Genet. Res., Camb. (1989), **54**, *pp.* 113–125 *With* 1 *text-figure Printed in Great Britain*

A study of ten families of transposable elements on *X* chromosomes from a population of *Drosophila melanogaster*

BRIAN CHARLESWORTH* AND ANGELA LAPID
Department of Ecology and Evolution, The University of Chicago, 1103 E. 57th St, Chicago, IL 60637, USA

(*Received 3 November 1988 and in revised form 16 February 1989*)

Summary

Data were collected on the distribution of ten families of transposable elements among fourteen *X* chromosomes isolated from a natural population of *Drosophila melanogaster*, by means of *in situ* hybridization to polytene chromosomes. It was found that, with the exception of *roo*, the copy number per chromosome followed a Poisson distribution. There was no evidence for linkage disequilibrium, either within or between families. Some pairs of families of elements were correlated with respect to the identity of the sites that were occupied in the sample, although there was no evidence for a correlation with respect to the sites at which elements attained relatively high frequencies. Elements appeared to be distributed randomly along the distal part of the *X* chromosome. There was, however, a strong tendency for elements to accumulate at the base of the chromosome. Element frequencies per chromosome band were generally low, except at the base of the chromosome where bands in subdivisions 19E and 20A sometimes had high frequencies of occupation. These results are discussed in the light of models of the population dynamics of transposable elements. It is concluded that they provide strong evidence for the operation of a force or forces opposing transpositional increase in copy number. The accumulation of elements at the base of the chromosome is consistent with the idea that unequal exchange between elements at non-homologous sites is such a force, although other possibilities cannot be excluded at present. The data suggest that the rate of transposition per element per generation is of the order of 10^{-4}, for the elements included in this study.

1. Introduction

There has been significant progress in both empirical and theoretical population studies of transposable elements (TEs) (Charlesworth, 1985, 1988; Brookfield, 1986; Engels, 1986; Hartl *et al.* 1986). These studies yield the conclusion that there is nothing that we know at present that is inconsistent with the view that TEs are maintained in populations as a result of transpositional increase in copy number, and that their spread is checked by one or more opposing forces. In other words, the concept that they are essentially intra-genomic parasites (Doolittle & Sapienza, 1980; Orgel & Crick, 1980) is supported. The empirical basis for this comes from studies of the frequencies of TEs at individual chromosomal sites into which they can insert (occupable sites). In *Drosophila*, the technique of *in situ* hybridization of probes to the polytene salivary gland chromosomes permits identification of the sites occupied by repeated DNA elements on chromosomes isolated from natural populations by standard genetic procedures. The resolution of this technique is somewhat coarse, since the locations of elements can only be determined down to the level of a salivary chromosome band. Nonetheless, a useful picture of the general properties of element frequencies can be obtained (Montgomery & Langley, 1983; Montgomery *et al.* 1987; Leigh Brown & Moss, 1987; Charlesworth, 1988; Hey, 1989). In most cases, element frequencies per band appear to be low, except for the four element families of *D. algonquin* and *D. affinis* studied by Hey (1989).

The results of these surveys can be compared with models of the population dynamics of TEs, permitting interferences to be made concerning the magnitudes of the forces affecting element frequencies (Charlesworth & Charlesworth, 1983; Langley *et al.* 1983; Kaplan & Brookfield, 1983). As a result of the interaction of genetic drift with deterministic forces affecting element abundances, a probability distribution of frequencies at occupable sites will be generated for a given family of TEs, which will tend to a stationary form over time.

* Corresponding author.

The following parameters provide a complete approximate description of the form of $\phi(x)$, the stationary probability density for the frequency of elements x at given site (Charlesworth, 1985, 1988). The probability per generation that an element in a given individual produces a new copy that inserts elsewhere is u; the probability that it spontaneously excises is v, and the selection coefficient against individuals carrying an element at a given site is s; $\tilde{n}$ is the expected value of the mean copy number per individual; m is the total number of occupable sites in a haploid genome; N_e is the effective size of a local population.

The joint effects of deterministic forces and drift are measured by the parameters

$$\alpha = 4N_e u\tilde{n}/(2m-\tilde{n}) \quad \text{and} \quad \beta = 4N_e(s+v).$$

If $\tilde{n} \ll 2m$, as is usually true in *Drosophila*, the effect of migration can be included by adding $4N_e M$ to β, where M is the frequency of immigrants into a local population. It turns out that the probability density ϕ is approximated by a beta distribution (Charlesworth & Charlesworth, 1983):

$$\phi(x) \approx \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}. \qquad (1)$$

The conclusion from the *in situ* data is that β is generally a large number, of the order of 10–40, indicating that transpositional increase in element abundance is opposed by other deterministic forces (Charlesworth, 1988, table 1).

These results leave open, however, the nature of the forces opposing transpositional increase in copy number. One possibility 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 in a variety of species (Charlesworth & Langley, 1986). This force can unquestionably lead to stabilization of copy numbers in a large population as a result of a dynamic equilibrium between transposition and excision (Charlesworth & Charlesworth, 1983), although direct evidence for its action is lacking for most *Drosophila* elements. Another possibility is selection on transposable elements through the deleterious mutational effects of the insertion of elements into or near genes. The empirical test of this model by Montgomery *et al.* (1987), by comparing the abundances of elements on *X* chromosomes sampled from a natural population with those on the autosomes, failed to show the predicted lower abundance of *X*-linked elements for two of the three element families studied.

Thus, although selection against insertional mutations may be a factor in stabilizing element frequencies in natural populations, it is unlikely to be the only force involved. Another possibility is that crossing over between homologous elements located at different chromosomal sites will lead to the production of deleterious chromosome rearrangements (Goldberg *et al.* 1983; Montgomery *et al.* 1987), and models of this process have been developed (Langley *et al.* 1988). These predict that, provided that the frequency of unequal exchange between a pair of TEs reflects the frequencies of meiotic exchange in the regions in which they are located, there should be a higher equilibrium abundance of elements in regions where crossing over is less frequent than normal. It is possible to test this prediction by examining the distribution of elements along the salivary chromosome map, since the frequency of exchange is lower near the tips and bases of *Drosophila* chromosomes than in the mid-sections (Lindsley & Grell, 1968; Lindsley & Sandler, 1977). The analysis by Langley *et al.* (1988) of data on the element *roo* showed a tendency for it to accumulate near the bases of the salivary chromosomes, particularly the *X*, in partial accordance with the models.

In view of the weak intensities of forces affecting element frequencies in natural populations, it seems clear that direct experimental tests to discriminate between these various possibilities will be difficult. The careful collection and analysis of quantitative population data should, however, continue to provide means of examining these questions, which are of fundamental significance for understanding the biological significance of TEs. The present paper describes the results of a survey of the distribution of ten families of transposable elements among a sample of *X* chromosomes isolated from a natural population of *D. melanogaster*, using the technique of *in situ* hybridization of element probes to salivary chromosomes.

2. Materials and methods

(i) *Genetic stocks and breeding procedures*

We isolated fourteen isogenic sets of independent *X* chromosomes from a population at Beltsville, Maryland. Dr Jerry Coyne kindly collected and mailed us wild-caught males in the autumn of 1986. 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, 1982)], and for the fourth chromosome marker spa^{pol}. These markers had previously been introduced by repeated backcrossing into a background of second and third chromosomes from the wild-type outbred laboratory stock *IV*, described by Charlesworth *et al.* (1985), in order to avoid hybrid dysgenesis on crossing the balancer stock to wild males. F_1 females heterozygous for *FM7* and the *X* chromosome transmitted from the wild male (+) were crossed to balancer stock males; in the next generation, $FM7/+;spa^{pol}/spa^{pol}$ and $+;spa^{pol}/spa^{pol}$ progeny were mated together to establish a stock homozygous for the original wild *X* chromosome, and marked with spa^{pol} as a precaution

https://doi.org/10.1017/S0016672300028482 Published online by Cambridge University Press

against contamination. Each stock was maintained in mass culture in vials, at 18 °C.

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

Slides of larval salivary gland chromosomes were prepared and hybridized to element probes labelled with biotinylated nucleotides. The sites of hybridization with the probes were detected by staining with diaminobenzidine and peroxidase, and chromosomes were stained with Giemsa. A modification of the protocol described by Montgomery *et al.* (1987) was used for these procedures. Slides were examined at a magnification of 680, under oil immersion. The procedure used to record the sites of hybridization of elements was to mark them on Xerox copies of Lefevre's (1976) photographic map of the salivary chromosomes. Elements were assigned to the bands shown on the photographic map. No attempt was made to assign elements to all of the bands shown in the standard Bridges' maps (Lefevre, 1976), since the hybridization procedure removes much of the detail that is visible in good conventional squash preparations. The most proximal band to which elements can be assigned is 20A3; staining is diffuse for the more proximal section of 20. In order to guard against contamination of the extracted chromosome stocks, and against errors in labelling of slides during hybridization, we scored two slides for each chromosome and element family, that had been prepared in different batches. Each of these was read independently, in such a way that the person reading the second slide had no knowledge of the result for the first slide. Any discrepancies between the two replicates were then studied by direct comparison of the two slides; irreconcilable discrepancies were resolved by reading new slides. Each wild *X* chromosome stock was scored for each of the ten elements described below. Copies of the original data sheets will be supplied on request.

(iii) *Element families used for* in situ *hybridization*

Clones of elements were supplied to us by Drs Charles Langley and William Eggleston. The following well-characterized families were selected for use in the population survey, on the basis of their having been found to give reliable staining: *copia*, *297*, *412* and *roo* (*B104*). Descriptions of these *copia*-like elements are found in Finnegan and Fawcett (1986). The remaining elements, *2156*, *2158*, *2161*, *2181*, *2210* and *2217*, are incompletely characterized, but are thought to be *copia*-like (Rubin *et al.* 1981). They have previously been used in species comparisons by Brookfield *et al.* (1984).

Table 1. *Tests of the Poisson distribution for ten families of* Drosophila melanogaster *elements for a sample of 14* X *chromosomes from a Maryland population*

Element	$\bar{n}$	V_n	$\bar{n}$	V_n	
roo	11·36**	4·40	*2210*	0·64	0·55
2156	0·50	0·58	*2217*	1·79	1·41
2158	0·21	0·18	*297*	4·43	4·88
2161	4·07	5·46	*412*	2·29	2·53
2181	1·21	0·49	*copia*	1·21	0·80

* $P < 0{\cdot}05$; ** $P < 0{\cdot}01$; *** $P <$ 0·0001.

3. Results

(i) *Distribution of elements between chromosomes*

Table 1 gives the means and variances of copy number per *X* chromosome for each family. The copy numbers shown omit the values for the base of the *X* chromosome [following Langley *et al.* (1988), this is defined here as salivary chromosome bands 18D1–20A3], since many elements show a strong differential accumulation of elements in this region, as will be discussed in section (iv) below. Differences in element frequencies between different regions of the *X* chromosome would lead to deviations from a Poisson distribution of copy number for the whole chromosome [Charlesworth & Charlesworth, 1983, eq.(4)]. It will be seen that, with the exception of *roo* and *2181*, the mean is approximately the same as the variance, consistent with the Poisson distribution expected on theoretical grounds for elements with low frequencies at each site and with no linkage disequilibrium between sites (Charlesworth & Charlesworth, 1983). The variance for *roo* is much smaller than the mean. The agreement of the distributions of copy numbers with the Poisson distribution was tested by χ^2, pooling adjacent classes as necessary to avoid low expected numbers. Only *roo* showed a significant deviation (see Table 1).

(ii) *Correlations in occupancy between adjacent sites*

One possible explanation of a reduction in variance below the Poisson expectation is a negative correlation between the frequencies of elements at different sites on the chromosome i.e. negative linkage disequilibrium (Charlesworth & Charlesworth, 1983). This was tested for in three ways. The first two of these involved pooling the sample of fourteen chromosomes, and recording for each band of the Lefevre (1976) photographic map whether or not an element is present at the site in question for a particular element family in the pooled sample. In order to maximize the amount of information, but at the same time avoiding distortion of the results from accumulation of elements at the base of the chromosome, divisions 19 and 20 were omitted from the analysis. A total of 165 bands were identified as present in this distal region of the *X* chromosome, and used as the units of observation in this analysis.

https://doi.org/10.1017/S0016672300028482 Published online by Cambridge University Press

Table 2. *Tests of the autocorrelation of occupancy between adjacent sites for ten families of* Drosophila melanogaster *elements for a sample of 14* X *chromosomes from a Maryland population*

Element	*r*	*z*	Element	*r*	*z*
roo	0·021	0·35	*2210*	0·121	1·64
2156	−0·045	0·50	*2217*	0·176	2·36*
2158	−0·019	0·16	*297*	−0·056	0·63
2161	0·339	4·69***	*412*	−0·029	0·29
2181	0·047	0·67	*copia*	0·065	0·91

* $P < 0·05$; ** $P < 0·01$; *** $P <$ 0·0001.

The first test was to calculate the autocorrelation of occupancy between adjacent bands, using a score of 0 for a band if the element is absent from the band in the whole sample, and a score of 1 if an element is present in that band in at least one chromosome. The rationale for this is that most sites have only at most one element present in the sample of chromosomes (see below), so that the strongest effect of linkage disequilibrium is likely to be a correlation between nearby sites in whether or not an element is present in the sample. Of course, factors other than linkage disequilibrium could lead to such correlations [see section 4 (i) below]. Unbiased estimates of the autocorrelations of these scores were calculated, and tested using a normal deviate test by comparison to their sample standard deviations, as described by Kendall, Stuart & Ord (1983, pp. 548–551). Clearly, caution is needed in interpreting the results of these tests, since the distribution of the autocorrelation is not known. The results are shown in Table 2. The entries labelled *r* are the estimates of the autocorrelations and those labelled *z* are the ratios of these to their sampling standard deviations. The only element to show a convincingly significant autocorrelation is *2161*; *2217* is significant at the 0·05 level, but this may not mean anything in view of the large number of tests.

An alternative method is to examine the evenness of the distribution along the chromosome of occupancy by elements belonging to a given family, assuming that each band has an equal probability of showing a score of 1 or zero. If the number of bands is infinitely large, and the probability of a score of 1 is p, then the probability P_n that the number of bands separating a pair of neighbouring elements is n is given by the geometric distribution as $P_n = p(1-p)^{n-1}$, noting that a separation of zero is not allowed. In practice, the number of bands is large but not infinite, and p has to be estimated from the fraction of bands that have a score of 1. This means that the formula is only an approximation, but it should be a sufficiently good one for present purposes given a total of 165 bands. The fit of this distribution to the data was tested by compiling the distribution of the numbers of bands separating adjacent elements for a given family, and comparing this by χ^2 to the distribution expected with the value of p estimated from the data, pooling adjacent classes as necessary to avoid small expected numbers. The only element to show a significant χ^2 value is *2161* ($\chi^2_5 = 16·9$, $P < 0·01$), in agreement with the results of the autocorrelation tests. The observed distribution for this element showed more cases of elements occupying nearby sites than expected, in accord with the positive autocorrelation found for this element.

The final test was to test directly for pairwise linkage disequilibrium between members of the same family in the sample of *X* chromosomes, by examining all possible 2×2 contingency tables formed by pairs of segregating sites. The 165 bands in the distal part of the chromosome were used as sites for this purpose, and the state of each site determined for each chromosome of the sample. The observed distribution of the product-moment correlation in element frequency between each pair of segregating sites was compared with the distribution expected on the null hypothesis of no linkage disequilibrium. The expected distribution was calculated by calculating the probabilities of all possible configurations of 2×2 tables corresponding to the marginals of the observed tables, computed from Fisher's hypergeometric formula (Fisher, 1958, p. 96), and weighting these by the numbers of occurrences of the relevant marginals. The numbers of observed and expected categories of correlation coefficients in intervals of width 0·1 were compared by χ^2 tests, pooling adjacent intervals as necessary in order to avoid low expected numbers. No significant deviations between observed and expected numbers were detected for the eight elements that were sufficiently abundant outside the base of the chromosome to make this test worthwhile (all except *2156, 2158* and *2210*); the sum of the χ^2 values for each element is 32·1 for 41 D.F. ($P > 0·05$). Fig. 1 displays the distributions of observed and expected numbers in each interval for the four most abundant element families, and it can be seen that they are in remarkably close agreement.

Of course, many of the comparisons involve pairs of very distant sites, for which linkage disequilibrium is unlikely. Three extensions of this test were therefore applied to the data. First, the analysis was performed on all neighbouring pairs of segregating sites. For the more abundant elements, such as *roo*, many of these are in fact physically close to each other on the chromosome. There was no evidence for departure from random expectation ($\chi^2_7 = 3·8$, $P > 0·05$). Second, the analysis was performed for all pairs of sites within the tip of the chromosome, where recombination is highly suppressed. Again, the data agreed well with the null hypothesis of no linkage disequilibrium, although the numbers of pairs of sites is limited. Third, the test was applied to the region close to the base of the chromosome, where recombination is also infrequent (Lindsley & Grell, 1968; Lindsley &

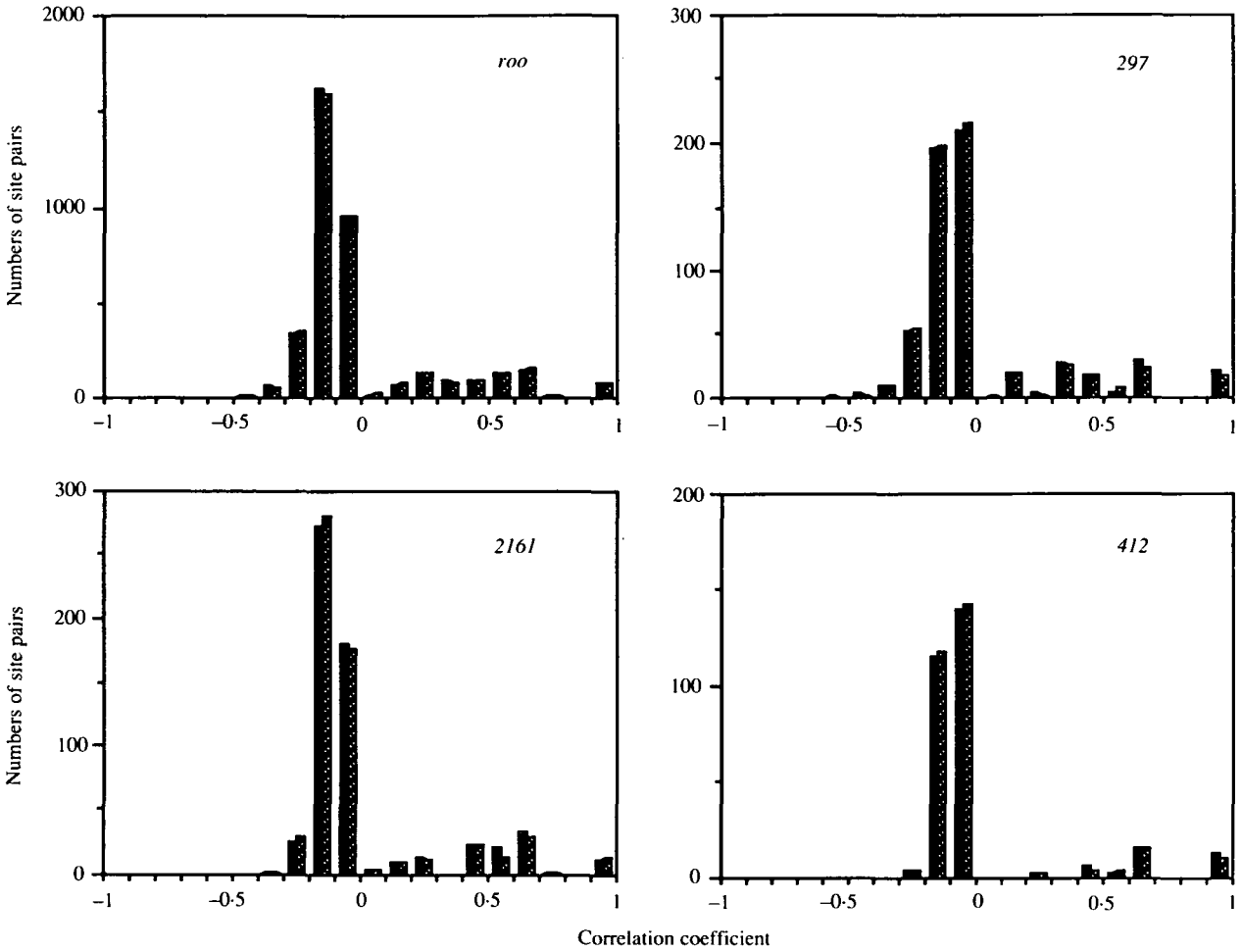

Fig. 1. The observed (full columns) and expected (hatched columns) numbers of pairs of sites with the correlation coefficients in element frequency shown on the *X* axis. The data are for the four most abundant families of elements, in the distal part of the *X* chromosome. The expected numbers are calculated on the hypothesis of no linkage disequilibrium between members of the same family at different sites, by the method described in the text.

Sandler, 1977). Salivary chromosome sub-divisions 16A–20A were used for this purpose, comprising 29 bands in total, and which will here be called the proximal region of the chromosome. The result of this analysis was again a failure to detect any deviation from randomness, despite the existence of high element frequencies for certain elements in divisions 19 and 20 (see Table 5).

(iii) *Correlations between sites occupied by different element families*

Montgomery & Langley (1983) found for a sample of *X* chromosomes from North Carolina that the elements *297* and *412* tended to co-occur at chromosomal sites more frequently than on the basis of random expectation. The present data set was tested for this type of effect in two ways. The first involved calculating the product-moment correlations in the scores for the pooled sample of *X* chromosomes, described in section 3(ii) above, for pairs of different families. This tests for a tendency for sharing between families of bands that are occupied at least once in the sample. The correlations are displayed as the entries in Table 3. Since the scores are not normal variates, the significance of the correlations was tested by Fisher's exact test [see section 3(ii)]. This procedure was used to obtain the significance levels indicated in Table 3. As before, divisions 19 and 20 were omitted.

It is clear from these results that there are significant correlations between some elements with respect to the identity of the bands that are occupied at least once in the sample. Although the number of comparisons is large (45), there is a sufficient number of highly significant results that there can be no doubt of the reality of some of the positive correlations.

The alternative method used was to perform 2×2 tests for pairs of families, treating the presence or absence of an element belonging to the two families on each chromosome in the sample as defining the rows and columns of a 2×2 table for each site where both elements are segregating. This was the procedure used by Montgomery & Langley (1983), and is formally equivalent to testing for linkage disequilibrium be-

https://doi.org/10.1017/S0016672300028482 Published online by Cambridge University Press

Table 3. *Correlation coefficients between pairs of families of* Drosophila melanogaster *elements with respect to identity of sites occupied in a sample of 14* X *chromosomes from a Maryland population*

	2156	*2158*	*2161*	*2181*	*2210*	*2217*	*297*	*412*	*copia*
roo	0·202**	0·040	0·308***	0·219**	0·142	0·184*	0·317***	0·189	0·176*
2156		0·196	0·190*	0·143	0·105	0·010	0·049	0·084	0·143
2158			0·155	0·115	−0·029	0·084	0·048	0·072	−0·043
2161				0·204*	0·190*	0·165*	0·356***	0·230***	0·204*
2181					0·143	0·195*	0·218*	−0·011	0·047
2210						−0·080	0·049	−0·087	0·143
2217							0·223*	−0·003	0·069
297								0·232**	0·058
412									0·109

* $P < 0{\cdot}05$; ** $P < 0{\cdot}01$; *** $P <$ 0·0001.

Table 4. *Proportions of elements found in the three sections of the* X *chromosome for ten families of elements sampled from a Maryland population of* Drosophila melanogaster

Element	Tip	Mid	Base	Total no. elements	Element	Tip	Mid	Base	Total no. elements
roo	0·108	0·708	0·185***	195	*2210*	0·000	0·750	0·250*	12
2156	0·000	0·259	0·741***	27	*2217*	0·066	0·767	0·167*	30
2158	0·000	0·150	0·850***	20	*297*	0·085	0·671	0·244***	82
2161	0·088	0·750	0·162**	68	*412*	0·061	0·592	0·347***	49
2181	0·125	0·406	0·469***	32	*copia*	0·111	0·833	0·056	18
Expected	0·12	0·81	0·07			0·12	0·81	0·07	

* $P < 0{\cdot}05$; ** $P < 0{\cdot}01$; *** $P <$ 0·0001.

tween pairs of families at each site where both families are segregating. Montgomery & Langley used the method of Mantel and Haenszel (Snedecor & Cochran, 1980, p. 213) for combining a set of 2 × 2 tables to test for a tendency for association that is shared between different tables. This provides a normal deviate Z that should equal zero on the null hypothesis of no association. Alternatively, the method of comparing observed and expected numbers of classes of table used above can be employed. Both of these methods were applied to the data set for the elements other than the low abundance *2156*, *2158* and *2210*, and agreed in revealing no significant associations between element families. Thus, at the level of individual chromosomes, there is no tendency for different kinds of elements to occupy the same site.

The comparison of observed and expected numbers of 2 × 2 tables can also be used to test for linkage disequilibrium between two different families of elements at different sites, on the lines of section 3(ii) for a single family. This was done for all of the above pairs of families for adjacent sites, for all pairs of sites, and for all pairs of sites within the tips and proximal regions of the chromosome. As before, there was no evidence for significant departures from random expectation.

(iv) *Distribution of elements along the* X *chromosome*

The procedure described by Langley *et al.* (1988) was used to examine the distribution of elements along the 14 Maryland *X* chromosomes (Table 4). The polytene chromosome map was divided into three sections, tip, middle and base, corresponding to the regions between band 1A1–3A4, 3A5–18C9 and 18D1–20A3, respectively. The total numbers of elements belonging to a given family that were found in these regions in the sample of 14 chromosomes were compared with the numbers expected if the numbers in each region were equal to the product of total number of elements in the sample and the proportion of polytene *X* chromosome DNA in the region, as measured by Rudkin (1965). The significance of deviations from expectation was tested by χ^2. Accumulation in the base of the chromosome was tested by pooling the tip and mid sections, and comparing observed and expected numbers in the pooled region and the base by χ^2 (1 D.F.). Accumulation in the tip of the chromosome was similarly tested by pooling the base and mid sections. Out of the 10 families, 7 (*roo*, *297*, *412*, *2156*, *2158*, *2161* and *2181*) showed a strongly significant excess of elements at the base of the X chromosome. In two other cases (*2210* and *2217*), there was a weakly significant deviation from random expectation in the

https://doi.org/10.1017/S0016672300028482 Published online by Cambridge University Press

Table 5. *Occupancy profiles for ten families of elements among fourteen* X *chromosomes sampled from a Maryland population of* Drosophila melanogaster

Element	Distal section Occupancy							Proximal section Occupancy													
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	9	10	11	12	13	14
roo	45	24	12	1	3	1	1	1	1	2	0	0	0	0	0	0	0	1	0	0	1
2156	8	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0
2158	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
2161	20	10	3	2	0	0	0	1	3	0	1	0	0	0	0	0	0	0	0	0	0
2181	12	3	0	0	0	0	0	2	0	0	0	1	0	1	0	0	0	0	0	0	0
2210	5	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2217	18	4	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0
297	22	7	1	2	3	0	0	1	3	0	0	0	0	0	0	0	0	0	0	1	0
412	18	6	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
copia	14	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

predicted direction. There was no evidence for an excess of elements at the tip; for this reason the tip and mid-sections of the chromosome were pooled for the purpose of the analyses presented previously.

Overall, these data provide strong evidence that TEs tend to be more abundant at the base of the *X* chromosome. For some elements, such as *2156* and *2158*, which have a mean copy number per *X* chromosome near one, this effect is so pronounced that nearly all copies are found at the base of the chromosome (see Table 5). A less strong pattern is seen for the higher copy number elements. The low copy number element *2210* shows no significant effect, however.

Inspection of the original data shows that the accumulation is most intense in the regions closest to the base. Band 20A1–2, for example, shows high element frequencies for several elements (*roo, 2156, 2158, 2181* and *297*). In the case of *roo*, indeed, all fourteen chromosomes in the sample appear to contain a copy of the element in this band. Band 20A3 also shows high frequencies of *2156, 2158, 2181* and *412* (11 out of 14 in the case of *412*). Band 19E3–4 also has a high frequency for *roo* (11 out of 14). These effects may be seen in the occupancy profiles shown in Table 5, which lists the numbers of bands that contain copies of each element in 1, 2, 3,..., 14 chromosomes in the sample. The profiles are given separately for the distal section of the chromosome (division 1–18), and for the proximal section (19–20A). The exact values of the occupancies in the proximal section are, however, somewhat uncertain, since it is sometimes hard to be sure of the exact position of a site of hybridization in this part of the chromosome, and it is possible that bands 20A1–2 and 20A3 may occasionally have been confused. Nevertheless, the fact of a strong tendency for element accumulation in this region, involving high element frequencies in bands 20A1–2 and 20A3, is unquestionable. Bands proximal to 20A3 cannot be reliably distinguished, although diffuse staining can be observed for many of these elements in the region between 20A3 and the chromocentre, indicating the presence of elements in some abundance.

(v) *Frequency distributions of element frequencies*

It is clear by inspection of the occupancy profiles for the distal section of the chromosome that all elements tend to be present at low frequencies in bands in this region. This can be quantified by obtaining estimates of the parameters α and β of the probability distribution of element frequencies (see section 1 of this paper, and Charlesworth & Charlesworth, 1983). These estimates for each element for the sampled *X* chromosomes are shown in Table 6. The proximal region of the *X* (divisions 19 and 20), where there is evidence for a strong differential accumulation of many elements [see section 3 (iv) above], is omitted. The values of α and β are estimated by the minimum χ^2 method of Charlesworth & Charlesworth (1983), except where infinite values of the two parameters give the best fit. This corresponds to equality of element frequencies at each occupable site, and the maximum likelihood procedure of Lewontin & Prout (1956) was used to obtain estimates of the element frequency per band and the total number of occupable bands on the *X* chromosome in these cases. The values of the parameters that yielded a minimum χ^2 value or maximum log likelihood were obtained by computer searches of the parameter space for the minimum or maximum. In two cases (*2156* and *2158*), all sites had an occupancy of one or zero, so that α is estimated as zero and β as infinity. In the case of finite estimates of α and β, the element frequency per occupable salivary chromosome band ($\bar{x}$) is estimated as $\alpha/(\alpha+\beta)$, and the total haploid number of occupable sites on the $X(m)$ is estimated as the mean copy number per chromosome divided by $\bar{x}$ (Charlesworth & Charlesworth, 1983).

https://doi.org/10.1017/S0016672300028482 Published online by Cambridge University Press

Table 6. *Parameters of the probability distributions of element frequencies for the distal section of the* X *chromosomes of* D. melanogaster *sampled from a Maryland population*

Element	α	β	$\bar{x}$	m	Element	α	β	$\bar{x}$	m
roo	0·8	12·5	0·060	191	*2210*	25	550	0·044	15
2156[a]	0	∞	0	∞	*2217*	0·4	40	0·009	204
2158[a]	0	∞	0	∞	*297*	0·0	5·5	0·000	∞
2161	2·5	35	0·067	58	*412*	17	380	0·043	55
2181[b]	∞	∞	0·028	46	*copia*[b]	∞	∞	0·019	63

[a] These elements occupied unique sites in each chromosome of the sample, so that no finite estimates of β and m can be obtained for them.
[b] The best fit for this element was obtained on the assumption of equal frequencies of elements at each occupable site, which implies infinite values of α and β (Charlesworth & Charlesworth, 1983). The values of $\bar{x}$ and m were estimated using a maximum likelihood procedure (Lewontin & Prout, 1956).

4. Discussion

(i) *Fit of the distribution of copy numbers to Poisson expectation*

As shown in Table 1, copy numbers for the combined tip and mid-section of the *X* in all but one of the elements (*roo*) are distributed between chromosomes of the sample in accordance with the Poisson distribution expected if element frequencies are low and uniform across different sites, and if there is no linkage disequilibrium (Charlesworth & Charlesworth, 1983; Langley *et al.* 1983). For *roo*, the observed variance in copy number is 4·40 and the Poisson expectation is 11·36, a discrepancy of 6·96. The estimated mean element frequency at the 61 segregating sites for the mid-section of the chromosome in the sample, is 0·19, which would contribute a reduction below Poisson expectation of $(11{\cdot}36)^2/85 = 1{\cdot}52$ to the variance [Charlesworth & Charlesworth, 1983, eq. (4)]. The sum of the values of the coefficient of linkage disequilibrium between all pairs of mid-section *roo* elements in the sample is $-2{\cdot}56$, although this is not statistically significant. The contribution of linkage disequilibrium to the copy number variance is twice this sum, and so the reduction in variance below binomial expectation attributable to this term is 5·12 if taken at face value. An additional reduction term comes from the variance in element frequencies between different sites, multiplied by the number of sites [Charlesworth & Charlesworth, 1983, eq. (4)]. In the present case, this is equal to 0·72. The net expected reduction from all these sources is 7·36 which is greater than the observed reduction of 6·96.

This suggests that the deviation from Poisson for *roo* is due to a chance accumulation of several terms. This interpretation is consistent with the fact that a reduction in variance below Poisson expectation was also observed in the 1984 North Carolina sample of Montgomery *et al.* (1987). Reanalysis of the original data, kindly provided by Elizabeth Montgomery, yielded a mean and variance of 8·83 and 4·26 for *roo* in the distal section of the *X* chromosome. The distribution of copy number in this data set did not differ significantly from Poisson, however ($\chi^2_8 = 4{\cdot}99$, $P > 0{\cdot}05$). The data of Montgomery *et al.* (1987) showed no reduction in variance below Poisson expectation for the second and third chromosomes for *roo*, and a reanalysis of this data for the distal sections of these chromosomes fails to show any deviation from Poisson expectation. This further suggests there is no special significance to the finding of significant departure from Poisson for *roo* on the *X* chromosome in the Maryland 1986 sample.

(ii) *Clustering of elements along the chromosome*

The analysis of the correlations in occupancy for the pooled sample of *X* chromosomes showed that only one element (*2161*) deviated from a random distribution along the chromosomes of sites that were occupied at least once [section 3 (ii) above], such that occupied sites tended to be closer together than would be expected on a uniform distribution for this element family. This result is unlikely to be due to linkage disequilibrium between neighbouring segregating sites, since the test for this fails to reveal any evidence for positive linkage disequilibrium of this kind for *2161* ($\chi^2_1 = 0{\cdot}84$). Indeed, the average correlation in element frequency between neighbouring segregating sites is negative ($-0{\cdot}02$) for *2161*. The most likely explanations of this result are either that *2161* has some preference for transposition into a limited number of sites that are relatively clustered, or that it is preferentially maintained in a clustered set of sites. Examination of the distribution of *2161* over the pooled sample of *X* chromosomes reveals that 9 of the 17 sites that it occupies outside divisions 19 and 20 are in divisions 15–18. This is within the zone where crossing over is relatively infrequent on the *X* (Lindsley & Grell, 1968; Lindsley & Sandler, 1977); the map

https://doi.org/10.1017/S0016672300028482 Published online by Cambridge University Press

distance between the distal ends of divisions 15 and 20 is only about 8 units, compared with a total map length of about 66. The accumulation in this region may thus reflect a tendency for elements to accumulate where crossing over is infrequent.

The general absence of any departure from a random distribution of elements for the distal section of the *X* chromosome suggests that there is little or no tendency for transposition events to produce insertions into nearby chromosomal sites, since any tendency for new copies to be inserted into bands close to the old one would tend to produce an autocorrelation in the distribution of occupied sites for a given element family in the pooled sample of chromosomes. Of course, it is possible that transpositions could produce insertions into physically adjacent sites within the same band, which would not be detectable with the resolution of *in situ* hybridization. There is evidence for this with new P-element insertions in *D. melanogaster* (Eggleston, pers. comm.). On the other hand, studies of natural populations of *D. melanogaster* by restriction fragment mapping of defined regions do not seem to reveal any tendency for multiple insertions of the same element within the same haplotype (Aquadro *et al.* 1986, 1988; Cross & Birley, 1986; Langley *et al.* 1982, 1988; Leigh Brown, 1982; Schaeffer *et al.* 1988).

(iii) *Correlations between element families in identity of sites occupied*

Table 3 shows that there are correlations between different element families with respect to the identities of the sites occupied at least once in the sample of chromosomes. There appears to be a group of elements that share occupied sites in common (*roo*, *2161*, *2181*, *2217*, *297* and *412*), whereas the remaining elements show small or negative correlations. The lack of many significant correlations involving *2156*, *2158* and *2210* is perhaps not surprising, since these elements are much rarer than the other elements outside the base of the chromosome (see below). This pattern is not, however, absolute; *412*, for example, shows positive correlations with *roo*, *2161*, and *297*, but no correlation with *2217*. Conversely, *copia* shows correlations with *roo* and *2161* but no other elements. This lack of absolute consistency in the pattern of correlations suggests that the apparent preference of several families of elements for certain bands is not due simply to some bands containing more DNA, and hence acting as larger targets for insertion than others.

There was no evidence for statistically significant correlations within individual chromosomes between different elements with respect to the identity of occupied sites, of the kind reported by Montgomery & Langley (1983) for *297* and *421*. The absence of a within-chromosome effect suggests that the correlations in occupied sites reflect some degree of shared site-specificity for insertions for the element families in question, or protection of certain regions against excision of these classes of element. It is clearly not due to the tendency for elements to accumulate at the base of the chromosome, since the region of strongest accumulation (divisions 19 and 20) was omitted from this analysis.

On the other hand, there is no evidence for any correlations between different element families with respect to the identities of the sites which had the highest element frequencies in the sample of chromosomes. This was tested for by a modification of the procedure described in section 3(iii), giving only the sites classed as having high element frequencies in the distal section of the chromosomes scores of 1. (In the case of *roo*, sites with occupancies $\geqslant 4$ were classed as high frequency; for *2161* and *297*, the cut-off was $\geqslant 4$, and for *2181*, *2217*, *412* and *copia* it was $\geqslant 2$.) This indicates that the high frequencies of elements at particular sites do not reflect some common affinity of different families for these sites.

(iv) *Accumulation of elements at the base of the* X *chromosome*

Langley *et al.* (1988) have discussed mechanisms that might lead to a tendency for elements to accumulate differentially in regions of reduced crossing over. In particular, they pointed out that crossing over between homologous elements located at different chromosomal sites that leads to the production of deleterious chromosome rearrangements. Provided that the frequency of unequal exchange between a pair TEs reflects the frequencies of meiotic exchange in the regions in which they are located, their models showed that there should be a higher equilibrium abundance of elements in regions where crossing over is less frequent than normal. Muller's ratchet is another population process that may lead to a build-up of transposable elements in regions of restricted crossing over (Charlesworth, 1985). Finally, if heterozygous elements can be excised as a result of gene conversion, as proposed by Holliday (1982), there could be a lower rate of such excision events in regions where conversion rates are lower, which would presumably encompass regions of restricted meiotic crossing over. There is evidence from yeast for this process for the *Ty* elements (Kupiec & Petes, 1988).

The prediction that elements will accumulate in regions of restricted crossing over was tested in section 3(iv), by examining the distribution of elements along the salivary chromosome map, since the frequency of exchange is lower near the tips and bases of *Drosophila* chromosomes than in the mid-sections (Lindsley & Sandler, 1977). The results of the tests showed a strong tendency for elements to accumulate at the bases of the chromosomes, but no tendency to accumulate at the tips, in partial accordance with the above predictions. The accumulation of *2161* in regions 15–18 supports the interpretation that the

https://doi.org/10.1017/S0016672300028482 Published online by Cambridge University Press

accumulation of elements is due to the suppression of recombination, rather than some specific tendency of elements to insert into divisions 19 and 20. The present results for *roo* agree with those of Langley *et al.* (1988) for *X* chromosomes sampled from a North Carolina population. Langley *et al.* (1988) discuss possible reasons for the fact that accumulation at the tip does not seem to occur, but at present this difference is a puzzle.

There is a possibility that the accumulation of elements at the base of the *X* could be due to an unusually high density of DNA here, providing a larger target size for element insertions, or to an unusually low density of functional loci, so that selection does not oppose the accumulation of elements as strongly as elsewhere. But, as shown in table 5 of Langley *et al.* (1988), the proportion of the DNA of the euchromatic *X*, the proportion of mutable loci, and the fraction of the total chromosome length represented by this region are all in good agreement. This is difficult to reconcile with these two explanations, and with the fact that not all elements show strong accumulation in this region. As noted in section 3(iv), the accumulation of elements at the base of the *X* chromosome is particularly marked for subdivisions 19E and 20A. Miklos *et al.* (1984, 1988) have isolated clones of DNA from these regions of the *X*, and have shown that they contain unusually high abundances of middle repetitive DNA sequences, some of which cross-hybridize with known TE families. Certain of these sequences are also abundant in the β-heterochromatin at the bases of the autosomes, and on the fourth chromosome, which are other regions where crossing over is restricted. Similarly, Barbara Wakimoto has cloned sequences containing the locus of *light* (located at the base of 2L, in subdivision 40F), and has detected the abundant presence of middle repetitive DNA sequences in these sequences (personal communication). Stephan & Langley (1989) have also reported an unusual abundance of insertion sequences in DNA from the centromeric region of the *X* chromosome of *D. ananassae*. These finer-scale studies are therefore in agreement with the results of our *in situ* results, and with the hypothesis that regions of reduced crossing over tend to accumulate TEs.

There are some reasons for thinking that the first of the mechanisms outlined above is most likely to be the cause of the observed build-up of elements in regions where crossing over is infrequent. First, as discussed by Langley *et al.* (1988), there is direct evidence from a number of species, including *D. melanogaster*, for the production of rearrangements in the postulated fashion. Second, as discussed by Maynard Smith (1978, p. 35), in the models of selection versus deleterious mutations normally used to analyse Muller's ratchet, it is found that the process is very sensitive to the relative values of the per genome mutation rate and the strength of selection against the mutations concerned. In the context of transposable elements inserting into the base of the *X*, the relevant 'mutation' rate *U* is the product of the per genome transposition rate (the mean copy number per haploid genome times *u*), and the probability of insertion into the base. If elements are distributed approximately randomly across the chromosomes, as found by Montgomery *et al.* (1987), the 'mutation' rate is thus equal to the product of *u*, the haploid copy number for the *X* and the proportion of the *X* represented by the base. In the case of *roo*, the most abundant element, this is equal to $1{\cdot}7u$. The values for other elements are smaller. If the population is approximately at equilibrium, and excision is absent (as is necessary for the ratchet to work), we have $u \approx s$. Hence, $U \approx 1{\cdot}7s$ for *roo*. According to Maynard Smith and the analysis of Haigh (1978), this is close to the boundary value for which the ratchet can operate in the total absence of recombination, even in very small populations. *D. melanogaster* is known to have reasonably high effective local population sizes and rates of migration between populations (Mukai & Yamaguchi, 1974; Singh & Rhomberg, 1987), and there are significant frequencies of crossing over in the regions of the *X* for which we have found evidence of element accumulation (Lindsley & Grell, 1968). As defined here, the tip of the chromosome encompasses about 2·5 map units and the base about 2·2. It is therefore theoretically rather implausible that a ratchet mechanism could have been responsible for the build-up of elements observed here.

The hypothesis of lower rates of excision by gene conversion in these regions is in conflict with the evidence that excision of *copia*-like elements is much less frequent than the recombinational events that generate rearrangements. The data reported by Davis *et al.* (1987) on recovery of different chromosome classes from females heterozygous for w^a, a mutation due to a *copia* insertion, show that no wild-type chromosomes that were non-recombinant for flanking markers (which would be produced by excision of w^a) were produced, whereas several chromosomes with *white* phenotypes occurred that were the product of meiotic unequal exchange between TEs. Furthermore, in yeast the excision of *Ty* elements by gene conversion seems to be primarily a mitotic event (Kupiec & Petes, 1988); since spontaneous mitotic exchange in *Drosophila* is not suppressed in the bases of the euchromatic sections of the chromosome (Becker, 1976), mitotic events are not likely to be capable of producing the observed pattern.

One final possibility should be discussed. In regions of restricted crossing over, there is a much greater opportunity for the occurrence of hitch-hiking effects, when advantageous alleles are fixed by natural selection (Maynard Smith & Haigh, 1974). This has the consequence that the effective population size in such regions is lower than elsewhere. Aquadro *et al.* (1988, p. 886) have suggested that TEs may be

expected to be more abundant when effective population size is low, due to the greater effectiveness of purifying selection with larger N_e, and proposed this as an explanation for the higher abundance of elements in *D. melanogaster* compared with *D. simulans*. If this argument were correct, the higher abundance of TEs in regions of restricted crossing over could be explained in the same way. The simulation study by Charlesworth & Charlesworth (1983, table 2) of the joint effects of selection, transposition and drift shows that there can indeed be a higher mean copy number per individual in small populations, but the effect is trivial unless N_e is of the order of 100 or less. This seems unlikely to be the case in *Drosophila*, even with restricted crossing over, unless hitch-hiking events are rather common. The fact that elements in the tip and base show little or no linkage disequilibrium suggests that such events are in fact rare, as hitch-hiking is known to create linkage disequilibrium between closely linked markers (Thomson, 1977). Furthermore, the results of Charlesworth and Charlesworth were obtained on the assumption that all sites are exposed to the same N_e; for the case in question, N_e is reduced only in the region of restricted crossing over, and so the overall effect is likely to be smaller.

(iv) *The distribution of element frequencies*

The data on element frequencies in the distal portion of the *X* chromosome, analysed in section 3(v), show clearly that elements of all 10 families are rare at individual chromosomal sites. Where finite estimates of the parameters α and β of the frequency distribution, defined in section 1, could be obtained, large values of β relative to α were found. It is clear in all these cases that the data can only be fitted by β values that are considerably greater than one. The maximum value of $\tilde{x}$ is 0·07. A similar conclusion was reached by Charlesworth (1988), on the basis of earlier studies of TEs in *D. melanogaster* populations.

It should be noted that the cases with very large values of β in Table 6 are probably artefactual; as can be seen these are cases where element abundances are relatively low, and the high values of β reflect the small number of sites with multiple occupation. It is probable that larger samples would turn up more sites with multiple occupancy, lowering the estimates of β. The minimum χ^2 values in these cases do not differ significantly from much smaller values for α and β. For example, for *412* the χ^2 statistic for goodness of fit to $\alpha = 0$, $\beta = 25$ is 1·51, compared with 0·2 for the minimum χ^2 estimate of $\alpha = 17$, $\beta = 380$ (2 D.F.). In contrast, the more abundant elements such as *roo*, *2161*, and *297* show much greater sensitivity of goodness-of-fit to the parameter values, and the parameter estimates for these can be regarded with some confidence.

For the more abundant elements, the number of occupable bands is either infinite or of similar magnitude to the total number of bands on the photographic map in this region (165). This suggests that the elements in this study do not have a high degree of site-specificity of requirements for insertions, at least at the level of salivary chromosome bands. The harmonic mean estimate of *m* over all elements is 66·7.

These findings strongly support the notion that the spread of elements as a result of transpositional increase in copy number is opposed by some force or forces. Additional evidence for this is provided by a re-analysis of the third chromosomes sampled by Montgomery *et al.* (1987) from a North Carolina population. The slides hybridized with *roo* were scored by us for the band positions of elements (excluding the bases of the chromosomes), and the following occupancy profiles were obtained:

No. of chromosomes (chr.) occupied	1	2	3	$\geqslant 4$
3L (12 chr.)	44	27	7	7
3R (11 chr.)	52	24	11	3

These data yield estimates for α of 3·4 and 2·5 for 3L and 3R respectively, and for β of 32 and 28, correcting those given by Charlesworth (1988). These are rather larger than the estimates obtained for *roo* for the *X* chromosome in our Maryland sample ($\alpha = 0{\cdot}8$, $\beta = 12{\cdot}5$), even taking into account the fact that the effective population size for *X*-linked genes is 0·75 that for autosomal genes. Correspondingly, the estimated mean element frequency $\tilde{x}$ is 0·096 for 3L and 0·082 for 3R, compared with 0·060 for the Maryland *X*. This may either reflect an inter-population difference, or a somewhat lower element frequency on the *X*, possibly due to the greater selection against insertional mutations for *X*-linked TEs (Montgomery *et al.* 1987). Given that their study failed to detect the predicted lower abundance of *roo* on the *X* relative to the autosomes on this hypothesis, the former possibility seems more likely. The estimated numbers of occupable sites (*m*) are 128 and 162 for 3L and 3R, in rough accord with the number of resolvable salivary chromosome bands.

As described by Charlesworth (1988), the rate of transposition *u* can be estimated from the values of β obtained here. Estimates of the value of $4N_eM$ in *D. melanogaster* are available from studies of the geographic distribution of electrophoretic alleles (Singh & Rhomberg, 1987), and these can be deducted from the values of β in order to obtain the contribution from excision and selection. If the expected mean number of elements per individual is unchanging over time, then $u \approx s+v$ and so the value of β corrected for migration is approximately equal to $4N_eu$. From Table 6, the harmonic mean of the uncorrected β values is 31·3 (the reason for using the harmonic mean is to deal with the cases where β is infinite or very large; as noted above, these tend to be the low copy

number elements that provide little information). The correction for $4N_e M$ for an *X*-linked gene for East Coast US populations is 6·4 (Singh & Rhomberg, 1987), yielding an overall β value of 24·9. The autosomal N_e for a local population is estimated to be about 20000 from data on lethal allelism rates (Mukai & Yamaguchi, 1974), so that N_e for a sex-linked locus is 15000. The overall value of u estimated from these numbers is 4×10^{-4}, which is consistent with what little is known about typical rates of transposition of *Drosophila* elements (Charlesworth & Charlesworth, 1983; Eggleston *et al.* 1988). The data on *roo* provide the most reliable information, since this is the most abundant element. For the *X* chromosome, an estimate of u of 1×10^{-4} is obtained; the values for 3L and 3R are $2{\cdot}9 \times 10^{-4}$ and $2{\cdot}4 \times 10^{-4}$ respectively. A similar procedure could be used to estimate u from the values of α and m, but the values of α are mostly small and hence very sensitive to error. Indeed, many of the non-zero α values in Table 6 are not significantly different from zero.

This work was supported by PHS grant GM-36405-01, and a grant from the Louis Block fund of The University of Chicago. We thank Darlene Canada and Ruby Payne for assistance, and Elizabeth Montgomery for supplying us with the set of slides for the analysis of third chromosomes. We also thank Charles Langley for many helpful discussions.

References

Aquadro, C. R., 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.

Aquadro, C. R., Lado, K. M. & Noon, W. A. (1988). The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA sequence variation and divergence. *Genetics* **119**, 875–888.

Becker, H. J. (1976). Mitotic recombination. In *The Genetics and Biology of Drosophila*, vol. 1c (ed. M. Ashburner and E. Novitski), pp. 1020–1087. Orlando: Academic Press.

Brookfield, J. F. Y (1986). Population biology of transposable elements. *Philosophical Transactions of the Royal Society of London*, **B312**, 217–226.

Brookfield, J. F. Y., Montgomery, E. A. & Langley, C. H. (1984). Apparent absence of transposable elements related to the *P* element of *Drosophila melanogaster* in other species of *Drosophila*. *Nature* **310**, 330–332.

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

Charlesworth, B. (1988). The maintenance of transposable elements in natural populations. In *Proceedings of the International Symposium on Plant Transposable Elements* (ed. O. J. Nelson), pp. 189–212. New York: Plenum.

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.

Cross, S. R. H. & Birley, A. J. (1986). Restriction endonuclease map variation in the *Adh* region in populations of *Drosophila melanogaster*. *Biochemical Genetics* **24**, 415–433.

Davis, P. S., Shen, M. W. & Judd, B. H. (1987). Asymmetrical pairing of transposons in and proximal to the *white* locus of *Drosophila* accounts for four classes of regularly occurring exchange products. *Proceedings of the National Academy of Sciences USA* **84**, 174–178.

Doolittle, W. F. & Sapienza, C. (1980). Selfish genes, the phenotype paradigm, and genome evolution. *Nature* **284**, 601–607.

Eggleston, W. B., Johnson-Schlitz, D. & Engels, W. R. (1988). P–M hybrid dysgenesis does not mobilize other transposable elements in *D. melanogaster*. *Nature* **331**, 368–370.

Engels, W. R. (1986). On the evolution and population genetics of hybrid-dysgenesis causing transposable elements in *Drosophila*. *Philosophical Transactions of the Royal Society of London*, **B312**, 205–215.

Finnegan, D. J. & Fawcett, D. H. (1986). Transposable elements in *Drosophila melanogaster*. *Oxford Surveys on Eukaryote Genes* **3**, 1–62.

Fisher, R. A. (1958). *Statistical Methods for Research Workers*, 13th edn, Edinburgh: Oliver and Boyd.

Goldberg, M. L., Sheen, J.-Y., Gehring, W. J. & Green, M. M. (1983). Unequal crossing-over associated with asymmetrical synapsis between nomadic elements in the *Drosophila melanogaster* genome. *Proceedings of the National Academy of Sciences USA* **80**, 5017–5021.

Haigh, J. (1978). The accumulation of deleterious genes in a population. *Theoretical Population Biology* **14**, 251–267.

Hartl, D. L., Medhora, M., Green, L. & Dykhuizen, D. E. (1986). The evolution of DNA sequences in *Escherichia coli*. *Philosophical Transactions of the Royal Society of London* **B 312**, 191–204.

Hey, J. (1989). The transposable portion of the genome of *Drosophila algonqin* is very different from that of *D. melanogaster*. *Molecular Biology and Evolution* **6**, 666–679.

Holliday, R. (1982). Gene conversion: a possible mechanism for eliminating selfish DNA. In *Molecular and Cellular Mechanisms of Mutagenesis* (ed. J. F. Lemont and W. M. Generoso), pp. 259–264. New York: Plenum.

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

Kendall, M. G., Stuart, J. D. & Ord, J. K. (1983). *The Advanced Theory of Statistics*, 4th edn, vol. 3. London: Charles Griffin.

Kupiec, M. & Petes, T. (1988). Allelic and ectopic recombination between *Ty* elements in yeast. *Genetics* **119**, 549–559.

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. L. (1983). Transposable elements in Mendelian populations. I. A theory. *Genetics* **104**, 457–472.

Langley, C. H., Montgomery, E. A. & Quattlebaum, W. F. (1982). Restriction map variation in the *Adh* region of *Drosophila*. *Proceedings of the National Academy of Sciences USA* **79**, 5631–5625.

Langley, C. H., Montgomery, E. A., Hudson, R. R., Kaplan, N. H. & Charlesworth, B. (1988). On the role of unequal exchange in the containment of transposable element copy number. *Genetical Research* **52**, 223–235.

Langley, C. H., Shrimpton, A. E., Yamazaki, T., Miyashita, N., Matsuo, Y. & Aquadro, C. F. (1988). Naturally occurring variation in the restriction map of the *Amy* region of *Drosophila melanogaster*. *Genetics* **119**, 619–629.

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. Orlando: Academic Press.

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 natural populations of *Drosophila melanogaster*. *Genetical Research* **49**, 121–128.

Lewontin, R. C. & Prout, T. (1956). Estimation of the number of different classes in a population. *Biometrics* **12**, 211–223.

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

Lindsley, D. L. & Sandler, L. (1977). The genetic analysis of meiosis in female *Drosophila melanogaster*. *Philosophical Transactions of the Royal Society of London*, **B277**, 295–312.

Lindsley, D. L. & Zimm, G. (1982). Computerized stock list 3. *Drosophila Information Service* **57**, 1–213.

Lindsley, D. L. & Zimm, G. (1986). The genome of *Drosophila melanogaster*, part 2: lethals; maps. *Drosophila Information Service* **64**, 1–158.

Maynard Smith, J. (1978). *The Evolution of Sex*. Cambridge: Cambridge University Press.

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

Miklos, G. L. G., Healy, M. J., Pain, P., Howells, A. J. & Russell, R. J. (1984). Molecular genetic studies on the euchromatin–heterochromatin junction in the *X* chromosome of *Drosophila melanogaster*. I. A cloned entry point near to the uncoordinated (*unc*) locus. *Chromosoma* **89**; 218–227.

Miklos, G. L. G., Yamamoto, M.-T., Davies, J. & Pirrotta, V. (1988). Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the β-heterochromatin of *Drosophila melanogaster*. *Proceedings of the National Academy of Science USA* **85**, 2051–2055.

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.

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.

Mukai, T. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila*. *Genetics* **82**, 63–82.

Orgel, L. E. & Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.

Rubin, G. M., Brorein, W. J., Dunsmuir, P., Flavell, A. J., Strobel, J. J., Toole, J. J. & Young, E. (1981). '*copia*'-like transposable elements in the *Drosophila* genome. *Cold Spring Harbor Symposium on Quantitative Biology* **45**, 619–628.

Rudkin, G. T. (1965). The relative mutability of DNA in regions of the *X* chromosome of *Drosophila melanogaster*. *Genetics* **52**, 665–681.

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.

Singh, R. S. & Rhomberg, L. R. (1987). A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. *Genetics* **115**, 313–222.

Snedecor, G. W. & Cochran, W. G. (1980). *Statistical Methods*, 7th edn. Ames, Iowa: Iowa State University Press.

Stephan, W. & Langley, C. H. (1989). Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. *Genetics* **121**, 89–99.

Thomson, G. (1977). The effect of a selected locus on linked neutral loci. *Genetics* **85**, 753–788.