arms (Charlesworth, Lapid & Canada, 1992). Furthermore, element frequencies at individual sites are almost always very low in this region of the genome. The data suggest that there is little site-specificity with respect to insertion probability, at least at the level of salivary chromosome bands. The pattern of distribution of element frequencies indicates that any increase in element abundance due to transposition must be opposed by a force or forces whose magnitude is much greater than that of genetic drift. These findings are in general agreement with the results of other studies of *Drosophila* transposable elements (Charlesworth & Langley, 1989).

The nature of the forces responsible for the containment of the spread of TEs in natural populations is, of course, not revealed by such distributional data. Three main possibilities have been discussed in the literature. The first is a reduction of the rate of transposition in response to increased copy number of elements within the host genome; there is evidence for such self-regulation of transposition probability for some classes of elements in a variety of species (Charlesworth & Langley, 1986). This can lead to stabilization of copy numbers in a large population as a result of a dynamic equilibrium between transposition and excision (Charlesworth & Charlesworth, 1983). The second is selection on transposable elements through the deleterious mutational effects of the insertion of elements into or near genes, although it is hard to account quantitatively for the *Drosophila* distributional data on this model (Charlesworth,

* Corresponding author.

1991). The last possibility is that ectopic crossing over between homologous elements located at different chromosomal sites leads to the production of deleterious chromosome rearrangements. This will have the effect of reducing the fitness of individuals with large numbers of elements belonging to a given family (Langley *et al.* 1988; Charlesworth & Langley, 1989; Montgomery *et al.* 1991).

In this paper, we attempt to discriminate between these alternative models using data on the distribution of the numbers of elements belonging to nine different families. We analyze the distribution of these elements within different sections of the major autosomes, between the autosomes and sex chromosomes, and between the fourth chromosome and the major autosomes. The general conclusion is that ectopic exchange probably plays a major role in controlling the abundances of these elements, but that it is not the only force involved.

2. Materials and methods

(i) *Genetic stocks and breeding procedures*

Sets of second and third chromosomes were isolated from a population at Beltsville, Maryland, as described by Charlesworth *et al.* (1992). Fourth chromosomes were extracted by the following breeding procedure. Single wild males were mated with females homozygous for the X chromosome balancer *FM7* [which is marked with *B*, w^a and *y* (Lindsley & Zimm, 1992)], and for the fourth chromosome marker spa^{pol}. These markers had previously been introduced by repeated backcrossing onto a background of second and third chromosomes from the wild-type outbred laboratory stock *IV*, described by Charlesworth & Charlesworth (1985), in order to avoid hybrid dysgenesis on crossing the balancer stock to wild males (Charlesworth & Lapid, 1989). Single *FM7* F_1 males from each cross, heterozygous for a wild-type fourth chromosome and spa^{pol}, were crossed to balancer stock females. In the next generation, *FM7*; $+/spa^{pol}$ females and males from each cross were mated together. Several single pair-matings of spa^+ flies were then established from each resulting culture. Crosses that segregated out homozygotes for spa^{pol} were discarded. This process was repeated for several generations until a set of lines that were unequivocally homozygous for different wild-type fourth chromosomes, on a background of *FM7* X chromosomes, was established. Each stock was maintained in mass culture in vials, at 18 °C. Thirteen fourth-chromosome lines were used in this study. Ten of these were isolated in 1987, and three in 1988.

(ii) *Preparation and scoring of* in situ *slides*

Procedures for preparing and scoring slides for *in situ* hybridization of transposable elements to the fourth chromosome lines were as described previously (Charlesworth & Lapid, 1989; Charlesworth *et al.* 1992). The set of nine TE families described in Charlesworth *et al.* (1992) were used for the fourth-chromosome hybridizations. The data for the X chromosomes and major autosomes analyzed here come from the experiments of Charlesworth & Lapid (1989) and Charlesworth *et al.* (1992) respectively.

3. Results

(i) *Distribution of elements within the major autosomes*

Charlesworth & Lapid (1989) found that many of the element families used in this study tended to be disproportionately abundant at the base of the X chromosome, but found no evidence for an overabundance of elements at the tip of the X, another region of restricted crossing over. They compared the proportion of the total number of elements in the sample belonging to a given family that were found to be located in the base of the chromosome to the proportion of the DNA of the polytene X chromosome represented by this region. A similar comparison was made by Langley *et al.* (1988) for all five major chromosome arms for the element *roo* from a North Carolina population.

Tables 1 and 2 summarize the results of a similar analysis for the four major autosomal arms, using the cytogenetic criteria of Langley *et al.* (1988) to delimit the tip, mid and basal (centromere–proximal) regions of the polytene chromosomes. The proportions of the chromosome arms represented by these regions were estimated from the DNA measurements of Bolshakov, Zharkikh & Zhimulev (1985). These estimates are used as the expectations for the proportions of elements in each region, and are very similar to the estimates of Langley *et al.* (1988) derived from measurements of the lengths of the regions in the Lefevre (1976) photographic maps of the polytene chromosomes. In some cases these estimates differ somewhat from those based on the densities of mutant genes (cf. Langley *et al.* 1988, Table 5). Since the latter are subject to considerable sampling error, the physical estimates probably provide more accurate estimates of the probabilities of insertion of elements into the respective chromosome regions, on the null hypothesis of random insertion.

The numbers of elements in each region of each set of chromosome arms in the sample were calculated from the raw data. The most conservative way to view the data is to omit contributions from four sites where there was apparent fixation of elements, since these may be artefacts due to hybridization of unique flanking sequences with the sites of cloning (Charlesworth *et al.* 1992). The results for these cases with the omission of the 'fixed' sites are shown within parentheses in the tables. Omitting these sites has no

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

Table 1. *Proportions of elements found in the three sections of chromosome 2 for nine families of elements sampled from a Maryland population of* Drosophila melanogaster

2L

Element	Tip	Mid	Base	Total no. elements	Element	Tip	Mid	Base	Total no. elements
roo	0·037	0·794	0·169	136	*2217*	0·000	0·949	0·051	39
2156	0·043	0·574	0·383***	47	*297*	0·000	0·635	0·365***	52
2158	0·030	0·576	0·394***	33	*412*	0·021	0·809	0·170	47
2161	0·038	0·886	0·076	79	*copia*	0·000*	0·761	0·239*	46
2181	0·030	0·794	0·176	34	*Total*	0·027	0·766	0·207***	513
Expected	0·06	0·82	0·12			0·06	0·82	0·12	

Heterogeneity χ^2 (8 D.F.) for comparison of proportion of elements in the basal region with expectation: 62·1***

2R

Element	Tip	Mid	Base	Total no. elements	Element	Tip	Mid	Base	Total no. elements
roo	0·057	0·823	0·120	141	*2217*	0·065	0·761	0·174	46
2156	0·000	0·514	0·486***	37	*297*	0·096	0·731	0·173	52
2158	0·000	0·361	0·639***	36	*412*	0·140	0·820	0·040*	50
2161	0·042	0·875	0·083	96	*copia*	0·051	0·641	0·308*	39
2181	0·286***	0·482	0·232	56	*Total*	0·058	0·738	0·204*	539
	(0·048	0·643	0·309*	42)[a]					
Expected	0·08	0·75	0·17			0·08	0·75	0·17	

Heterogeneity χ^2 (8 D.F.) for comparison of proportion of elements in the basal region with expectation: 103·1***

[a] This omits site 60B1, where this element is apparently fixed. Such apparent fixation may be an artefact due to hybridization with flanking sequences cloned simultaneously with the element.
* $P < 0·05$; ** $P < 0·01$; *** $P < 0·001$.

effect on the overall conclusions. Significance levels for individual elements were obtained by χ^2 tests of observed versus expected numbers for the three sections of the chromosome arms, using the products of the total number of elements in the sample for the given chromosome arm and element family, and the expected frequencies at the foot of each table, as expected numbers. Accumulation at the tip of the chromosome was tested by comparing the abundances in the tip and mid sections with the conditional expectations for these regions. In view of the general lack of significance for the tip (see below), accumulation at the base was tested by pooling the tip and mid sections to form the 'distal' section of the chromosome, and calculating expected numbers for the proximal and distal regions (Charlesworth & Lapid, 1989). For each chromosome arm, the numbers for all families of elements were also pooled, and tests for deviations from random expectations performed. Heterogeneity chi-squares for the proximal versus distal comparison for each arm were also calculated by subtracting the χ^2 for the pooled data from the sum of the individual element χ^2 values. These test for differences in the deviation from random expectation among element families within arms.

There were only two cases in which a significant contribution to the overall chi-squared value at the 1% level or above could be attributed to an excess of elements at the tip of the chromosome arm in question, *2161* and *2181* on 3R (Table 2). The results for *2161* and *2181* are both due to a high frequency of these elements at single distal sites (band 100C3, frequency of *2161* of 0·69; band 100B3, frequency of *2181* of 0·38). The only other suggestion of a significant effect is an apparent deficiency of *copia* at the tip of 2L ($P < 0·05$), but this has little meaning in view of the large number of tests performed. 3R is the only chromosome with an excess at the tip over random expectation for pooled elements ($\chi^2 = 27·7$, $P < 0·001$), with a ratio of observed to expected numbers of 1·62. 2L has an apparent deficiency of pooled elements at the tip ($\chi^2 = 7·2$, $P < 0·01$). The mean value of the ratios of observed to expected numbers for the tip over all five chromosome arms, including the X chromosome data of Charlesworth & Lapid (1989), is 0·97. This suggests that there is no overall tendency for elements to deviate from random expectation as far as the tip of the chromosome is concerned.

The pattern is very different for the base. For each autosomal arm, there are always at least two element families with an excess of elements at the base at the $P < 0·01$ level (*2156* and *2158*). Some other families (e.g. *297*) show significant excesses for some arms but not others. Other families (e.g. *roo*, *2217*) fail to show a significant effect for any autosomal arm. In one case (*2161* on 3R) there is a highly significant deficiency of elements at the base. The heterogeneity tests are all highly significant, indicating that there are differences

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

Table 2. *Proportions of elements found in the three sections of chromosome* 3 *for nine families of elements sampled from a Maryland population of* Drosophila melanogaster

3L

Element	Tip	Mid	Base	Total no. elements	Element	Tip	Mid	Base	Total no. elements
roo	0·071	0·729	0·200	140	*2217*	0·024	0·762	0·214	42
2156	0·069	0·310	0·621***	29	*297*	0·078	0·660	0·302**	53
2158	0·000	0·364	0·636***	33	*412*	0·028	0·833	0·139	36
	(0·000	0·600	0·400**	20)[a]					
2161	0·036	0·714	0·250*	84	*copia*	0·078	0·706	0·216	51
2181	0·000	0·958	0·042	24	*Total*	0·052	0·687	0·244***	479
Expected	0·06	0·78	0·16			0·06	0·78	0·16	

Heterogeneity χ^2 (8 D.F.) for comparison of proportion of elements in the basal region with expectation: 48·4***

3R

Element	Tip	Mid	Base	Total no. elements	Element	Tip	Mid	Base	Total no. elements
roo	0·041	0·765	0·194	196	*2217*	0·050	0·725	0·225	40
2156	0·040	0·640	0·320**	50	*297*	0·012	0·741	0·247	85
	(0·054	0·514	0·432***	37)[b]		(0·039	0·694	0·292**	72)[c]
2158	0·000	0·280	0·770***	25	*412*	0·018	0·860	0·123	57
2161	0·162***	0·769	0·069**	130	*copia*	0·000	0·738	0·262	61
2181	0·107**	0·760	0·133	75	Total	0·065***	0·727	0·208***	693
Expected	0·04	0·79	0·17			0·04	0·79	0·17	

Heterogeneity χ^2 (8 D.F.) for comparison of proportion of elements in the basal region with expectation: 88·5***

[a] This omits site 80B, where this element is apparently fixed.
[b] This omits site 87C1-2, where this element is apparently fixed.
[c] This omits site 99E1-2, where this element is apparently fixed.
These cases of apparent fixation may be artefacts due to hybridization with flanking sequences cloned simultaneously with the element.

among families in their relative abundances at the base. Overall, each arm shows a significant excess of pooled elements at the base (although 2R only has $P < 0·05$), and the mean value of the ratio of observed to expected numbers at the base for the autosomal arms is 1·42.

(ii) *Differences between autosomes and X chromosomes with respect to the proximal concentration of elements*

The above mean value for the ratio of observed to expected numbers of elements for the autosomal bases is considerably lower than the value for the X, 3·61. In addition, the fraction of families showing significant autosomal excesses is much lower than for the X chromosome data of Charlesworth & Lapid (1989), where 7 out of the 9 families studied here show highly significant ($P < 0·01$) excesses at the base (this is true even if the site where *roo* is apparently fixed, 20A1, is excluded). At most, 3 out of 9 families show effects that are significant at this level on any one autosomal arm.

This suggests that there may be a difference between the X and autosomes with respect to the tendency for elements to accumulate in the proximal euchromatin, which might be relevant to tests of the models described in section 1. A difficulty with this comparison is that the sizes of the regions defined as the bases by Langley *et al.* (1988) vary considerably between arms, from 7% of the euchromatin for the X to 17% for 2R and 3R. The apparent greater excess at the base of the X could thus be due entirely to the use of a segment of DNA that is more proximal than for the autosomes, with a correspondingly greater accumulation of elements. In order to remedy this, the data were reanalyzed by dividing each chromosome arm into standard polytene divisions, and comparing the proportions of elements for that arm that were located in a given division with the proportion of the arm's DNA represented by the division in question, estimated by Bolshakov *et al.* (1985).

In order to test for a significant difference in the level of accumulation at the base between the X and the autosomes, further combining of polytene chromosome divisions was carried out, to produce proximal regions of approximately equal sizes as proportions of the euchromatin. The most proximal three polytene chromosome divisions of each arm were found to be the most useful for this purpose, except for 2R in which only the most proximal two divisions

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

Table 3. *Comparisons of abundances of elements in the proximal regions of the X and major autosomes for pooled data for nine families of elements sampled from a Maryland population of* Drosophila melanogaster

Arm	X[a]	2L	2R	2L	3R
Proximal no.	141	106	97	107	103
Total no.	517	513	539	479	693
% arm DNA at base	10·6	12·2	13·0	13·0	9·9
Ratio of obs. to expected nos. at base	2·57	1·69	1·38	1·87	1·50
χ^2 (1 D.F.)	54·8***	62·8***	70·1***	64·2***	68·8***

The proximal region of an arm is defined here as the most proximal three polytene chromosome divisions, except for 2R for which the two most proximal divisions were used. Sites at which elements were apparently fixed were omitted when compiling the data.

[a] The family *2210* was omitted from the X chromosome data, since it was not studied on the autosomes.

were pooled. Table 3 shows the results obtained by defining the proximal regions of the arms in this way, for the pooled element data. The mean ratio of observed to expected abundances for the autosomal arms is 1·61, and the value for the X is 2·57. The significance of this difference between the X and the autosomes can be assessed by a 2×2 contingency χ^2 test, using proximal *vs.* distal regions, and X *vs.* pooled autosomes, as rows and columns. This test is conservative, since the proportion of the X represented by the proximal region is still slightly lower than the mean for the autosomes. The result of this test is $\chi^2 = 19{\cdot}7, P < 0{\cdot}001$, indicating a highly significant bias towards accumulation of elements at the base of the X compared with the autosomes. The implications of this result are discussed in section 4.

(iii) *Analyses of occupancy profiles for the basal versus distal regions*

An alternative method of examining the concentration of elements in the bases of the chromosomes is to contrast the occupancy profiles for the distal and proximal regions, reported in Table 5 of Charlesworth & Lapid (1989) and Tables 3 and 4 of Charlesworth *et al.* (1992). Since these encapsulate information on frequencies of elements per polytene chromosome band, the results of such a comparison should be insensitive to errors in the estimates of the sizes of the regions. The contrasts were conducted using the definition of the base of section 3(i) (Langley *et al.* 1988), omitting sites where elements were apparently fixed but including the X chromosome data on *2210*.

The first question that can be asked concerns the proportions of bands that are occupied at least once in the sample of chromosomes. The means and standard errors of these over all element families and all five chromosome arms are $0{\cdot}21 \pm 0{\cdot}02$ for the distal regions, and $0{\cdot}25 \pm 0{\cdot}02$ for the bases. A paired t test gives $t = 2{\cdot}4$ (45 D.F., $P < 0{\cdot}05$ on a 2-tailed test), suggesting a borderline significant result in favour of a higher frequency of sites segregating for elements in the bases. For elements which show no significant excess at the base at the $P < 0{\cdot}01$ or higher levels, the values are $0{\cdot}25 \pm 0{\cdot}02$ and $0{\cdot}24 \pm 0{\cdot}02$ respectively, $t = 0{\cdot}42$ (27 D.F., $P > 0{\cdot}05$). For the remaining sites, which show highly significant excesses at the base, the values are $0{\cdot}15 \pm 0{\cdot}03$ and $0{\cdot}26 \pm 0{\cdot}02$, $t = 5{\cdot}0$ (17 D.F., $P < 0{\cdot}01$). Accumulation at the base of a chromosome is thus in part associated with a higher probability that a band is found to be occupied.

The second question concerns the proportion of sites which are occupied more than once, given that they are occupied at least once. The means and standard deviations for this measure of conditional multiple occupancy are $0{\cdot}23 \pm 0{\cdot}02$ for the distal regions, and $0{\cdot}49 \pm 0{\cdot}04$ for the bases, $t = 6{\cdot}1$ (45 D.F., $P < 0{\cdot}001$). For sites which show no significant excess at the base at the $P < 0{\cdot}01$ or higher levels, the values are $0{\cdot}26 \pm 0{\cdot}03$ and $0{\cdot}39 \pm 0{\cdot}04$ respectively, $t = 2{\cdot}9$ (27 D.F., $P < 0{\cdot}01$). This suggests that there is a tendency for higher element frequencies at the base, even for elements which do not yield a clear-cut result on the test above. The values for the sites with highly significant excesses at the base are $0{\cdot}18 \pm 0{\cdot}03$ and $0{\cdot}63 \pm 0{\cdot}05$, $t = 8{\cdot}1$ (17 D.F., $P < 0{\cdot}001$). High element frequencies at sites which are occupied thus contribute greatly to the excess of elements at the bases of the chromosomes.

The levels of accumulation of elements at the bases of the X and autosomes can also be compared by this method. While there is no obvious difference with respect to the frequency of occupied sites between the proximal regions of the X and autosomes (the mean for the X is 0·27 compared with the autosomal mean of 0·25), there is a borderline significant result for the

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

Table 4. *Abundances of elements on chromosome 4*

Element	Number on 4	Expected	Total number on autosomes	χ^2
roo	2	7·44	615	4·03*
2156	12	1·96	162	52·06***
2158	4	1·43	118	4·68*
2161	5	4·76	393	0·12
2181	1	2·13	176	2·16
2217	5	2·08	172	4·15*
297	11	2·90	240	22·88***
412	1	2·31	191	0·75
copia	1	2·40	198	0·82
Total	42	27·41	2266	7·85**

Heterogeneity χ^2 (8 D.F.): 84·8***.

conditional multiple occupancy frequencies (0·64 ± 0·10 for the X *vs.* 0·45 ± 0·04 for the autosomes), t = 2·2 (44 D.F., $P < 0·05$).

(iv) *Accumulation of elements on the fourth chromosome*

Another test of the possibility that elements accumulate in regions of restricted recombination is provided by comparing the abundances of elements on chromosome four with the other chromosomes. Chromosome four appears to lack meiotic exchange under normal conditions (Hochman, 1976). Given the evidence presented above for differences between the X and the major autosomes, comparisons of copy number were made between chromosome four and the major autosomes, using the thirteen fourth chromosome lines isolated as described in section 2(i).

The expected fraction of autosomal elements located on chromosome four was calculated as follows. Hochman (1976, p. 905) estimated that chromosome four represents approximately 1% of the polytene bands of a haploid set. Direct measurement of the lengths of the autosomes on the Lefevre (1976) map of the polytene chromosomes gives a slightly lower estimate of 0·88%; the more conservative value of 1% will be used here. Given that the X contributes about 20% of the total euchromatin (Langley *et al.* 1988), the fraction of the autosomal euchromatin represented by chromosome four is thus approximately 0·01/0·8 = 0·0125. In order to correct for the different sizes of the major autosomes, and the fact that 14 second chromosomes and 13 third and fourth chromosomes were sampled, the following procedure was adopted for testing for accumulation of elements on chromosome four. An estimate of the relative contributions of the major autosomal arms can be obtained by calculating a weighted mean sample size, using the sizes of each arm as the weights. The total number of bands given in Charlesworth *et al.* (1992) and the measured arm lengths of the Lefevre photographic map yield the same mean sample size for the major autosomes, 13·47. The expected proportion of autosomal elements on chromosome four, on the null hypothesis of a random distribution, is thus $13 \times 0·0125/(13 \times 0·0125 + 13·47 \times 0·9875) = 0·012$.

The results of comparing the observed numbers of elements on chromosome four with this expectation are shown in Table 4. Highly significant excesses are found for elements *2156* and *297*, and marginally significant excesses for *2158* and *2217*. A marginally significant deficiency is observed for *roo*. The excess for pooled elements is highly significant, with a ratio of observed to expected numbers of 1·50. It proved very difficult to identify the locations of elements on the fourth chromosome to the level of the bands shown in the Lefevre photographic map, so that it was not possible to determine element frequencies with any confidence. There was no indication that the high abundances of elements such as *2156*, *2158* and *297* were due to fixation at individual sites, however.

(v) *Abundances of elements in polymorphic inversions*

The ectopic exchange model predicts that polymorphic inversions in *Drosophila* populations may be associated with excess abundance of transposable elements, provided that (i) the inverted sequence is much rarer than the standard sequence and (ii) exchange is suppressed as a result of failure of pairing in inversion heterozygotes (Eanes, Wesley & Charlesworth, 1992). Data on P elements in an African population of *D. melanogaster* suggest that they are indeed more abundant in regions covered by inversions (Eanes *et al.* 1992). The chromosomes from the Maryland population studied here were examined for effects of this kind. Unfortunately, the small sample sizes, and evidence for some heterogeneity in element abundances between chromosomes sampled in different years [see section 3(vi) below], preclude any definite conclusions.

In the case of the third chromosome, there was no firm evidence for between-year heterogeneity, and so the data for all years could be pooled. One copy of 3L carried an inversion with breakpoints 66C-70F, presumed to be *In(3L)M* (Lindsley & Zimm, 1992). The mean number of elements in the region covered by the inversion in the inverted and the standard chromosomes were 9 and 7·9 ± 1·0 respectively, which are not significantly different. One copy of 3R carried an inversion with breakpoints 87A/B-96E/F, presumed to be *In(3R)K* (Lindsley & Zimm, 1992). Another carried *In(3R)P*. The mean copy numbers for the inversion regions were 27 and 14 for *In(3R)K* and *In(3R)P*, and 22·45 ± 1·4 and 12·8 ± 1·1 for the corresponding regions of the standard sequence chromosomes. None of these differences are significant, but there is a suggestion that the inverted chromosomes may carry larger numbers of elements in the region covered by the inversion than do the standard chromosomes. The mean of the ratio between

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

Table 5. *Mean numbers of copies per autosomal arm for each year of sampling for all families of elements pooled*

	1986 (3 chrs.)	1987 (11 chrs.)
Second chromosome		
2L: Whole arm	44·67 ± 2·19	34·45 ± 1·36
Mid-section	34·33 ± 0·88	26·45 ± 1·04
2R: Whole arm	45·33 ± 4·37	36·64 ± 2·19
Mid-section	31·33 ± 4·26	27·63 ± 1·84
	1987 (6 chrs.)	1988 (7 chrs.)
Third chromosome		
3L: Whole arm	26·67 ± 2·60	33·29 ± 2·87
Mid-section	23·00 ± 3·22	28·71 ± 2·63
3R: Whole arm	55·83 ± 2·27	51·14 ± 2·75
Mid-section	42·50 ± 1·67	35·86 ± 2·21

the inverted and standard chromosomes of the mean numbers of elements in the inversion region is 1·14.

(vi) *Between-year heterogeneity in element abundances*

Because of a high frequency of sterility in the initial crosses with wild males, not all chromosomes used in this study were sampled at the same time (Charlesworth *et al.* 1992). The X chromosomes were all sampled in 1986; 3 second chromosomes were sampled in 1986, and 11 in 1987; 6 third chromosomes were sampled in 1987 and 7 in 1988; 10 fourth chromosomes were sampled in 1987, and 3 in 1988.

Table 5 shows the means and standard errors of the total copy numbers for all element families pooled, for the major autosomal arms and their mid-sections for each year of sampling. The means for 2L and 2R are both greater for 1986 than for 1987. If the means for 2L and 2R for each family are pooled, 7 out of 9 families have a higher mean copy number in 1986 than 1987, with binomial probability 0·004. Overall, there is thus a strong indication that copy numbers are indeed higher overall in 1986 than 1987.

In contrast, the means for 3L and 3R vary in opposite directions across years, and the differences between years are non-significant. The chromosome three data can therefore be treated as homogeneous across years.

(vii) *Comparison of abundances of elements on the X chromosome and autosomes*

Following Montgomery *et al.* (1987) and Langley *et al.* (1988), the relative abundances of elements on the X chromosome and major autosomes can be used to make inferences concerning the mechanism of containment of TE copy numbers. In accord with the approach of Langley *et al.* (1988), only data on the mid-sections of the major chromosome arms were used, since these have comparable rates of meiotic exchange per unit physical length. This eliminates any biases due to accumulation of elements in regions of restricted exchange. The fraction of elements expected to be found on the mid-section of the X among a sample of X chromosomes and autosomes was calculated for the four different hypotheses examined by Langley *et al.* (1988): a random distribution of elements between X and autosomes, as would be expected with elimination of elements solely by excision (H_0), elimination of elements solely by selection against the deleterious effects of insertional mutations (H_1), elimination of elements solely by ectopic exchange that occurs independently of the physical locations of the elements involved (H_2), and elimination solely by ectopic exchange that occurs only between elements located in the same region (H_3). Details of the method of calculating the expectations on the various hypotheses are given by Langley *et al.* (1988).

These expected frequencies must be adjusted by correcting for unequal sizes of the arms and unequal sample sizes for different autosomes, as in section 3(iv). In this case, the relative sizes of the second and third chromosomes derived from band counts and measurements of the lengths of the mid-sections of the polytene chromosomes on the Lefevre photographic map differ slightly (1·15 *vs.* 1·12). The mean value of the two estimates of physical size was used to calculate the mean sample size for the autosomes for the sample as a whole (13·47).

The indication that the 1987 chromosomes may have lower copy numbers than those from 1986 means that the tests may be biased in favour of overestimating the copy numbers on the X compared with the autosomes, which include chromosome sampled from years in which copy numbers appear to be lower than in 1986. Without correcting for this bias, H_0 may be incorrectly accepted, and hypothesis H_1 incorrectly rejected in some instances. In order to avoid this problem, the expected values of the frequency of elements on the X *vs.* the autosomes were calculated by multiplying the expectation for the X and 3/14 of the second chromosomes by 1·21, the ratio of the 1986 and 1987 mean element abundances for the second chromosome mid-sections. This is a somewhat rough-and ready procedure, as this estimate of the change between years in element abundances is subject to a high sampling error, and there is evidence for heterogeneity between elements in the extent of their changes. The results of individual significance tests should thus be treated with caution, and firm conclusions should not be drawn from deviations that are significant only at the 5% or even 1% levels.

The results of a comparison of the observed abundances of elements on the X chromosome and the corrected expectations are shown in the right-hand portion of Table 6. Sites where elements were apparently fixed were treated as lacking elements, as explained in section 3(i). The left-hand part of the

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

Table 6. *Abundances of elements on the mid-section of the X versus the mid-sections of the major autosomes*

Element	El. freqs X	El. freqs A	Prop. on X	Total no.	χ^2 (1 D.F.) H_0	H_1	H_2	H_3
roo	0·070	0·065	0·225	614	0·68	20·57***	3·98*	0·28
2156	0·004	0·010**	0·086	78	10·37**	3·12	5·91*	8·03*
2158	0·002	0·007**	0·056	54	9·99**	4·26*[a]	6·55*	8·21**
2161	0·026	0·043***	0·140	365	19·78***	0·92	6·65*	12·54***
2181	0·007	0·018***	0·088	157	18·32***	5·35*	10·32**	14·13***
2217	0·012	0·018*	0·147	156	7·19**	0·13	2·08	4·33*
297	0·028	0·021	0·261	211	0·54	16·72***	6·20*	2·49
412	0·015	0·022	0·155	187	7·24**	0·01	1·73	4·10*
copia	0·008	0·019***	0·096	156	17·50***	4·49*[a]	9·39**	13·23***
Total	0·019	0·025***	0·169	1971	52·41***	1·99	7·01**	42·12***
Expected proportion of elements on X:					0·24	0·16	0·19	0·22
Heterogeneity χ^2 (8 D.F.):					39·21***	55·55**	52·81***	43·56***

[a] This hypothesis gives the best fit of the four.
H_0 is the hypothesis that elements are distributed at random; H_1 assumes that element abundance is limited by the fitness effects of insertional mutations; H_2 and H_3 are two different models of limitation by ectopic exchange.

table also shows the frequencies of elements per band in the mid-sections of the X and autosomes, derived from the numbers of elements sampled in the mid-sections of the arms and the corresponding numbers of bands. Differences between the X and autosomes in these element frequencies were tested by 2×2 contingency χ^2 tests, yielding the significance levels shown by the column headed A. This provides a test for a difference between X and autosomes which does not use the correction for heterogeneity between years. It is thus a more conservative test of H_0 than the previous one.

The abundance of pooled elements on the X chromosome is much less than predicted by H_0. H_2 and H_3 are also apparently rejected (although H_2 is rejected only at the 1 % level), whereas H_1 is accepted. The more conservative test based on element frequencies per band also indicates that the set of pooled elements is in significant deficit on the X chromosome. There is highly significant heterogeneity between TE families in their agreement with the various hypotheses, implying that none of the models provides an explanation of all of the data. It is interesting to note that *297* and *roo* both depart strongly from H_1, and do not deviate from the other hypotheses at a high level of significance. *412*, on the other hand, fits best with H_1 and rejects H_0 at the 1 % level. This is in good agreement with the results obtained for these TE families in the earlier studies of Montgomery *et al.* (1987) and Langley *et al.* (1988).

4. Discussion

(i) *Accumulation of elements at the bases of the chromosomes*

The results presented in sections 3(i–iii) show clearly that there is an overall tendency of elements to accumulate in the proximal euchromatin of the autosomes, in excess of the expectation on the hypothesis of random insertion. This is in agreement with what was found earlier for the X chromosome for this set of families (Charlesworth & Lapid, 1989). There is evidence that the level of this excess is much greater for the X than for the autosomes: the ratio of the observed number of all elements to that expected with random insertion is 2·57 for the X, and 1·61 for the autosomes [section 3(ii)]. This difference is highly significant. The excess at the base is due both to a larger fraction of sites where elements are found to be segregating in the sample, as well as to higher element frequencies at the segregating sites. There is also a significant accumulation of elements on chromosome four, relative to its abundance on the other autosomes [section 3(iv)]. Owing to its small size and lack of meiotic exchange, the entire euchromatic material of this chromosome is physically and genetically proximal to the centromere. There is no evidence for a significant accumulation of elements at the tips of the chromosomes [section 3(i)], apart from the isolated cases of high frequencies of *2161* and *2181* at the tip of 3R.

Any theory of the accumulation of elements in the proximal euchromatin must include an explanation of these observations. Several viable theories can be imagined; some others have been discussed and rejected by Charlesworth & Lapid (1989).

(*a*) There is preferential insertion of elements into the proximal euchromatin, e.g. due to a higher density of specific recognition sequences in this region of the genome. While there is evidence for sequence specificity of insertion sites for some element families, many families do not appear to have strong requirements of this kind (Bingham & Zachar, 1989). In *D. annanassae*, the *tom* retroviral-like element is closely related to *297* and *17.6* of *D. melanogaster* and has a similar target-site specificity for TATAT (Tanda

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

et al. 1988). This rapidly-transposing element does not appear to insert preferentially into proximal regions (Shrimpton, Montgomery & Langley, 1986). Finally, this model provides no explanation for the patterns revealed by the correlates of the heterogeneity of the degree of proximal accumulation, described in section 4(ii) below.

(*b*) There are fewer sites where insertions produce deleterious effects on fitness in the basal euchromatin than in the distal section of the chromosome, perhaps because this is a region where euchromatin grades into heterochromatin in which functional genes are less abundant (Miklos & Cotsell, 1990). The equilibrium abundance of elements would then be relatively high proximally, if element abundance is primarily controlled by a balance between transposition and selection against insertional mutations (cf. Charlesworth, 1991). The fact that the proximal region of the X shows a greater accumulation of elements relative to the distal region than is seen for the major autosomes could be explained by the stronger selection against deleterious insertional mutations in X-linked sites (Montgomery *et al.* 1987; Langley *et al.* 1988).

The difficulty with this theory is that there is no evidence that the density of functional gene loci is indeed greatly lower in the basal regions of the chromosomes. Langley *et al.* (1988) give the relative sizes of the bases as defined in Tables 1 and 2, measured both by the numbers of loci detected by mutant genes, and by their lengths on the Lefevre photographic map of polytene chromosomes (which are very similar to the DNA-based values given in Tables 1 and 2). The relative genetic sizes of the bases are very similar to the relative physical sizes, except for 3R, where the genetic size is significantly smaller than the physical size. In the case of the X, for which the genetic data are the best, and where the base is the smallest, the genetic and physical estimates agree almost perfectly. Similarly, the available molecular genetic evidence on the density of genes at the base of the X chromosome does not suggest any difference from the more distal part of the chromosome (Yamamoto *et al.* 1990). These facts strongly militate against the insertional mutation explanation, unless ancillary hypotheses such as greater selective effects of insertions into proximal genes are erected.

(*c*) Hitch-hiking effects due to periodic increases in the frequencies of favourable mutant alleles (Maynard Smith & Haigh, 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992) may influence the abundance of elements at the base of the chromosome, where the average degree of linkage between a favourable mutation that arises in this region and polymorphic elements is much greater than for more distal regions. Hitch-hiking events reduce the effective size of the population for the chromosome region concerned, due to the selected genes dragging linked sections of chromosome to high frequency as they spread through the population. As shown by Charlesworth & Charlesworth (1983), elements subject to negative selection may equilibrate at higher levels of abundance in smaller populations. Surveys of nucleotide site variation show lower levels of silent polymorphism for both extreme proximal and extreme distal loci, as expected if hitch-hiking is occurring (Begun & Aquadro, 1992), so that it is conceivable that hitch-hiking could explain the excess of elements at the bases of chromosomes. In the accompanying paper, we have presented evidence for effects of hitch-hiking events on TEs in the proximal regions of the chromosomes (Charlesworth *et al.* 1992).

The hitch-hiking model for the proximal accumulation of TEs can be examined as follows. The spread of a favourable mutation that is very closely linked to a set of segregating elements that are initially at low frequency will drive to low frequency elements that are absent from the chromosome in which the mutation occurred, and cause a large increase in the frequency of any elements that happened to be present in this chromosome. The net consequence is a reduction in the proportion of sites at which elements are expected to be segregating, and an increase in the frequencies of elements at those sites where elements remain segregating. Since the data fail to show any obvious increase in the fraction of fixed sites at the bases of the chromosomes [section 3(iii)], or on chromosome four [section 3(iv)], and also indicate that there is a higher frequency of occupied sites at the bases of the chromosomes [section 3(iii)], it would seem that the hitch-hiking model is insufficient to explain the results.

(*d*) Excision of retroviral elements with long terminal repeats (LTRs) takes place by recombination between the LTRs of an element, and is thus less frequent in regions where recombination is restricted. While several examples of such excision events have been reported (Bingham & Zachar, 1989), there are two reasons for rejecting this as a major factor in causing proximal accumulation of elements. First, experiments capable of detecting reversions of insertional mutations due to excision events suggest that they are much less frequent than ectopic exchange events (Davis, Shen & Judd, 1987; Montgomery *et al.* 1991), so that it is likely to play a much less important role. Second, the net rate of excision of members of a TE family rises only linearly with mean copy number, and hence excision processes cannot in themselves maintain a stable equilibrium (Charlesworth & Charlesworth, 1983). Since the evidence on element frequencies suggests that they are held close to an equilibrium with low frequencies per site (Charlesworth & Langley, 1989; Charlesworth *et al.* 1992), LTR recombination is unlikely to be the major factor cotaining element abundance. It could, however, contribute to the generation of several of the patterns described below.

(*e*) Proximal accumulation of elements is due to lower rates of meiotic ectopic exchange between

Table 7. *Relative proximal abundances of elements on the major chromosomes, and abundances of elements on chromosome four relative to the proximal major autosomes*

Element	X	2L	2R	3L	3R	4
roo	1·44 ± 0·60	1·43 ± 0·64	0·70 ± 0·39	1·29 ± 0·59	1·33 ± 0·56	0·24 ± 0·33
2156	21·83 ± 18·79	4·46 ± 2·66	5·16 ± 3·38	11·95 ± 9·56	6·28 ± 4·25	1·76 ± 1·08
2158	43·30 ± 53·15	4·58 ± 3·23	10·20 ± 6·93	3·98 ± 3·56	19·32 ± 17·01	0·65 ± 0·66
2161	2·10 ± 1·24	0·57 ± 0·48	0·21 ± 0·24	1·99 ± 1·01	0·08 ± 0·16	1·66 ± 1·57
2181	9·99 ± 7·22	1·49 ± 1·32	2·35 ± 1·65	0·26 ± 0·53	0·98 ± 0·77	0·40 ± 0·80
2217	2·66 ± 2·13	0·36 ± 0·51	1·15 ± 0·94	1·49 ± 1·16	1·37 ± 1·30	2·27 ± 2·20
297	2·92 ± 1·47	3·85 ± 2·18	1·21 ± 0·93	2·29 ± 1·48	2·86 ± 1·54	1·86 ± 1·20
412	4·48 ± 2·68	1·41 ± 1·07	0·28 ± 0·40	0·97 ± 0·94	0·65 ± 0·66	0·52 ± 1·06
copia	0·51 ± 1·03	2·10 ± 1·43	2·77 ± 1·90	1·83 ± 1·23	1·94 ± 1·28	0·22 ± 0·44

The values in the table under X, 2L, 2R, 3L and 3R are the ratios of proximal to mid mean copy numbers for each arm, divided by the ratio of the corresponding sizes of the regions, together with their approximate 95% confidence limits. The values under 4 are the ratio of the mean copy numbers on the fourth to the mean for the pooled proximal autosomes, divided by the ratio of their sizes.

members of the same element family in this region of the genome (Langley *et al.* 1988; Charlesworth & Lapid, 1989). Several patterns in the data appear to be consistent with the predictions of this model, and will be discussed in section 4(ii) and (iii).

(ii) *Patterns of element distribution in relation to the ectopic exchange model*

It is useful to consider first the information provided by the degree of accumulation of elements on chromosome four on the possible relative magnitudes of meiotic ectopic exchange and other forces removing elements, such as spontaneous excision, selection against insertional mutations, or mitotic ectopic exchange. It is reasonable to presume that meiotic ectopic exchange is absent on the fourth chromosome, since regular meiotic exchange is undetectable under normal circumstances (Hochman, 1976). Modifying equation (6) of Langley *et al.* (1988), the ratio of the equilibrium mean copy number per fourth chromosome, $\bar{n}_4$, to the mean copy number for a defined region j of the major autosomes, $\bar{n}_j$, is given by

$$\frac{\bar{n}_4}{\bar{n}_j} = \frac{p_4(v+s_A+\kappa_j)}{p_j(v+s_A)}, \qquad (1)$$

where v is the probability per generation of elimination of an element by excision/mitotic ectopic exchange, s_A is the average selection coefficient against an autosomal insertion, κ_j is the probability per generation of elimination of an element by meiotic ectopic exchange for region j, p_4 is the probability of insertion of an element into the fourth chromosome, and p_j is the probability of insertion of an element into region j. The ps are taken to be proportional to the physical sizes of the respective regions, so that their ratio is given by the ratios of the corresponding sizes.

For comparison with the data, it is useful to take j to be the sum of the proximal regions of the major autosomes, defined as in section 3(ii) in order to equalize the sizes of the bases across chromosomes. Using the weighting procedure described in section 3(iv), the proportion of the total euchromatin in the proximal regions of the major autosomes in the sample is estimated to be 0·12. Taking 0·0125 as the fraction of the autosomal euchromatin represented by chromosome four [section 3(iv)], we have $p_4/p_j = 0{\cdot}104$. Hence, $(v+s_A+\kappa_j)/(v+s_A)$ for each element family can be estimated by dividing $\bar{n}_4/\bar{n}_j$ by 0·104. An approximate 95% confidence range for this ratio can be obtained by assuming that the variance of copy number is Poisson, and calculating the variance of the ratios of the two means. As shown in the right-hand column of Table 7, the individual estimates of $(v+s_A+\kappa_j)/(v+s_A)$ are subject to considerable error, but there is no indication that this ratio is ever very large relative to 1. Its mean over all families is 1.01, with 95% confidence limits ± 0·62. The data thus suggest that the rate of elimination of elements by meiotic ectopic exchange in the proximal euchromatin, κ_{AP}, occurs at a very low rate compared with the rate of elimination due to other factors. This is consistent with Ising's comparison of the physical and genetic locations of autosomal insertions of the TE element, reported by Ashburner (1989, chap. 11). This shows an almost complete lack of regular meiotic exchange in the proximal regions defined here.

Similarly, we can estimate R_X, the ratio of the mean number of elements of a given family in the base of the X chromosome to the mean for the mid-section, divided by the ratio of the appropriate p values (given in Table 4 of Charlesworth & Lapid, 1989). On the ectopic exchange model, and assuming the same selection coefficients and rates of excision in the proximal and mid sections,

$$R_X = (v+s_X+\kappa_{XM})/(v+s_X+\kappa_{XP}),$$

where s_X is the average selection coefficient against an X-chromosome insertion, and κ_{XM} and κ_{XP} are the

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

rates of elimination of elements by meiotic ectopic exchange for the mid and proximal sections of the X respectively. The estimates and their 95% confidence limits are given in Table 7. The mean value of R_X is 9·91. Two of the R_X values, those for *2158* and *copia*, are estimated with very little accuracy, due to low copy numbers in the mid section of the X. If these are omitted, the mean is reduced to 6·49, which is probably more accurate. The results suggest that, if the ectopic exchange model is correct, $\kappa_{XM} \gg \kappa_{XP}$ for most families, although (as already noted) there is considerable heterogeneity between families in the relative importance of ectopic exchange and other forces.

The same type of comparison of base to mid can be done for the autosomes, giving estimates of

$$R_A = (v+s_A+\kappa_{AM})/(v+s_A+\kappa_{AP}),$$

where κ_{AM} is the rate of elmination by meiotic ectopic exchange in the mid-section of the autosomes. The values for individual autosomal arms are given in Table 7, using the p values of Tables 1–3. A pooled autosomal value of R_A can also be obtained, using the same weighting system as for the comparison with chromosome four. The mid-sections constitute approximately 78% of the total euchromatin of the major autosomes, giving a p ratio for proximal/mid of 0·154 for the pooled autosomes. The mean of the pooled R_A values for all families is 2·57.

This analysis shows that element abundances are generally only modestly elevated on the fourth chromosome and proximal regions of the major chromosomes over the values for the mid-sections of the major chromosomes, where meiotic exchange is relatively free. This implies that meiotic ectopic exchange cannot be the only force involved in regulating element abundances, since otherwise element abundances would increase without limit in regions such as the fourth chromosome.

(ii) *Patterns in the distribution of elements within chromosomes*

Further quantitative analysis of the patterns in the data on the levels of accumulation at the bases of the X and major autosomes requires consideration of the heterogeneity between families. This reveals several patterns that appear to be consistent with the ectopic exchange model. First, if there is variation between families in the rate at which they undergo ectopic exchange, it would be expected that families with higher levels of accumulation at the base of the chromosome would equilibrate at lower abundances in the regions of the genome subject to recombination, since they would be subject to a higher rate of elimination from the population than other families (Langley *et al.* 1988).

There is evidence that the overall abundance of an element family is inversely related to the degree of its proximal accumulation, as measured by these ratios.

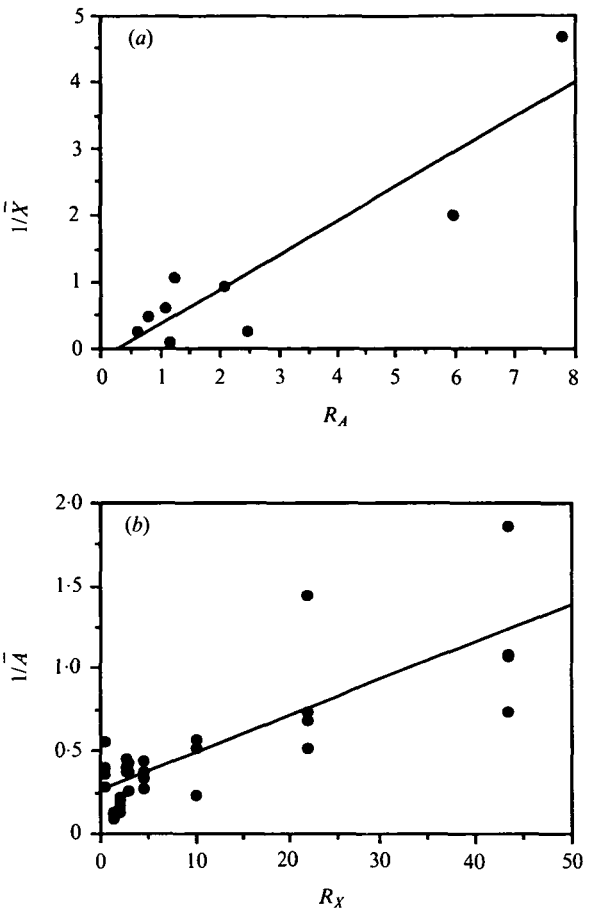

Fig. 1. The relations between the reciprocal of mean copy number of each family in the mid-section of the X and the corresponding value of R_A for the pooled autosomes (*a*), and between the reciprocal of mean element copy number in the mid-section of each major autosomal arm and the corresponding value of R_X (*b*). The lines are the regression equations ($y = 0{\cdot}51x - 0{\cdot}16$ and $y = 0{\cdot}03x + 0{\cdot}25$, respectively). The r^2 values are 0·83 (*a*) and 0·66 (*b*).

Fig. 1 *a* shows the regressions of the reciprocal of the mean copy number for each family in the mid-section of the X on the corresponding estimate of R_A for the pooled autosomes. (The comparisons between independent chromosomes avoid the introduction of spurious correlations induced by sampling error; the reciprocal of mean copy number was used in order to obtain a more linear relation.) This regression is positive and highly significant, both on the normal deviate test for the significance of the slopes, and on a jack-knife test (the jack-knife gives $t_8 = 8{\cdot}77$, $P < 0{\cdot}001$). The latter test should be robust against deviations from the usual assumptions of normal-based regression tests, such as homogeneity of error variances (Mosteller & Tukey, 1977), and hence is more trustworthy. Fig. 1 *b* shows the regression of the reciprocal of the mean copy number for each family for the mid-section of each autosome on the corresponding estimate of R_X. Omitting of the two elements with dubious R_X values (*2158* and *copia*) yields a jack-knife significance of $P < 0{\cdot}05$ ($t_6 = 2{\cdot}59$).

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

It is hard to interpret these relations quantitatively, however. Equation (5) of Langley *et al.* (1988) implies that the equilibrium abundance of a particular family in a given chromosomal region is clearly influenced by the values of rates of transposition and excision, which may vary among families, which complicates the interpretation of the relation between copy number and level of proximal accumulation. A test that is independent of these confounding parameters is provided by the relation between the measures of proximal accumulation on the autosomes and X chromosome, R_A and R_X. From the relations developed above, and assuming a lack of ectopic exchange at the bases of the chromosomes and similar rates of excision on the X and autosomes, it is easily seen that

$$\frac{(R_A-1)}{(R_X-1)} = \frac{(v+s_X)\kappa_{AM}}{(v+s_A)\kappa_{XM}}. \quad (2)$$

From the facts that (i) the rate of crossing over in the middle of the X is about 23% higher per unit physical length than for the autosomes and (ii) two-thirds of the X chromosomes are in females, and exposed to meiotic recombination, compared with only one-half of the autosomes, we would expect the ratio κ_{AM}/κ_{XM} to be approximately $3/(2\times 2\times 1{\cdot}23) = 0{\cdot}61$ (cf. Langley *et al.* 1988). In the effective absence of selective elimination of elements due to insertional mutations, as suggested by the facts discussed below, the regression coefficient of R_A on R_X is thus expected to be approximately 0·61. The jack-knifed estimate of the regression coefficient, omitting the two element families with dubious R_X values (see above), and grouping the data into sets of four for each family for different autosomes, is highly significant ($t_6 = 8{\cdot}15$, $P < 0{\cdot}001$), and the 95% confidence interval for the jack-knifed estimate of the slope nearly overlaps the expected value ($b = 0{\cdot}46 \pm 0{\cdot}14$). In addition, the expected value of R_A for $R_X = 1$ is 1; the value predicted by the jack-knifed regression equations is $0{\cdot}61 \pm 0{\cdot}61$. The slope is probably the best overall measure of the tendency for proximal accumulation to be stronger on the X than the autosomes.

Given the uncertainties in the data, this level of quantitative fit is encouraging, and the qualitative relation between R_A and R_X is in accord with the ectopic exchange hypothesis. These results suggest, however, that selection against the deleterious effects of insertional mutations must play a very minor role in regulating the abundances of the elements studied here. The expected slope was calculated on the assumption of equal rates of elimination by forces other than ectopic exchange on the X and autosomes. If there is selection against insertional mutations, the expected slope is even larger than 0·61, since s_X is expected to be greater than s_A (Langley *et al.* 1988), thus worsening the fit to the data. This conclusion about the role of insertional mutations is consistent with the observation from restriction mapping studies of small regions of the genome, which show that element insertions in samples from natural populations are nearly always located in flanking sequences rather than in genes (Charlesworth & Langley, 1989). This suggests that elements detected in small population samples are those that have survived screening by selection against insertional mutations, and are located in sites where insertions are close to neutrality. Similarly, ectopic exchange in the base of the X must be occurring at a low rate.

(iii) *Comparisons of TE abundances on the X chromosome and the autosomes*

The comparison of element abundances between the X chromosome and the autosomes in section 3(vii) was intended to help discriminate between the alternative models of containment of element abundance (Montgomery *et al.* 1987; Langley *et al.* 1988). All three classes of model other than self-regulation of transposition rate (selection against insertional mutations, ectopic exchange between elements independent of their degree of physical separation, and ectopic exchange between nearby elements) predict varying degrees of reduced abundances of elements on the X relative to the autosomes.

Both the direct contingency χ^2 tests for a difference in element frequency per band between X and autosomes, and the less conservative test for a deficiency of element abundance on the X (relative to the null hypothesis of a random distribution of elements between X and autosomes), indicate that overall there is a highly significant deficit of elements on the X (Table 6). There is, however, considerable heterogeneity among element families with respect to the magnitude or even existence of this deficit. At first sight, the results of Table 6 suggest that the model of selection against deleterious mutations, H_1, provides the best overall fit to the data. But the expectations under this model were calculated with the deliberately conservative assumption of a dominance coefficient of 0·35 for deleterious mutations of small effect (Montgomery *et al.* 1987; Langley *et al.* 1987). This is the value estimated for newly-arisen viability mutations in *D. melanogaster*, while chromosomes isolated from natural populations indicate a value of 0·2 (Crow & Simmons, 1983). The difference presumably reflects variation in the heterozygous effects of mutations, such that more recessive alleles persist longer in the population and contribute disproportionately to the estimate of dominance coefficient (Crow & Simmons, 1983). Since we are dealing with naturally occurring elements, the lower value is probably more appropriate for calculating the expected abundance of elements on the X. If the data are reanalyzed using the lower dominance coefficient (which yields an expected equilibrium proportion of elements on the X of approximately 0·10 instead of 0·13), the overall χ^2 for

Table 8. *Abundances of elements in the proximal region of the X versus the proximal regions of the autosomes*

Element	El. freqs X	El. freqs A	% on X	Total no.	χ^2 (1 D.F.)
roo	0·124	0·087	24	110	0·30
2156	0·095	0·071	23	88	0·08
2158	0·081	0·063	22	78	0·00
2161	0·067	0·031*	32	44	2·78
2181	0·081	0·060***	40	42	8·96**
2217	0·038	0·022	27	30	0·47
297	0·100	0·053*	26	80	1·07
412	0·081	0·020***	47	36	14·11***
copia	0·005	0·047**	2	46	10·18**
Total	0·75	0·048**	25	554	5·12*

Expected % on X: 21·5. Heterogeneity χ^2 (8 D.F.): 37·95***.

the deviation from H_1 in the data set corrected for between-year heterogeneity increases to 50·1 ($P < 0·001$). Given the uncertainties associated with the calculations of the expected frequencies, it would be unwise to use these results to discriminate with confidence between the insertional mutation and ectopic exchange models. It seems safe to conclude, however, that most elements are significantly less abundant on the X than expected on the hypothesis of a purely random distribution among the mid-sections of the X and the autosomes, and that this effect varies widely among element families.

This strongly suggests that some form of natural selection opposes the spread of elements, and that it operates more strongly on X-linked insertions than autosomal ones. The following analysis shows that this effect cannot be accounted for by the insertional mutation model. On this model, the degree of deficiency of elements in the proximal region of X chromosome should be similar to that for the mid-section, given their similar gene density [see section 4(i) above]. The X/autosome comparison for the proximal region is complicated by the between-year heterogeneity in abundance, leading to artefactually higher element frequencies on the X than on the autosomes. This can be corrected for, at least partially, by using the multiplier of 1·21 for the abundance of elements in 1986 relative to that in 1987 and 1988 [section 3(vii)]. The expected proportion of elements found on the proximal region of the X versus that of the autosomes under the hypothesis of a random distribution of elements is 0·215, correcting for between-year heterogeneity and for different sizes of the proximal regions of each chromosome.

Table 8 shows the results of comparing observed and expected proximal abundances for the X chromosome versus the autosomes, together with comparisons of the (uncorrected) frequencies of elements per band in the proximal regions of the X and autosomes. For all elements except *copia*, the uncorrected element frequencies per band are higher for the proximal X than for the proximal autosomes, in contrast to the pattern for the mid-sections where the frequencies are lower in most cases (Table 6). The comparisons using the corrected expected abundances mostly show agreement between observed and expected, apart from *copia* (in deficit) and *2181* and *412* (in excess). Overall, there is a highly significant excess on the X over expectation. It is difficult to judge whether this is a real phenomenon, or an artefact of an inadequate correction for the between-year effect. With the exception of *copia*, the data strongly reject the hypothesis that there is a deficiency of elements in the proximal region of the X, thus rejecting an important role for selection against insertional mutations. A similar test for agreement with the predictions of the insertional mutation model indicates a generally highly significant disagreement.

An additional test of whether or not ectopic exchange or insertional mutation is the most plausible explanation of the deficiencies of elements on the X is provided by examination of the relation between R_A for a given family and the ratio of its abundance relative to expectation in the mid-section of the autosome to that for the X chromosome. Relative abundance is measured here by the ratio of the observed mean number of elements for the X chromosome or major autosome mid-section, divided by the expected number calculated on the null hypothesis of random insertion (H_0 in Table 6). From equation (2) the autosome/X ratio for this quantity should be equal to $(1+\kappa_{XM})/(1+\kappa_{AM})$, if there is negligible selection against deleterious mutations. As we have seen, we expect $\kappa_{XM} > \kappa_{AM}$. If the ratio κ_{XM}/κ_{AM} is approximately constant across families, while the overall amount of ectopic exchange varies, then the autosome/X ratio should increase with R_A. This is indeed found to be the case for the mid-sections of the chromosomes, as shown in Fig. 2. The jack-knifed regression coefficient is $0·42 \pm 0·05$ ($P < 0·001$). This cannot be due to an artefact introduced by correlated errors in the estimates of abundances for the mid-sections of the autosomes, since these would cause a negative relation. In contrast, there is no significant relation between R_A and the autosome/X ratio for the proximal regions.

(iv) *Between-year heterogeneity in element abundances*

As discussed in section 3(vi), there is evidence for a decrease in mean copy number between 1986 and 1987. Estimates of the rate parameters of transposition, excision, etc. affecting element abundances suggest that they are normally of the order of 10^{-4} or less per generation (Charlesworth & Langley, 1989; Charlesworth *et al.* 1992). It is thus impossible that the observed changes in element abundance could be

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

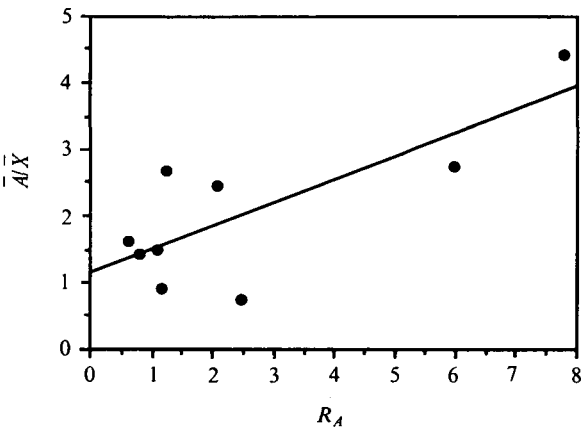

Fig. 2. The relation between R_A and the ratio of the abundance (relative to random expectation) of a family in the mid-sections of the autosomes to that for the mid-section of the X. Relative abundance is measured by the ratio of the observed mean number of elements for the X chromosome or major autosome mid-section, divided by the expected number calculated on the null hypothesis of random insertion (Table 6). The line is the regression equation $y = 0{\cdot}35x + 1{\cdot}14$. The r^2 value is 0·60.

produced as a result of shifts in the relative values of these parameters over one year, which must involve less than 20 generations in the wild. It seems far more likely that the shift was brought about by immigration of flies from a population with lower mean copy numbers. There is good evidence that East Coast *D. melanogaster* individuals are capable of travelling several kilometres in a few generations (Coyne & Milstead, 1987). There is little published information that would allow one to judge whether there is indeed sufficient between-population divergence in element abundance to support this interpretation. The data of Montgomery *et al.* (1983, 1987) on chromosomes from a North Carolina population do not show any obvious differences from our results for the elements *roo*, *297*, *412* and *copia*, but the data are not adequate to reveal the relatively small differences in abundances detected here.

(v) *The relative sizes of the autosomal arms*

Inspection of the mean numbers of all element families in each major autosomal arm (Table 5) indicates that 3R seems to have a higher overall copy number than any of the other arms. This raises the question of whether or not there is any tendency for elements to accumulate differentially on 3R, which would cast doubt on the validity of many of the tests used in previous sections. This question may best be answered by using data on the mid-sections of the chromosomes, which removes any difficulties caused by differences between arms in the extent of accumulation at the base or tip. If the second chromosomes from 1987 only are used (in order to eliminate the effect of between-year heterogeneity), and the 1987 and 1988 third chromosomes are pooled, the mean copy numbers for the autosomal mid-sections are as follows: 26·4 (2L), 27·6 (2R), 26·1 (3L) and 38·9 (3R). Comparing the mean for 3R with the mean for the next highest arm (2R) gives a t value of 3·71 (22 D.F., $P < 0{\cdot}01$), indicating that this difference is real. Inspection of the means for individual elements indicates that this pattern of a higher abundance of elements on 3R is fairly consistent across element families, suggesting that 3R presents a greater target size for insertions than the other arms. Comparison of the arm with the lowest mean (3L) and 2R gives no significant difference ($t = 0{\cdot}47$, $P > 0{\cdot}05$), suggesting that the other arms are of comparable size.

The hypothesis that this difference between arms is due simply to a larger physical size of 3R can be tested by comparing the element abundances with the length measurements of the chromosomes in the Lefevre photographic polytene chromosome maps, or with the numbers of bands per arm given by Charlesworth *et al.* (1992). The element densities per cm of polytene map for the mid-sections are 0·79 (2L), 0·86 (2R), 0·87 (3L) and 0·88 (3R), and the corresponding densities per band are 0·22, 0·22, 0·20 and 0·23. Very similar relative densities are obtained if the total number of elements per arm instead of the mid-sections are compared.

The similarities between arms in these densities of elements suggests that it is indeed likely that 3R is substantially larger than the other autosomal arms, and that this is the only reason for the greater abundance of elements on 3R. This is consistent with the larger number of mutable loci reported in Lindsley & Grell (1968) for the mid-sections of the autosomes: 3R has 75 such loci as opposed to a mean of 56·3 for the other autosomal arms (Langley *et al.* 1988). A further test of this would be provided if measures of total levels of mutability per chromosome arm were available, e.g. for recessive lethals. Unfortunately, this does not appear to be the case, although measurements of total lethal mutation rates for each autosome have been reported (Wallace, 1968). The ratio of the sizes of the third and second chromosomes is variously estimated here as 1·20 (element abundance), 1·22 (polytene maps), and 1·12 (number of bands given by Charlesworth *et al.* 1992). These estimates are consistent with each other and with the estimate of Gowen & Gray (1933) from the oogonial mitotic chromosome lengths (1·27), suggesting that the third chromosome should have a higher overall mutability than the second. Wallace's data do not indicate any difference between the second and third chromosomes in recessive lethal mutation rates, however. This may simply be due to sampling error. The results of surveys of the effects of homozygosity on viability of chromosomes extracted from natural population, summarised by Simmons & Crow (1977), indicate that the mean homozygous lethal loads for third and second chromosomes from natural populations have a relative value of 1·65; the ratio of the mean detrimental loads is 1·20. Finally, Charlesworth, Coyne & Barton (1987) noted

https://doi.org/10.1017/S0016672300030809 Published online by Cambridge University Press

that the third chromosome contributes disproportionately to the response to artificial selection for a number of different characters in *D. melanogaster*. Taken together, these data suggest that the third chromosome is larger than the second, probably by about 20%. This is close to the value of 26% recently obtained from the coverages of these chromosomes by sequences cloned into yeast artificial chromosomes (Hartl *et al.* 1992).

(vi) *Conclusions*

The results discussed above seem to rule out the hypotheses of self-regulation of transposition rate and selection against mutations induced by TE insertions, as general mechanisms for containing the spread of the elements studied in these population samples. The first of these models cannot account for the deficiency of elements on the mid-section of the X chromosome compared with the mid-sections of the autosomes; the second cannot account for the lack of such a deficiency in the proximal region of the X chromosome. Neither of them provides an explanation of the accumulation of elements in the proximal regions of the chromosomes, and on chromosome four. The model of elimination of elements as a result of the production of deleterious chromosome rearrangements by ectopic exchange between TEs of the same family successfully explains these features of the data, and several other patterns described above. The modest level of element accumulation on the chromosome four indicates, however, that meiotic ectopic exchange cannot be the only force regulating element abundance. The analysis of the level of proximal accumulation on the X compared with the autosomes suggests that it is unlikely that selection against insertional mutations is a major force acting in addition to ectopic exchange to regulate the elements that appear in our samples. It is, in fact, likely that these elements are a subset of all naturally occurring TE insertions: i.e. they are those which have entered sites where their direct fitness effects are minimal.

The major difficulty with the ectopic exchange model is that consistency with the population data appears to require that ectopic exchange occurs freely at the tip of the X chromosomes, but is effectively absent in the proximal X. This is inconsistent with the distribution of regular exchange along the X, which is strongly suppressed at the tip and is non-zero (though much reduced) proximally (Ashburner, 1989). Possible explanations for the lack of accumulation of elements at the tip of the X are discussed by Charlesworth & Lapid (1989) and Montgomery *et al.* (1991). An additional possibility is suggested by the fact that hitch-hiking events appear to have depleted nucleotide site variability at the extreme tip of the X chromosome (Begun & Aquadro, 1992). These may have perturbed the abundance of TEs at the tip from their equilibrium values, although tests for hitch-hiking events on TEs at the tip of the X similar to those described in section 4(i) yielded negative results.

This work was supported by PHS grant 1-R01-GM36405-01 and NSF grant BSR-8906059. We thank Effi Ofek, Helen Clinkenbeard and Ruby Payne for technical assistance, and Jerry Coyne and Gerald Wilkinson for collecting the wild flies used in this work. We also thank Charles Langley and William Eggleston for providing probes, and Deborah Charlesworth and Charles Langley for useful discussions and comments on the manuscript. Two anonymous reviewers made valuable suggestions for improving the paper. B.C. thanks Charles Langley and the Center for Population Biology, University of California at Davis for their hospitality during preparation of the manuscript.

References

Ashburner, M. (1989). *Drosophila. A Laboratory Handbook*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Begun, D. J. & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rate in *D. melanogaster*. *Nature* **356**, 519–520.

Bingham, P. M. & Zachar, Z. (1989). Retroposons and the FB transposon from *Drosophila melanogaster*. In *Mobile DNA* (ed. D. E. Berg and M. M. Howe), pp. 485–502. Washington, D.C. : American Society of Microbiology.

Bolshakov, V. N., Zharkikh, A. A. & Zhimulev, I. F. (1985). Intercalary heterochromatin in *Drosophila*. II. Heterochromatic features in relation to local DNA content along the polytene chromosomes of *Drosophila melanogaster*. *Chromosoma* **92**, 200–208.

Charlesworth, B. (1991). Transposable elements in natural populations with a mixture of insertion sites. *Genetical Research* **57**, 127–134.

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

Charlesworth, B. & Charlesworth, D. (1985). Genetic variation in recombination in *Drosophila*. I. Responses to selection and preliminary genetic analysis. *Heredity* **54**, 71–84.

Charlesworth, B. & Langley, C. H. (1986). The evolution of self-regulated transposition of transposable elements. *Genetics* **112**, 359–383.

Charlesworth, B. & Langley, C. H. (1989). The population genetics of *Drosophila* transposable elements. *Annual Review of Genetics* **23**, 251–287.

Charlesworth, B. & Lapid, A. (1989). A study of ten families of transposable elements on X chromosomes from a population of *Drosophila melanogaster*. *Genetical Research* **54**, 113–125.

Charlesworth, B., Lapid, A. & Canada, D. (1992). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genetical Research* **60**, 103–114.

Charlesworth, B., Coyne, J. A. & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist* **130**, 113–146.

Coyne, J. A. & Milstead, B. (1987). Long-distance migration of *Drosophila melanogaster*. 3. Dispersal of alleles from a Maryland orchard. *American Naturalist* **130**, 70–82.

Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila*. In *The Genetics and Biology of Drosophila*, vol. 3c (ed. H. L. Carson, M. Ashburner and J. N. Thomson), pp. 1–35. London: Academic Press.

Eanes, W. F., Wesley, C. & Charlesworth, B. (1992).

Accumulation of P elements in minority inversions in natural populations of *Drosophila melanogaster*. *Genetical Research* **59**, 1–9.

Gowen, J. W. & Gay, E. H. (1983). Gene number, size and kind in *Drosophila*. *Genetics* **18**, 1–31.

Hartl, D. L., Ajioka, J. W., Cai, H., Lohe, A. R., Lozovskaya, E. R., Smoller, D. A. & Duncan, I. W. (1992). Towards a *Drosophila* genome map. *Trends in Genetics* **8**, 70–75.

Hochman, B. (1976). The fourth chromosome of *Drosophila melanogaster*. In *The Genetics and Biology of Drosophila*, vol. 1b (ed. M. Ashburner and E. Novitski), pp. 903–928. London: Academic Press.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitchhiking effect' revisited. *Genetics* **123**, 887–889.

Langley, C. H., Montgomery, E. A., Hudson, R. R., Kaplan, N. L. & Charlesworth, B. (1988). On the role of unequal exchange in the containment of transposable element copy number. *Genetical Research* **49**, 31–41.

Lefevre, G. (1976). A photographic representation of the polytene chromosomes of *Drosophila melanogaster* salivary glands. In *The Genetics and Biology of Drosophila*, vol. 1a (ed. M. Ashburner and E. Novitski), pp. 31–36. London: Academic Press.

Lindsley, D. L. & Grell, E. H. (1968). *Genetic Variations of Drosophila melanogaster*. Washington: Carnegie Institution.

Lindsley, D. L. & Zimm, G. (1992). *The Genome of Drosophila melanogaster*. San Diego: Academic Press.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.

Miklos, G. L. G. & Cotsell, J. N. (1990). Chromosome structure at interfaces between major chromosomal types: α-heterochromatin and β-heterochromatin. *Bioessays* **12**, 1–7.

Montgomery, E. A. & Langley, C. H. (1983). Transposable elements in Mendelian populations. II. Distribution of three *copia*-like elements in a natural population. *Genetics* **104**, 473–483.

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

Montgomery, E. A., Huang, S.-M., Langley, C. H. & Judd, B. H. (1991). Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics* **129**, 1085–1098.

Mosteller, F. & Tukey, J. W. (1977). *Data Analysis and Regression*. Reading, Mass.: Addison-Wesley.

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

Shrimpton, A. E., Montgomery, E. A. & Langley, C. H. (1986). *Om* mutations in *Drosophila ananassae* are linked to insertions of a transposable element. *Genetics* **114**, 125–135.

Stephan, W., Wiehe, T. H. E. & Lenz, M. W. (1992). The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theoretical Population Biology* **41**, 237–254.

Tanda, S., Shrimpton, A. E., Ling-Ling, C., Itayama, H., Matsubayashi, H., Saigo, K., Tobari, Y. N. & Langley, C. H. (1988). Retrovirus-like features and site specific insertions of a transposable element, *tom*, in *Drosophila ananassae*. *Molecular and General Genetics* **214**, 405–411.

Wallace, B. (1968). Mutation rates for autosomal lethals in *Drosophila melanogaster*. *Genetics* **60**, 389–393.

Yamamoto, M.-T., Mitchelson, A., Tudor, M., O'Hare, K., Davies, J. A. & Miklos, G. L. G. (1990). Molecular and cytogenetic analysis of the heterochromatin-euchromatin junction region of the *Drosophila melanogaster* X chromosome. *Genetics* **125**, 821–832.