Genet. Res., Camb. (1992), **60**, *pp.* 103–114

The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution

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 second and third chromosomes isolated from a natural population of *Drosophila melanogaster*, by means of *in situ* hybridization of element probes to polytene chromosomes. It was found that the copy numbers per chromosome in the distal sections of the chromosome arms followed a Poisson distribution. Elements appeared to be distributed randomly along the distal sections of the chromosome arms. There was no evidence for linkage disequilibrium in the distal sections of the chromosomes, but some significant disequilibrium was detected in proximal regions. There were many significant correlations between different element families with respect to the identity of the sites that were occupied in the sample. There were also significant correlations between families with respect to sites at which elements achieved relatively high frequencies. Element frequencies per chromosome band were generally low in the distal sections, but were higher proximally. 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 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

Studies of the population properties of transposable elements (TEs) in *Drosophila* suggest that that TEs are maintained as a result of transpositional increase in copy number, and that their spread is checked by one or more opposing forces (Charlesworth & Langley, 1989). The chief evidence for this conclusion is provided by studies of the population frequencies of TEs at individual chromosomal sites in *Drosophila*, by means of *in situ* hybridization of TE probes to the polytene salivary gland chromosomes. In most cases, element frequencies per band appear to be low (Charlesworth & Langley, 1989), except for the four element families of *D. algonquin* and *D. affinis* studied by Hey (1989). Restriction mapping studies of intrapopulation variation of defined genomic regions confirm this result: with rare exceptions, TE insertions are present at very low frequencies at individual nucleotide sites in *Drosophila* population samples (Charlesworth & Langley, 1989).

Fits of the population frequency data from surveys using *in situ* data to a model of TE population dynamics indicate that β, the product of $4N_e$ (where N_e is the effective size of a local population) and the net rate of removal of an element from a chromosomal site, is generally a large number, of the order of 10–40 (Charlesworth & Langley, 1989, Table 2). The magnitude of the deterministic forces affecting element frequencies is thus much greater than that of genetic drift. This information does not permit identification of the nature of these forces. Several hypotheses concerning their nature, and means of testing these hypotheses, have been proposed in the literature (Charlesworth & Langley, 1989; Charlesworth, 1991).

The present paper describes the results of a survey of the distribution of nine families of transposable elements among a sample of second and third chromosomes isolated from a natural population of *D. melanogaster*. *In situ* hybridization was used to identify the locations of TEs, at the level of polytene chromosome bands. In agreement with our earlier study of a sample of X chromosomes from the same population (Charlesworth & Lapid, 1989), we find that elements are distributed randomly along the distal sections of the chromosome arms. Element frequencies in these regions are almost always very low, but tend to be higher in the centromere-proximal

* Corresponding author.

regions of the euchromatin. This suggests that the force or forces opposing the spread of TEs may operate more weakly in the proximal regions than distally. In the accompanying paper, we present evidence on the nature of these forces (Charlesworth, Lapid & Canada, 1992).

2. Materials and methods

(i) *Genetic stocks and breeding procedures*

We isolated fourteen homozygous viable and fertile isogenic independent second chromosomes from a population of *D. melanogaster* at Beltsville, Maryland in 1986 and 1987, and thirteen homozygous viable and fertile isogenic third chromosomes in 1987 and 1988. Wild male flies were kindly provided by Drs Jerry Coyne and Gerald Wilkinson. Single wild males were mated with females carrying either the second chromosome balancer *SM1* heterozygous over *Pm*, or the third chromosome balancer *TM6*, heterozygous over *Sb* (Lindsley & Zimm, 1992). The balancer stock females were also marked with the fourth chromosome recessive gene spa^{pol}. The marked chromosomes 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. From each cross, a single F_1 male heterozygous for the balancer and a wild-type chromosome transmitted from the wild male (+) was crossed to balancer stock females; in the next generation, balancer spa^{pol}/spa^{pol} females were crossed to $+/balancer$; spa^{pol}/spa^{pol} males. Females and male progeny heterozygous for the balancer and wild-type chromosome, and homozygous for spa^{pol}, were intercrossed. Their $+/+$; spa^{pol}/spa^{pol} progeny were mated together to establish a stock homozygous for the original wild-type chromosome, and marked with spa^{pol} as a precaution against contamination. Recessive lethal or infertile chromosomes were discarded. 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 or 1000, 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 gland polytene 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.

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, which 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 chromosome stock was scored for each of the nine elements described below. For numerical analysis of the data, the recognizable bands for each chromosome were numbered, and the state of each band for each isolated chromosome recorded in computer files. Copies of the data files and the key to the numbering of the bands 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 Lindsley & Zimm (1992). The remaining elements, *2156*, *2158*, *2161*, *2181* and *2217*, are incompletely characterized, but are thought to be retroposon-like (Rubin *et al.* 1981; Strobel, 1982). They have previously been used in species comparisons by Brookfield *et al.* (1984). There is evidence from DNA sequencing of the probe that *2161* contains sequences homologous to the element *jockey* (C. F. Aquadro, pers. comm.). It can thus tentatively be identified with this element, which is a retroposon with no long terminal repeats (Finnegan, 1989).

On completion of the population surveys, four cases were found in which elements were apparently present at the same band in all chromosomes sampled [see section 3(v) below]. These are *2181* at 60B1 on 2R, *2158* at 80B on 3L, *2156* at 87C1-2 on 3R, and *297* at 99E1-2 on 3R. In the X chromosome data of Charlesworth & Lapid (1989), only *roo* showed evidence of fixation (at 20A). This raises the question of whether or not these cases of apparent fixation are due to similar hybridization of flanking sequences with the original sites from which the probes were cloned, or whether they represent genuine high frequencies or fixation of elements at these chromosomal sites. The fact that no element is apparently

Table 1. *Tests of the Poisson distribution for nine families of* Drosophila melanogaster *elements for a sample of second and third chromosomes from a Maryland population*

	2L		2R		3L		3R	
Element	$\bar{n}$	V_n	$\bar{n}$	V_n	$\bar{n}$	V_n	$\bar{n}$	V_n
roo	8·07	6·69	8·86	6·59	8·62	8·92	12·15	12·31
2156	2·07	3·76	1·36	1·48	0·85	1·31	1·62	2·42
2158	1·43	0·57	0·93	0·84	0·92	0·58	0·54	0·94
2161	5·21	4·18	6·29	8·68	4·85	9·31	9·31	12·90
2181	2·00	1·69	2·07	1·76	1·77	2·53	5·00	4·50
2217	2·64	1·79	2·71	3·60	2·54	3·94	2·38	2·26
297	2·36	2·09	3·07	4·22	2·85	2·47	3·92	3·41
412	2·79	1·87	3·43	3·03	2·38	2·26	3·85	3·64
copia	2·50	2·73	1·93	1·30	3·08	3·91	3·46	5·44

Note. No significant departures from the Poisson distribution were detected.

fixed at more than one site suggests that the former is more likely to be the case. We have not answered this question yet, because it requires using independently isolated clones of these elements to see if there is still hybridzation at the sites in question. We have therefore taken the conservative approach of treating all these cases as artefacts, and scoring the sites in question as lacking elements for the purposes of statistical analysis. (This applies also to *roo* in the re-analysis of the X-linked data in the accompanying paper, which appeared to be fixed in band 20A1.) This procedure has little effect on our conclusions, since few sites are involved compared with the total number of segregating sites.

(iv) *Definitions of the cytogenetic regions used to partition the chromosomes*

Since previous results suggested that elements may accumulate in regions of restricted recombinational exchange (Charlesworth & Langley, 1989), we have divided the polytene chromosome arms into tip, mid and base, using the cytogenetic criteria of Langley *et al.* (1988, Table 2). (The tips and bases of the autosomes correspond roughly to the most distal standard chromosome division, and to the proximal three divisions, respectively.) Exchange is close to normal at most of the tips of the autosomes (Ashburner, 1989, chap. 11), and we find little evidence for accumulation of elements at the tips (Charlesworth *et al.* 1992). For most purposes, we have thus combined the tips and mid-sections of the chromosomes to form a 'distal section'. The numbers of identifiable bands in the tip, mid and basal regions were as follows: X (20, 140, 12), 2L (7, 120, 22), 2R (14, 125, 25), 3L (15, 130, 21), 3R (11, 169, 34).

3. Results

(i) *Distribution of elements between homologous chromosomes*

If element frequencies are low at each site, as is the case here [section 3(iv)], and if there is no linkage disequilibrium among elements belonging to the same family, element copy numbers per arm should follow the Poisson distribution, with equality of mean and variance (Charlesworth & Charlesworth, 1983). Since elements often accumulate in excess of random expectation at the bases of the chromosomes (Charlesworth & Langley, 1989; Charlesworth *et al.*, 1992), tests for agreement with the Poisson expectation have been carried out on the distal sections of the chromosome arms alone [see section 2(iv)].

Table 1 gives the means and variances of copy number for the distal section of each chromosome arm for each family. The agreement of the distributions of copy number per arm with the Poisson distribution was tested by χ^2, pooling adjacent classes as necessary to avoid low expected numbers. In no case was there any significant deviation from the Poisson expectation, suggesting that the conditions for the Poisson distribution are met, at least approximately. Similar results have been found in most other studies (Charlesworth & Langley, 1989).

(ii) *Correlations in occupancy between adjacent sites*

A test for the independence across chromosomal sites of the distribution of elements belonging to the same family is to examine the randomness of the distribution of sites that are occupied at least once in the sample, across sites within a chromosome arm. These tests were carried out as described by Charlesworth & Lapid (1989).

Table 2 shows the autocorrelations in occupancy (r) between adjacent bands, together with their ratios to their sampling standard deviations (z), calculated

https://doi.org/10.1017/S0016672300030792 Published online by Cambridge University Press

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

Element	r	z	r	z
		2L		2R
roo	0·079	0·90	−0·061	0·72
2156	0·029	0·32	−0·053	0·62
2158	−0·029	0·33	0·112	1·32
2161	0·092	1·05	−0·007	0·08
2181	0·028	0·32	0·811	1·04
2217	0·150	1·69	0·106	0·13
297	−0·178	0·20	0·130	1·54
412	−0·269	0·30	−0·072	0·85
copia	−0·165	1·87	−0·263	0·31
		3L		3R
roo	−0·016	0·19	0·228	3·06*
2156	0·033	0·40	0·043	0·58
2158	0·016	0·19	−0·041	0·55
2161	0·048	0·59	0·071	0·96
2181	0·165	1·99	0·067	0·90
2217	−0·086	1·04	−0·101	1·36
297	0·074	0·89	0·067	0·90
412	−0·132	1·59	−0·139	1·88
copia	0·096	1·16	0·069	0·93

* $P < 0·05$.

according to the formulae of Kendall, Stuart & Ord (1983, pp. 548–551). (Occupancy for a given family for a band is scored as zero if no elements are present in that band in the sample, and as one if the band is occupied at least once by a member of the family.) The significance of deviations from randomness was assessed by comparing the distribution of the numbers of bands separating occupied sites with the geometric distribution by means of a χ^2 statistic (Charlesworth & Lapid, 1989). The only element to show an apparently significant χ^2 value is *roo* for 3R ($\chi^2_5 = 12·1$, $P < 0·05$), but this may be a result of sampling error in view of the large number of tests conducted. There is therefore no evidence for any deviation from a random pattern of element insertion within the distal sections of the chromosomes.

(iii) *Linkage disequilibrium between elements belonging to the same family*

Pairwise linkage disequilibrium between members of the same family of TEs was tested as described by Charlesworth & Lapid (1989), by examining all possible 2×2 contingency tables formed by pairs of segregating sites. The distal sections of each chromosome arm were tested separately, yielding four sets of tests for each TE family. 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. This distribution was obtained 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.

No significant deviations between observed and expected numbers were detected for the seven TE families that were sufficiently abundant outside the bases of the chromosome to make this test worthwile (all except *2156* and *2158*), except for *roo* on 3R where χ^2 was 21·1 (12 D.F., $P < 0·05$). This marginally significant result is probably due to chance, in view of the number of tests performed (28 in total). In all other cases, the distributions of observed and expected numbers in each interval were in remarkably close agreement. In addition, linkage disequilibrium that is consistently in the same direction was tested for by the method of Mantel & Haenszel (Snedecor & Cochran, 1980, p. 213), and no significant result was obtained.

Of course, many of the comparisons involve pairs of very distant sites, for which linkage disequilibrium is unlikely. Two 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 genetically close to each other on the chromosome. There was no evidence for departure from random expectation. Second, the analysis was performed for all pairs of sites within the bases of each chromosome arm, where recombination is highly suppressed (Ashburner, 1989, chap. 11). In this case, element families *2156* and *2158* were included in the tests, as they are fairly abundant at the bases of the chromosomes (see Tables 4 and 5 below). Again, the data mostly agreed well with the null hypothesis of no linkage disequilibrium, although the numbers of pairs of sites is limited. *412* on 2L showed evidence of positive linkage disequilibrium, with a value of 3·87 ($P < 0·001$) for the Mantel–Haenszel normal deviate statistic, and a χ^2 of 9·5 (1 D.F., $P < 0·01$). *2156* on 3R gave a Mantel–Haenszel statistic of 4·12 ($P < 0·001$), with an indication of overall positive linkage disequilibrium, although the χ^2 for goodness of fit was not significant ($\chi^2 = 5·2$, 2 D.F.).

No separate analysis of linkage disequilibrium was carried out for the tips of the autosomes, as was done for the X (Charlesworth & Lapid, 1989), since the suppression of exchange at the tips of the autosomes is much less marked than for the X (Ashburner, 1989, chap. 11).

Linkage disequilibrium between members of pairs of different families was also tested for by similar methods (described by Charlesworth & Lapid, 1989). Distal and proximal regions of each chromosome arm were examined separately, for all pairs of distinct sites in both regions, and for all pairs of adjacent sites

Table 3. *Correlation coefficients between pairs of families of* Drosophila melanogaster *elements with respect to the identity of sites occupied at least once*

	roo	*2156*	*2158*	*2161*	*2181*	*2217*	*297*	*412*	*copia*
roo		0·140	0·167	0·182	0·000	0·032	−0·031	0·180	0·031
		0·244***	0·202*	0·366***	0·155	0·412***	0·179*	0·278***	0·274**
2156	0·028		0·443***	0·120	−0·053	−0·022	0·210	0·277	0·124
	0·087		0·562***	0·187*	0·125	0·249**	0·102	0·102	0·143
2158	0·249**	0·230**		0·087	0·000	−0·053	0·031	0·020	0·336**
	0·095	0·112		0·159	0·001	0·170	0·137	0·137	0·013
2161	0·228**	0·003	0·156		0·265*	0·237*	0·307**	0·092	0·371**
	0·407***	0·116	−0·085		0·141	0·245**	0·065	0·134	0·216*
2181	0·236**	−0·035	0·104	0·076		0·109	−0·012	−0·023	−0·015
	0·261***	0·139	0·025	0·236**		0·018	0·160	0·246**	0·195*
2217	0·161*	−0·005	0·047	0·116	0·177*		0·220*	0·201*	0·298*
	0·148*	0·277**	0·090	0·163*	0·015		0·139	0·260**	0·002
297	0·229**	0·063	−0·082	0·004	0·225*	0·075		0·159	0·253*
	0·265***	0·109	0·031	0·147*	0·211**	0·183*		0·007	0·054
412	0·005	−0·060	−0·137	0·194*	0·105	0·106	0·062		0·159
	0·250***	0·187*	0·028	0·247**	0·108	0·138	0·315***		0·099
copia	0·240**	−0·083	0·031	0·080	0·197*	0·088	0·129	0·205*	
	0·203**	0·044	0·072	0·141*	0·017	0·147	0·287***	0·205**	

* $P < 0·05$; ** $P < 0·01$; *** $P < 0·001$.
Note. The entries above the diagonal are for chromosome 2; the entries below the diagonal are for chromosome 3. The upper set in each row are for the left arm; the lower are for the right arm.

within the distal regions. There was only one case of a clearly significant deviation from random expectation, involving *2161* and *2217* on 3L for all pairs of distal sites (95 cases with a correlation of 0·9–1·0, instead of 64 expected; the Mantel–Haenszel z was 4·42 ($P < 0·0001$). The meaning of this result is obscure, in view of the lack of a significant effect for pairs of adjacent sites.

In addition, linkage disequilibrium was tested for indirectly by examining the correlations in copy numbers per sampled chromosome for pairs of element families. In a random-mating population, linkage disequilibrium is the source of such correlations (Bulmer, 1980, p. 227). For the purposes of this analysis, chromosome arms were divided into distal and proximal regions, as before, which were analyzed separately. Elements *2156* and *2158* were excluded from the analyses of the distal regions, on account of their low abundances. Standard normal-based tests of the significance of individual correlations are unreliable in this case, except for high copy number families, due to the Poisson nature of the distribution copy numbers between families. For this reason, the means and variances of the correlations for all pairs of elements for a given chromosome arm were calculated. In order to test the hypothesis of a zero correlation between element families, the former can be compared with an expected value of zero; the latter with an approximate expected value of $1/(n-1)$, where n is the number of chromosomes sampled. For large numbers of correlations, normal-based tests of these statistics should be valid.

For the distal regions of the chromosomes, both means and variances of the correlations were close to their expected values, confirming the impression of a lack of linkage disequilibrium. A similar result held for the correlations between the two distal sections of each chromosome arms (here, correlations in copy numbers for the same family across arms, and between pairs of different families across arms, were calculated).

In the case of the proximal sections of the arms, the results were slightly different. For 3L, there were several large, positive correlations between families (e.g. *2156/copia*, 0·537; *2161/297*, 0·811; *2181/copia*, 0·741). Overall, the mean correlation was 0·207, with standard error of 0·049, $z = 4·22$ ($P < 0·001$). Similarly, several large, positive correlations in copy numbers per chromosome between the proximal sections of 3L and 3R were obtained (e.g. *2181/412*, 0·566, *2181/copia*, 0·548), and the mean correlation was 0·095 (standard error 0·031, $z = 3·08$, $P < 0·01$). Since there is no exchange in the proximal heterochromatin, these between-arm correlations for the proximal regions of the euchromatin (as well as the correlations within the proximal regions) must represent the effects of linkage disequilibrium. As would be expected from the large map distances involved, there is no evidence for significant correlations between the right and left distal arm sections.

(iv) *Correlations between sites occupied by different families*

Charlesworth & Lapid (1989) reported that there was a tendency for different families of TEs to share polytene chromosome bands that were occupied at least once in a sample of X chromosomes. Their

https://doi.org/10.1017/S0016672300030792 Published online by Cambridge University Press

Table 4. *Occupancy profiles for nine families of elements among fourteen second chromosomes sampled from a Maryland population of* Drosophila melanogaster

	Distal section[a] occupancy							Proximal section occupancy													
Element	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2L																					
roo	23	6	9	2	0	0	0	4	3	3	1	0	0	0	0	0	0	0	0	0	0
2156	9	3	0	0	0	0	0	3	0	0	0	0	0	1	1	0	0	0	0	0	0
2158	7	1	0	0	0	0	0	3	0	0	1	0	1	0	0	0	0	0	0	0	0
2161	18	4	5	0	1	1	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0
2181	8	1	1	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0
2217	9	6	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
297	14	2	1	0	0	0	0	4	1	1	1	0	1	0	0	0	0	0	0	0	0
412	15	3	0	0	0	0	0	2	3	0	0	0	0	0	0	0	0	0	0	0	0
copia	17	0	0	0	0	0	0	3	2	0	1	0	0	0	0	0	0	0	0	0	0
2R																					
roo	31	20	9	4	2	0	0	6	2	1	1	0	0	0	0	0	0	0	0	0	0
2156	14	1	1	0	0	0	0	4	0	0	2	0	1	0	0	0	0	0	0	0	0
2158	9	2	0	0	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0
2161	31	11	8	0	1	1	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0
2181[b]	22	2	1	0	0	0	0	6	1	0	0	1	0	0	0	0	0	0	0	0	0
2217	23	6	1	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0
297	28	6	1	0	0	0	0	4	1	1	0	0	0	0	0	0	0	0	0	0	0
412	26	6	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
copia	20	2	1	0	0	0	0	4	0	0	0	0	0	0	1	0	0	0	0	0	0

[a] This includes salivary chromosome divisions 21–32 and band 33A1; the more proximal regions 33–37 have been excluded, due to the difficulty of localizing elements to bands here. This also applies to sections 3(iii)–(iv).
[b] One distal site (60B1) is apparently fixed for this element; this may be due to hybridization with flanking sequences cloned simultaneously with the element.

method of analysis was applied here to the data on each autosomal arm. Product-moment correlations of the occupancy score of section 3(ii) were calculated between pairs of families in the distal sections of the chromosome arms; the significance of the associations was tested using Fisher's exact test [see section 3 (iii)]. The results are displayed in Table 3. It will be seen that there are numerous examples of highly significant pairwise associations between element families, in agreement with the results for the X chromosome.

In addition, tests for associations between pairs of families at the level of individual chromosomes were carried out, by applying the methods of section 3 (iii) for testing for linkage disequilibrium to pairs of element families at the same distal site. There was no indication of any significant associations for the distal sites (*2156* and *2158* were excluded). Thus, as in the case of the X chromosome (Charlesworth & Lapid, 1989), the correlations in occupancy were not due to any tendency of elements belonging to different families to be found as insertions into the same band.

(v) *The frequency distributions of element families*

The information on the frequencies of elements at individual chromosomal locations is summarized in Tables 4 and 5 in terms of occupancy profiles for each family for each chromosome. These give the numbers of sites for which elements are present 1, 2, 3, ... times at the same band in the sample. In view of previous results indicating that some families of elements achieve higher frequencies at the bases of the chromosomes (Charlesworth & Langley, 1989), the occupancy profiles are given separately for the distal sections and the bases. It can be seen from the tables that this pattern of higher basal frequencies holds also for some families in the present data set, especially *2156*, *2158* and *297*. High frequencies of elements at the chromosomal base are associated with higher mean copy numbers than are expected with random insertions of elements into this region; this is analysed in more detail in the accompanying paper (Charlesworth *et al.* 1992).

The rest of the discussion here thus concerns the frequencies of elements in the distal parts of the autosomal arms. The occupancy profiles for the distal section of the chromosome show that elements tend to be present at low frequencies in bands in this region. There are three cases where an element is apparently fixed at a distal location (*2181* on 2R, *2156* and *297* on 3R). Reasons have been given in section 2(iii) above for believing that these apparent fixations may be due to hybridization with flanking sequences in the probe DNA, and so these sites are treated as though elements are absent from them in the following analysis.

The stationary probability distribution of element frequencies in the population at individual chromosomal sites for a given family of TEs is expected to follow a β-distribution, such that the probability density of element frequency x is proportional to

https://doi.org/10.1017/S0016672300030792 Published online by Cambridge University Press

Table 5. *Occupancy profiles for nine families of elements among thirteen third chromosomes sampled from a Maryland population of* Drosophila melanogaster

	Distal section[a] occupancy							Proximal section occupancy												
Element	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8	9	10	11	12	13
3L																				
roo	43	14	7	5	0	0	0	5	5	0	2	1	0	0	0	0	0	0	0	0
2156	9	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0
2158[a]	10	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1
2161	31	7	2	3	0	0	0	2	2	1	0	0	2	0	0	0	0	0	0	0
2181	19	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2217	27	3	0	0	0	0	0	2	0	1	1	0	0	0	0	0	0	0	0	0
297	25	4	0	1	0	0	0	3	3	1	1	0	0	0	0	0	0	0	0	0
412	23	4	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0
copia	27	5	1	0	0	0	0	3	2	0	1	0	0	0	0	0	0	0	0	0
3R																				
roo	47	16	15	6	2	0	0	7	4	3	1	2	0	0	0	0	0	0	0	0
2156[b]	19	1	0	0	0	0	0	4	1	0	0	0	0	0	0	0	1	0	0	0
2158	7	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	1	0	0	0
2161[c]	33	15	8	1	3	1	0	3	2	1	0	0	0	0	0	0	0	0	0	0
2181	29	7	3	2	1	0	0	4	1	0	0	1	0	0	0	0	0	0	0	0
2217	23	3	0	2	0	0	0	4	1	1	0	0	0	0	0	0	0	0	0	0
297[d]	32	6	1	1	0	0	0	5	3	0	1	0	1	0	0	0	0	0	0	0
412	32	8	1	1	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0
copia	17	8	2	0	0	1	0	7	0	3	0	0	0	0	0	0	0	0	0	0

[a] One proximal site (80B) is apparently fixed for this element.
[b] One distal site (87C1-2) is apparently fixed for this element.
[c] One distal site (100C3) has an occupancy of 9.
[d] One distal site (99E1-2) is apparently fixed for this element.
The cases of apparent fixation may be due to hybridization with flanking sequences cloned simultaneously with the element.

Table 6. *Parameters of the probability distributions of element frequencies for the distal sections of the second chromosome of* D. melanogaster *sampled from a Maryland population*

	2L[a]					2R			
Element	α	β	$\hat{x}$	m	Element	α	β	$\hat{x}$	m
roo	2·6	30·5	0·079	64	*roo*	3·2	31·5	0·092	96
2156[b]	∞	∞	0·037	30	*2156*	0	26·0	0	∞
2158[b]	∞	∞	0·026	25	*2158*[b]	∞	∞	0·027	35
2161	0	5·9	0	∞	*2161*	0·4	10·5	0·037	56
2181	0	16·5	0	∞	*2181*	0	34·0	0	∞
2217[b]	∞	∞	0·054	28	*2217*[b]	∞	∞	0·036	75
297	0	23·0	0	∞	*297*[b]	∞	∞	0·031	98
412[b]	∞	∞	0·023	64	*412*	0	15·0	0	∞
copia	0	∞	0·000	∞	*copia*	0	31·0	0	∞

[a] Only the salivary divisions 21–32 and band 33A1 were used for 2L, due to the difficulty of identifying sites in the remainder of the distal section of this arm.
[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 $\hat{x}$ and m were estimated using a maximum likelihood procedure (Lewontin & Prout, 1956).

$x^{\alpha-1}(1-x)^{\beta-1}$ (Charlesworth & Charlesworth, 1983). The parameters α and β respectively measure the product of $4N_e$ and the probability of insertion of a newly-transposed element into a given chromosomal site, and the product of $4N_e$ and the rate of removal of elements from an occupied site in the population (by excision, selection, etc.). N_e is the effective size of a local breeding population. The magnitudes of the forces affecting element frequencies may therefore be quantified by estimates of these two parameters for each element family, obtained by fitting the observed occupancy profiles to the theoretical distribution.

https://doi.org/10.1017/S0016672300030792 Published online by Cambridge University Press

Table 7. *Parameters of the probability distributions of element frequencies for the distal sections of the third chromosome of* D. melanogaster *sampled from a Maryland population*

	3L					3R			
Element	α	β	$\hat{x}$	m	Element	α	β	$\hat{x}$	m
roo	0·8	15·0	0·051	170	*roo*	1·2	14·5	0·076	159
2156[a]	∞	∞	0·017	51	*2156*[a]	∞	∞	0·008	203
2158[a]	∞	∞	0·015	62	*2158*[b]	0	∞	0	∞
2161	0	9·5	0	∞	*2161*	0	4·0	0	∞
2181[a]	∞	∞	0·015	119	*2181*	0	8·5	0	∞
2217[a]	∞	∞	0·015	165	*2217*	0	16·5	0	∞
297	0	18·0	0	∞	*297*	0	18·0	0	∞
412[a]	∞	∞	0·023	106	*412*[a]	∞	∞	0	119
copia	6·0	200	0·029	106	*copia*	0	6·5	0	∞

[a] 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 $\hat{x}$ and m were estimated using a maximum likelihood procedure (Lewontin & Prout, 1956).
[b] 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.

These estimates for each element for the sampled autosomal arms are shown in Tables 6 and 7. 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 arm 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 some cases, all sites had a maximum occupancy of one, so that α is estimated as zero and β as infinity. In the case of finite estimates of α and β, the expected element frequency per occupable salivary chromosome band ($\hat{x}$) is estimated as $\alpha/(\alpha+\beta)$, and the total haploid number of occupable sites on an arm (m) is estimated as the mean copy number per chromosome divided by $\hat{x}$ (Charlesworth & Charlesworth, 1983).

4. Discussion

(i) *Distributions of copy numbers and linkage disequilibrium*

The general picture revealed by these data on the distribution of numbers of elements per chromosome for autosomal arms is very similar to that found for X chromosomes sampled from the same population (Charlesworth & Lapid, 1989), and for the few other studies of autosomal elements (Charlesworth & Langley, 1989). For the distal section of a given chromosome arm, members of a given element family are distributed in accordance with the Poisson distribution, suggesting that causes of departure from the Poisson distribution (such as variation between sites in element frequencies, and linkage disequilibrium) are too small to be detected.

Direct tests for linkage disequilibrium among members of the same family of elements yielded little evidence for any significant linkage disequilibrium among members of the same family, with only two exceptions [section 3(iii)]. Both involved elements in proximal regions of the chromosomes, where recombination is restricted in frequency (Ashburner, 1989, chap. 11). Similarly, there were significant, positive correlations in numbers of copies of different elements only for contrasts involving the proximal regions of the third chromosome. These correlations indicate a general tendency toward positive linkage disequilibrium among members of different families in the proximal euchromatin. The one case of significant linkage disequilibrium between members of different families in the distal sections of the arms (for *2161* and *2217* on 3L) is an anomaly, particularly as no such effect was detected for comparisons of adjacent occupied sites, where disequilibrium is likely to be strongest. It is presumably a statistical outlier.

Obviously, with the small samples we have studied, the ability to detect linkage disequilibrium is limited. As far as they go, the results suggest that the evolutionary forces affecting element frequencies are sufficiently weak that haplotype frequencies are close to random expectation for the distal sections of the chromosomes, where recombination is relatively free. The most likely interpretation of the positive associations between elements in the proximal chromosome regions is that they are generated by hitchhiking events. Since recombination is restricted here,

https://doi.org/10.1017/S0016672300030792 Published online by Cambridge University Press

variation may be affected by events at physically distant loci. On this view, haplotypes with several element insertions have been dragged to high frequencies by alleles at linked loci which have increased in frequency rapidly as a result of selection (Maynard Smith & Haigh, 1974; Thomson, 1977; Kaplan, Hudson & Langley, 1989; Stephan, Wiehe & Lenz, 1992). The other possible explanations for the linkage disequilibria are selection or genetic drift. But the form of selection needed to stabilize element frequencies (a negative second-derivative of log fitness with respect to element copy number: Charlesworth & Charlesworth, 1983) generates negative rather than positive linkage disequilibrium between elements (Charlesworth, 1990). The effect of genetic drift is to generate linkage disequilibria that are random in sign (Robertson & Hill, 1983), which is unlikely to produce the positive correlations in copy numbers detected here, which reflect the sum of the disequilibria across many pairs of sites.

Hitch-hiking is thus the only viable candidate to explain the disequilibrium detected. The sign and magnitude of the linkage disequilibrium generated by selection at a pair of linked neutral loci depends in a complex way on the frequencies of the alleles at the loci in question (Thomson, 1977). Hitch-hiking events that cause changes in the frequencies of TE insertions at linked sites could thus result in disequilibria that are positive on average, depending on the accidents of which haplotypes were initially associated with selectively favoured alleles. Recent studies of molecular variation in regions of restricted recombination in *Drosophila* suggest that there is often a considerable reduction in nucleotide site variability, as expected if hitch-hiking has been occurring (Begun & Aquadro, 1992).

(ii) *Distribution of elements along the distal chromosome arms*

As in the case of the X chromosomes studied by Charlesworth & Lapid (1989), elements were nearly always distributed randomly along the distal sections of the autosomal arms, as far as the positions of the sites which were occupied at least once are concerned [section 3(ii)]. In addition to providing further evidence against linkage disequilibrium in the distal sections of the chromosomes, this suggests that the sites of new insertions must be effectively independent of the presence of elements belonging to the same family at neighbouring sites (Charlesworth & Lapid, 1989).

The existence of some type of interaction between members of different families is, however, indicated by the correlations between different families with respect to the identities of distal sites that they occupy at least once in the sample (Table 3). This effect was also found for the X chromosome (Charlesworth & Lapid, 1989). As for the X chromosome, these correlations do not seem to reflect common tendencies to insert into the same sites on individual chromosomes, since tests of pairwise linkage disequilibria between elements from different families with respect to identical sites yielded no significant effects [section 3(iv)]. In contrast to the findings for the X chromosome, there was some evidence that different families were correlated with respect to the identities of distal sites where elements attained relatively high frequencies. (The definition of high frequency was an occupancy of not less than 2 or 3, depending on the chromosome arm and the element in question.) The combined results for all arms (44 pairs of sites) yielded one association with $P < 0{\cdot}001$, two with $P < 0{\cdot}01$, and five with $P < 0{\cdot}05$, on Fisher's one-tailed exact test. These are clearly too numerous to be the result of chance. The mean correlations and their standard errors were as follows: $0{\cdot}165 \pm 0{\cdot}114$ (2L, $n = 3$), $0{\cdot}143 \pm 0{\cdot}032$ (2R, $n = 10$), $0{\cdot}068 \pm 0{\cdot}038$ (3L, $n = 10$) and $0{\cdot}134 \pm 0{\cdot}024$ (3R, $n = 21$).

The positive mean correlations between families for both the identity of occupied bands and of bands occupied at relatively high frequency suggest that there may be common affinities of different element families for insertion into certain bands, e.g. because of different physical sizes of the bands. Alternatively, differences between sites with respect to the rates of elimination of elements from the population could be involved. The reason for the difference from the X chromosome results for the high frequency sites is unclear; it may simply reflect the fact that only nine pairs of correlations could be estimated in that case (Charlesworth & Lapid, 1989).

There is no contradiction between the findings of a random distribution of occupancy and correlations between pairs of families, provided that the inhomogeneities that cause the latter are themselves distributed at random along the chromosome.

(iii) *Distribution of element frequencies*

Apart from four sites with apparent fixation of elements, element frequencies in the distal sections of the autosomes are low, suggesting that some force or forces are responsible for holding element frequencies to a low level. The largest estimate of the expected element frequency for a TE family is 0·092 for *roo* on 3R. As in the case of the X chromosome (Charlesworth & Lapid, 1989), higher element frequencies are often observed at the bases of the chromosomes. Possible reasons for this are discussed in the accompanying paper (Charlesworth *et al.* 1992). The following discussion is therefore confined to the information provided by the frequency data for the distal sections of the chromosomes, notably the estimates of the parameters α and β of the probability distribution of element frequencies (Tables 6 and 7).

While the individual parameter estimates are subject to considerable uncertainty, there is no question that the value of β is always highly significantly different

from zero. In most cases, β is at least an order of magnitude greater than unity. The harmonic mean of β across all element families is 23·4 (the harmonic mean is used to cope with the cases when β is infinite or very large, which tend to be the low copy number elements for which the data are least informative). Since this parameter measures the product of $4N_e$ and the sum of the magnitudes of the systematic forces tending to remove elements from the population, it is clear that the distribution of element frequencies requires that the magnitude of these forces be much greater than that of genetic drift.

Following Charlesworth & Lapid (1989) and Charlesworth & Langley (1989), this value of mean β can be combined with the estimate of N_e for an East Coast population of *D. melanogaster* (Mukai & Yamaguchi, 1974), to estimate the magnitude of the forces removing elements from the population. If the population is approximately in equilibrium, this should be equal to the mean rate of transposition per element per generation, u (Charlesworth & Lapid, 1989). The estimate of the product of the rate of gene flow into the population and $4N_e$ must be deducted from the estimate of β in order to calculate the component due to other forces, and hence $4N_e u$; from Singh & Rhomberg (1987), this product is 8·8 for autosomal loci. Hence, $4N_e u = 14{\cdot}6$. Using the estimate of N_e of 20000, we have $u = 1{\cdot}8 \times 10^{-4}$, compared with the value of 4×10^{-4} estimated from X chromosome data by Charlesworth & Lapid (1989). The present estimate is more accurate, since it is based on four chromosome arms instead of one. It is consistent with the small number of published laboratory estimates of rates of movement of non-dysgenic *Drosophila* elements (Charlesworth & Langley, 1989).

One possible reservation about these conclusions that needs discussion comes from the fact that we are using polytene bands as the unit for measuring element frequencies. As has been pointed out before (Kaplan & Brookfield, 1983), multiple occupancy of a polytene band may not reflect multiple occupancy at the level of nucleotide sites, since the number of nucleotide sites per band is very large. Insertions within the same band may thus be at different sites. Restriction-map surveys indeed show that TE insertions are nearly always at unique sites within a sample of chromosomes (Charlesworth & Langley, 1989). At first sight, this implies that the estimates of β discussed above are seriously dowardly biased. The analysis given in the Appendix shows, however, that (under quite light conditions) the use of data in which TE frequencies are estimated by collapsing many different sites into a single unit does not, in fact, lead to a bias in β.

Inspection of Tables 6 and 7 suggests that there is considerable heterogeneity between families in the value of β, but that β estimates for a given family are fairly consistent across chromosome arms. The statistical significance of this heterogeneity was assessed as follows. There is an indication from the Tables that the value of β for a TE family is inversely related to its abundance as measured by mean copy number in the distal sections of the chromosomes (Table 1). In order to remove infinite values, and to improve the linearity of the relation, $1/\beta$ and the natural logarithm of the reciprocal of distal mean copy number for a given family for a given autosomal arm were used as variables. There is a significant negative correlation between the two (Kendall's $\tau = -0{\cdot}455$, $P < 0{\cdot}001$). This relation is probably artefactual, reflecting the fact that elements with low copy numbers tend to produce occupancy profiles with maximum occupancies of 1 or 2, which lead to infinite β values. If the cases with zero values of $1/\beta$ are deleted, the correlation is no longer significant, although still negative ($\tau = -0{\cdot}292$, $P = 0{\cdot}07$).

There is, however, evidence for heterogeneity among families with similar copy numbers with respect to β. For example, if the high copy number families *roo* and *2161* are compared, neither of which have any infinite β values, a Mann–Whitney U test gives a probability of 0·02. The significance of the heterogeneity among β values was tested for the complete autosomal data set by a Kruskal–Wallis non-parametric analysis of variance of the residuals of $1/\beta$ from the regression of $1/\beta$ on ln mean copy number, treating element families as groups. The H statistic was 17·6 ($P = 0{\cdot}024$). This suggests that there is indeed heterogeneity among TE families with respect to the value of β, even after adjusting for the effect of copy number. This implies variation between TE families in the rate of transposition.

As already mentioned, there were four cases of apparent fixation of elements in our data, each involving a separate TE family (Tables 4 and 5). One of these (*2158* on 3L) is in the proximal region of the chromosome, where element frequencies are often high. This may thus represent a case of a genuinely high frequency. The other cases, involving distal sites, are is sharp contrast to the low frequencies of elements generally observed. As mentioned in section 2(iii), we are at present unable to exclude the possibility that these cases of fixation are artefacts of hybridization of flanking sequences in the probes with the original cloning sites, and have proceeded with the analyses on the assumption that this is indeed the case. Since over 1500 polytene bands were scored as segregating for elements in the set of sampled X chromosomes and autosomes, the numerical effect of discarding the fixed sites is very small, and does not materially affect our quantitative conclusions. It is, of course, possible that these cases may involve genuine fixation or high frequencies of elements, due to selection for beneficial mutations induced by insertion of the elements concerned (McDonald, 1990), or to hitch-hiking with closely-linked favourable mutations. We plan to carry out experiments to distinguish between these possibilities in the near future.

The results shown in Tables 6 and 7 can also be used to ask whether or not there is any limitation on the numbers of sites into which elements can insert, at the level of the polytene chromosome bands resolved in this study. Since the individual estimates of m, the number of bands for the chromosome region in question, are unreliable, the harmonic means of m across TE families were calculated for each chromosome arm, and compared with the relevant numbers of bands. The results were as follows, showing the estimated and actual band numbers as the first and second members of each bracketed pair: 2L (64, 80), 2R (120, 139), 3L (120, 145), 3R (459, 180). There is little indication here of any major differences between the numbers of bands available to TEs for insertion and the numbers of bands in the distal sections of the autosomal arms. This agrees with the results on the distributions of occupancies of TEs along the distal sections of the arms, which are essentially random (Table 2). Similar results were obtained earlier for the X chromosome (Charlesworth & Lapid, 1989). Elements thus appear to be capable of inserting into virtually any band in the distal sections of the major chromosomes.

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.

Appendix

Effects of pooling element frequency information across different sites

Consider the situation when the population frequency x_i of elements belonging to a given family of TEs at polytene band site i is the sum of the frequencies of elements at the array of nucleotide sites within the band in question, i.e. $x_i = \Sigma_j x_{ij}$, where x_{ij} is the frequency of elements at the jth site in band i. If element frequencies at all nucleotide sites are low, and the effects of element insertions on fitness are independent of chromosomal location, then the deterministic change per generation in x_{ij} is given by the standard equation (Charlesworth, 1985)

$$\Delta x_{ij} \approx x_{ij}\frac{\partial \ln \bar{w}}{\partial \bar{n}} + \frac{u\bar{n}}{2m} - vx_{ij}, \qquad \text{(A 1)}$$

where $\bar{n}$ is the mean copy number per diploid individual, $\bar{w}$ is the mean fitness of the population, v is the rate of excision of elements from insertion sites, and the other variables have been defined in the text.

Summing over all sites within band i, we obtain the expression for the change in net element frequency for this band as

$$\Delta x_i \approx x_i\left(\frac{\partial \ln \bar{w}}{\partial \bar{n}} - v\right) + \frac{m_i u\bar{n}}{2m}, \qquad \text{(A 2)}$$

where m_i is the number of nucleotide sites in band i. If m_i is approximately constant across bands, this equation implies that the deterministic expression for element frequency per band has the same form as the more basic expression for an individual nucleotide site, substituting the total number of bands for the total number of sites in the term involving the rate of transposition u.

From the properties of the multinomial distribution, the variance in the change in element frequency at the jth site in band i caused by drift at a particular nucleotide site is $x_{ij}(1-x_{ij})/2N_e$ and the covariance between changes at two different sites j and k is $-x_{ij}x_{ik}/2N_e$. Provided that the frequencies per site and band are sufficiently low, the covariances can thus be neglected compared with the variance, and the variance in the change in the net element frequency for band i is approximated by $x_i/2N_e$.

This proves that, under suitable conditions, both the deterministic and stochastic changes in the element frequencies per band can be approximately described by equations of the same form as for individual sites. The condition of low element frequencies per site and band seems to be met in the data analyzed here (see Tables 4 and 5). It is less clear that the number of sites per band is uniform across bands, as assumed here. However, the exact form of the probability distribution of element frequencies is only important for cases when the parameters α and β are to be fitted by the minimum χ^2 technique of Charlesworth & Charlesworth (1983). The results of this invariably yields estimates of α that are close to zero (Tables 6 and 7). This suggests that variation between bands in the value of α ($= 4N_e m_i u\bar{n}/2m$) is unlikely to have a major effect on the estimates of the parameters, since it will be close to zero for all bands.

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.

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.

Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Genetics*. Oxford: Oxford University Press.

Charlesworth, B. (1990). Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**, 199–221.

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. (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*. II. Inferences on the nature of selection against elements. *Genetical Research* **60**, 115–130.

Finnegan, D. J. (1989). F and related elements in *Drosophila melanogaster*. In *Mobile DNA* (ed. D. E. Berg and M. M. Howe), pp. 519–521. Washington, D.C.: American Society of Microbiology.

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

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

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

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

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

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.

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

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

McDonald, J. F. (1990). Macroevolution, gene regulation and retro-viral like elements. *Bioscience* **40**, 183–191.

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

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.

Mukai, T. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila*. XI. Genetic variability in a natural population. *Genetics* **82**, 63–82.

Robertson, A. & Hill, W. G. (1983). Population and quantitative genetics of many linked loci in finite populations. *Proceedings of the Royal Society of London* **B219**, 253–264.

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

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**, 313–322.

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

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.

Strobel, E. (1982). Mobile dispersed repeated DNA elements in the *Drosophila* genome. *Federation Proceedings* **41**, 2656–2658.

Thomson, G. (1977). The effect of a selected locus on linked neutral loci. *Genetics* **85**, 789–814.