Genet. Res., Camb. (1994), **63**, *pp.* 197–212 *With 2 text-figures* Copyright © 1994 Cambridge University Press

Population dynamics of the copia, mdg1, mdg3, gypsy, and P transposable elements in a natural population of *Drosophila melanogaster*

C. BIÉMONT[1]*, F. LEMEUNIER[2], M. P. GARCIA GUERREIRO[1], J. F. BROOKFIELD[3], C. GAUTIER[1], S. AULARD[2] AND E. G. PASYUKOVA[4]

[1] *Laboratoire de Biométrie, Génétique, Biologie des populations, URA C.N.R.S. 243, Université Lyon 1, 69621 Villeurbanne Cedex, France*
[2] *Laboratoire de Biologie et Génétique évolutives, C.N.R.S., 91198, Gif-sur-Yvette Cedex, France*
[3] *University of Nottingham, Department of Genetics, Queen's Medical Centre, Nottingham NG7 2UH, UK*
[4] *Institute of Molecular Genetics, Russia Academy of Science, Kurchatov Sq. 46, 123182, Moscow, Russia*

(*Received 8 October 1993 and in revised form 7 January 1994*)

Summary

The insertion site polymorphism of the copia, mdg1, mdg3, gypsy, and P transposable elements was analysed by *in situ* hybridization to the polytene chromosomes in genomes of males from a natural population of *Drosophila melanogaster*. Parameters of various theoretical models of the population biology of transposable elements were estimated from our data, and different hypotheses explaining TE copy number containment were tested. The copia, mdg1 and gypsy elements show evidence for a deficiency of insertions on the X chromosomes, a result consistent with selection against the mutational effects of insertions. On the contrary, mdg3 and P copy numbers fit a neutral model with a balance between regulated transposition and excisions. There is no strong evidence of a systematic accumulation of elements in the distal and proximal regions of the chromosomes where crossing over and ectopic exchanges are reduced. For all chromosome arms but 3L, however, the TE site density increases from the proximal to the distal parts of the chromosomes (the centromeric regions were excluded in this analysis) with sometimes a sharp decrease in density at the extreme tip, following in part the exchange coefficient. The way the copy number of TEs is contained in genomes depends thus on the element considered, and on various forces acting simultaneously, indicating that models of TE dynamics should include details of each element.

1. Introduction

The discovery that transposable elements (TEs) are common genomic components of many organisms has stimulated debate about the nature of the forces that affect their distributions in populations. Theoretical and empirical studies yield the conclusion that these transposable elements are maintained in populations as a result of a balance between the transpositional increase in copy number and some opposing forces. These forces include the regulation of the rate of transposition with increasing copy number, selection against insertional mutations, and induction of deleterious chromosome rearrangements by ectopic exchange. Thus far, however, considering the data acquired in *Drosophila*, it seems impossible to decide definitively which of these forces plays the major rôle in maintaining copy number in populations (see the various syntheses published on different aspects of the TE characteristics: Brookfield, 1986; Charlesworth, 1985, 1988; Ajioka & Hartl, 1989; Echalier, 1989; Finnegan, 1989; Mackay, 1989; Biémont, 1992).

Although the accumulation of data on insertion site polymorphism in *Drosophila* has made some statistical analyses possible, most of the data are restricted to the X chromosomes (Kaplan & Brookfield, 1983; Montgomery & Langley, 1983; Ronsseray & Anxolabéhère, 1986; Leigh-Brown & Moss, 1987; Charlesworth & Lapid, 1989). There have been a few studies in which the sites of insertion were localized on all the major chromosomes, for a number of elements; they were done mainly on inbred lines or laboratory strains (Strobel, Dunsmuir & Rubin, 1979; Belyaeva, Ananiev & Gvozdev, 1984; Biémont, 1986; Biémont & Gautier, 1988; Biémont, Gautier & Heizmann, 1988; Pasyukova *et al.* 1988; Biémont *et al.* 1990*b*), or in small samples of second, third and fourth chromo-

* Corresponding author.

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

somes extracted from a natural population in different years and maintained en masse as laboratory stocks (Charlesworth, Lapid & Canada, 1992*a*, *b*). All other studies with wild populations assessed copy numbers only and did not establish the site positions (Montgomery, Charlesworth & Langley, 1987; Yamaguchi *et al.* 1987; Ronsseray, Lehmann & Anxolabéhère, 1989). Here, we report the localization of insertions for four transposable elements of the copia-like type (copia, mdg1, mdg3, and gypsy), and for the P element, on all the chromosome arms of individuals from a natural population of *Drosophila melanogaster*. The distribution of elements across the chromosomes and its various associated parameters (mean and variance of insertion site numbers, site occupancy frequencies) were analysed and used to estimate parameters of various theoretical models of transposable element evolutionary dynamics (Charlesworth & Charlesworth, 1983; Kaplan & Brookfield, 1983; Langley, Brookfield & Kaplan, 1983; Charlesworth, 1985; Kaplan, Darden & Langley, 1985; Charlesworth & Langley, 1986; Langley *et al.* 1988; Stephen & Langley, 1992).

2. Materials and methods

(i) Drosophila *strains*

Line 16: a highly inbred line maintained by brother–sister mating (except for some generations when the flies were maintained in small mass cultures to build up their numbers and to increase viability). At the 146th and 165th generations larvae were analysed and female flies crossed with wild males from the Valence natural population.

Gruta strain: a long-established laboratory strain that is devoid of P elements and has an M cytotype according to the nomenclature used in the hybrid dysgenesis phenomenon (it lacks the ability to regulate P movements). This Gruta strain was used to determine the status of the Valence population with respect to the P–M system of hybrid dysgenesis (see below).

Valence population: wild adult flies and larvae were collected in 1990 and 1991 near an orchard in Valence (Drôme, France). The sample collected in October 1990 was analysed for the two elements copia and mdg1, the sample collected in October 1991 was analysed for the mdg3 and P elements. Flies were collected in a single trapping, both in 1990 and 1991. The status of the Valence population in the P–M system of hybrid dysgenesis was found to be M′; it had an M cytotype as determined by the standard crosses proposed by Kidwell (1983) using the true M strain Canton S (males Valence crossed with females Canton S gave 3·5% F1 female sterility) and the strong P inbred strain Harwich (males Harwich crossed with females Valence gave 20% F1 female sterility).

(ii) *Mating system*

Males were crossed with virgin females from the inbred line 16 (to analyse copia, mdg1, mdg3 and gypsy location) or gruta (to analyse P element location). The pairs so formed were set up individually in small vials containing axenic medium and were allowed to lay eggs. The insertion site patterns of the hybrid F1 larvae analysed (we analysed four to ten larvae per male) were thus the combination of the two parental patterns. The different elements were examined in progeny from different males. The insertions were localized to numbered and lettered subdivisions of the chromosomes, following the Lefevre photographic map. Because many larvae were analysed, it was possible to compare replicate slides of the same genotype so as to limit error in assigning element positions. The insertion profiles of the males were obtained by subtracting the characteristic profile of the highly inbred line 16 from that of the hybrid larva (this step was not necessary in crosses involving the gruta strain which was devoid of P elements). Because there is no recombination in *Drosophila* males, we recovered intact chromosomes. The only minor drawback with this technique is that bands common to both line 16 and the wild male under study are excluded from the insertion profile of the male. However, by analysing copia insertions in larvae resulting from some crosses between males from the wild and females from both the gruta line and line 16, we estimated that 0 to 2 bands were common to line 16 and each of the males; this cannot greatly perturb either the global view of the genome or the statistical analysis. Although the common sites may appear as 'cold spots', this concerns only a few individual bands and not large regions of the chromosomes.

We chose this particular type of cross instead of the classical chromosome extraction procedure with balancer strains for two reasons: (1) we cannot eliminate the possibility that crossing wild males with marker strains may mobilize the elements under study, making the insertion profile estimate of the wild males inaccurate. Such interstrain cross mobilization during the process of extracting homozygous chromosomes of wild males has been actually observed not only with P elements but with copia-like elements as well (gypsy and copia: Mevel-Ninio, Mariol & Gans, 1989; mdg1, mdg3 and copia: Pasyukova *et al.* 1988; stalker: Georgiev *et al.* 1990; Ulysses: Scheinker *et al.* 1990), (2) the procedure of extracting homozygous chromosomes is slow and difficult when the 2 major autosomes of individual males are to be made homozygous simultaneously.

(iii) *Probes used*

We used the probe cDm5002 containing the copia element (5 kb; Dunsmuir *et al.* 1980; Levis, Dunsmuir

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

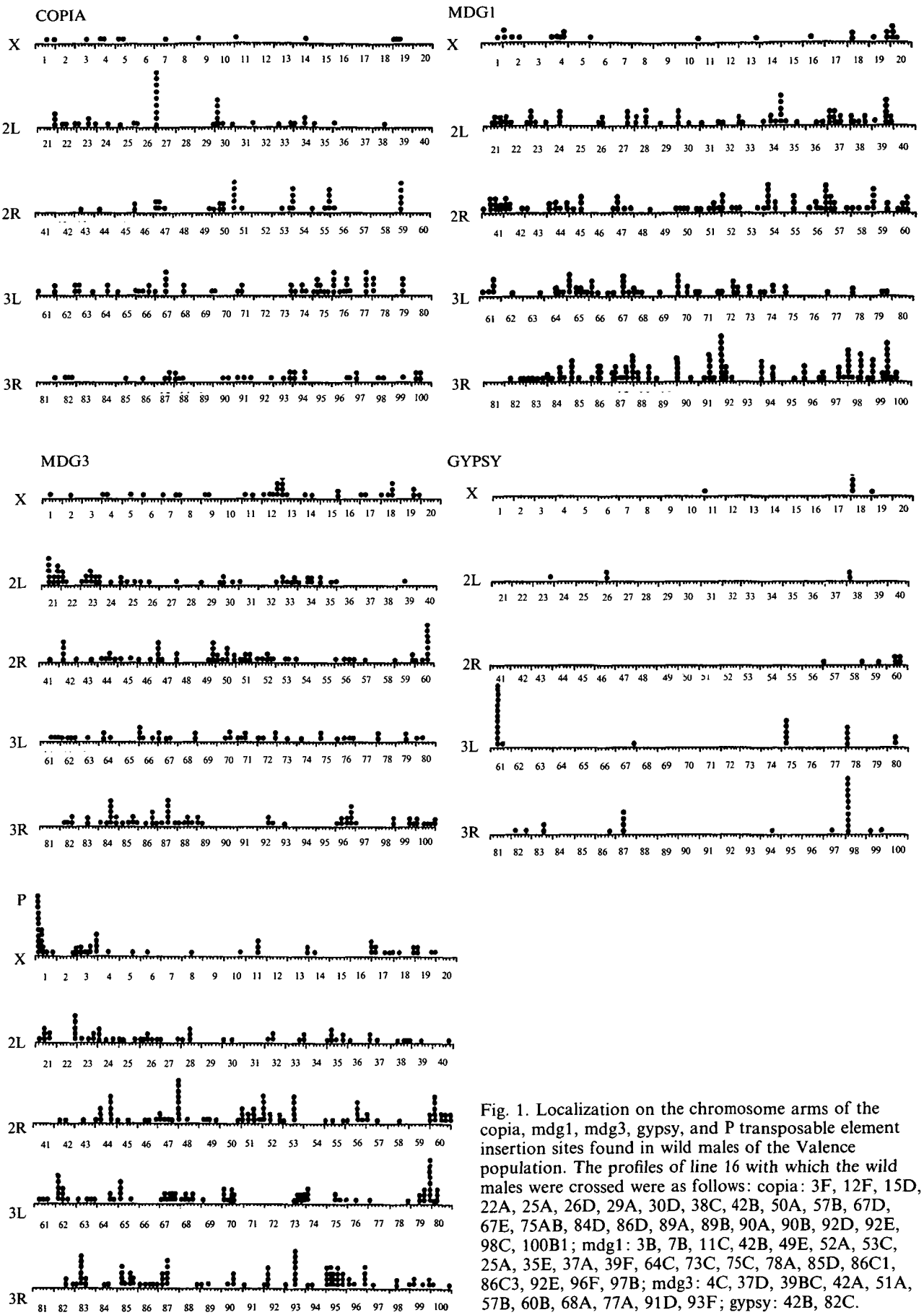

Fig. 1. Localization on the chromosome arms of the copia, mdg1, mdg3, gypsy, and P transposable element insertion sites found in wild males of the Valence population. The profiles of line 16 with which the wild males were crossed were as follows: copia: 3F, 12F, 15D, 22A, 25A, 26D, 29A, 30D, 38C, 42B, 50A, 57B, 67D, 67E, 75AB, 84D, 86D, 89A, 89B, 90A, 90B, 92D, 92E, 98C, 100B1; mdg1: 3B, 7B, 11C, 42B, 49E, 52A, 53C, 25A, 35E, 37A, 39F, 64C, 73C, 75C, 78A, 85D, 86C1, 86C3, 92E, 96F, 97B; mdg3: 4C, 37D, 39BC, 42A, 51A, 57B, 60B, 68A, 77A, 91D, 93F; gypsy: 42B, 82C.

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

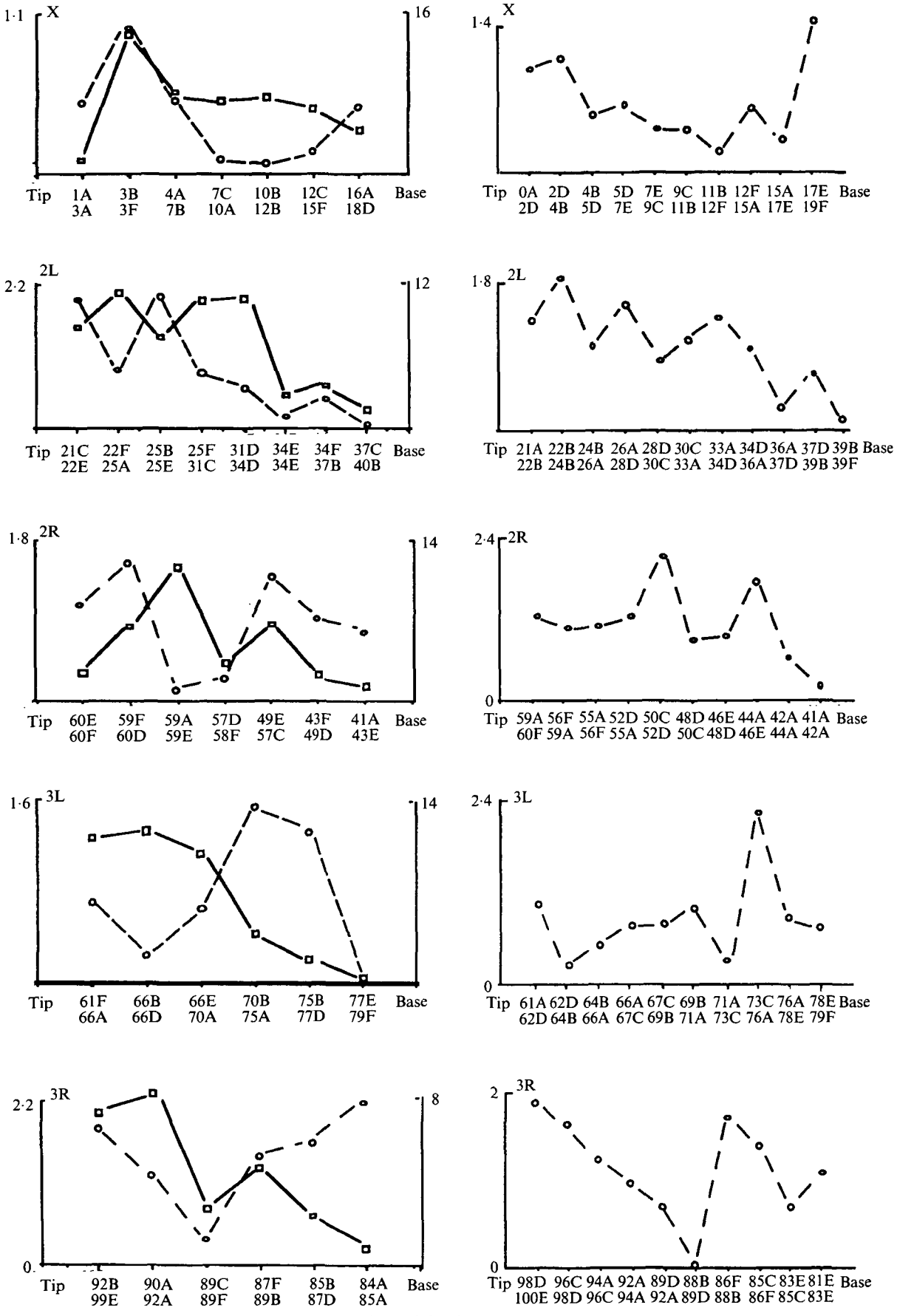

Fig. 2. TE density over chromosome arms. Left part: distribution of TE density (all elements pooled) and coefficient of exchange along the chromosome arms X, 2L, 2R, 3L, 3R. The regions are those defined in Lindsley & Sandler (1977). For the sake of simplicity the regions of the chromosomes were represented on the figures as if they had the same size,

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

& Rubin, 1980); the A fragment of the mdg1 element incorporated at the *Hind* III site of the pBR322 plasmid (Ilyin, Chmeliauskaite & Georgiev, 1980; Tchurikov *et al.* 1981); the *Hind* III fragment of the mdg3 element in pBR322 (probe Dm86: Tchurikov *et al.* 1981); the clones Dm111 containing the full-length gypsy element incorporated at the *Bam*H 1 site of pBR322 (Bayev *et al.* 1984); the plasmid pπ 25·1 (O'Hare & Rubin, 1983) to detect all kinds of P elements.

(iv) In situ *hybridization*

Polytene chromosome spreads from salivary glands of third instar larvae were treated with nick-translated, biotinylated DNA probes, as described in Biémont & Gautier (1988). Insertion sites were visualized as brown bands resulting from a dye-coupled reaction with peroxidase substrate and diaminobenzidine, the hybridization signal being amplified by silver grains (DAB enhancement kit of Amersham) when necessary. Genomic DNA associated with the copia and P probes, hybridized to the 5A and 17C regions on the X chromosome, respectively; a signal at these regions was thus used as positive control indicator of hybridization for these elements.

3. Results

(i) *Distribution of element insertion sites on chromosome arms*

Figure 1 shows the distribution of the insertion sites along the chromosomes from 20, 24, 36, 51, and 20 autosomes, and from 10, 14, 25, 51, and 12 X chromosomes for the elements copia, mdg1, mdg3, gypsy, and P respectively. At first sight it appears that the distributions are not uniform in that some chromosomes, some regions, and some sites have more insertions than others. For example, the 1A and 1B sites of the X chromosomes have many P insertions, as already reported in various populations and inbred lines (Ajioka & Eanes, 1989; Biémont *et al.* 1990*a*). This region corresponds to the insertions of specific P elements capable of eliciting regulation of the P elements (Ronsseray, Lehmann & Anxolabéhère, 1991). Gypsy is another element with apparent high frequency of insertion in regions 61C and 98A. We cannot, however, eliminate the hypothesis of a very small number of potential insertion sites for this element and thus an accumulation in some sites, an hypothesis sustained by the low number of gypsy insertions per haploid genome (Table 2). Indeed, as reported many times, gypsy sequences hybridize to only few bands on the euchromatin but with great intensity to the chromocenter in which non-autonomous copies seem to accumulate (Modolell, Bender & Meselson, 1983; Bayev *et al.* 1984; Peifer & Bender, 1988; Lyubomirskaya *et al.* 1990).

To analyse the distribution of TE insertions over the chromosomes, and to test whether it is related to recombination rate as postulated in some analyses (Langley *et al.* 1988; Charlesworth & Langley, 1991; Charlesworth, Lapid & Canada, 1992*a*, *b*), we have correlated the number of insertion sites along the chromosome with the recombination rates as calculated by Lindsley & Sandler (1977). It is well known that the pericentromeric regions of the chromosomes contain a high proportion of middle repetitive DNA sequences (Ananiev *et al.* 1984; Belyaeva, Ananiev & Gvozdev, 1984) which are in general restricted to the β-heterochromatin (Miklos *et al.* 1988), which seems to be a preserve for many non-mobile ancient copies of TEs (Vaury, Bucheton & Pélisson, 1989). Such an accumulation of insertions in the centromeric region is easily seen by *in situ* hybridization with all our TEs probes except P. Note also that for some elements insertions exist in pericentromeric regions in fixed positions (see Vaury, Bucheton & Pélisson, 1989 for the I element). We thus have omitted the regions 20, 40, 41, 80, and 81 from our analysis and considered only the euchromatic regions studied for recombination in the work of Lindsley & Sandler (1977). Each chromosomal region was associated with its density of TE insertion sites (the number of sites divided by the amount of DNA as determined by Bolshakov, Zharkikh & Zhimulev (1985)). We used the number of sites instead of the total number of insertions in the sites (a site is thus represented only once even if an insertion at that site is detected in various chromosomes: see the site distribution in Fig. 1) so as to eliminate any bias due to high site frequencies resulting from inbreeding, drift or preferential insertion into specific regions. No significant correlation was detected between the TE densities and the coefficient of exchange, although for some elements there is a tendency for both criteria to follow a similar trend when they are represented along the chromosome arms. As seen in Fig. 2, which represents the value of coefficient of exchange and density of TEs (all elements combined) along the chromosome arms, the site density for some arms tends to decrease from the distal to the proximal parts of the chromosomes following the decrease of the coefficient of exchange. This tendency is particularly evident for copia as reported in Table 1 which shows the Spearman's rank

and from the tip of the chromosome to the base. Although this may change the visual appearance of the curves it does not change the values of the spearman's rank correlation coefficient calculated in Table 1. Right part: distribution of TE density (all elements pooled) over the chromosome. The regions of the chromosomes were determined by dividing the chromosome arms into 10 regions of close DNA content. See text for details. Exchange coefficient, □——□; TE site density, ○——○.

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

Table 1. *Spearman's rank coefficients for TE (copia, mdg1, mdg3, gypsy, P, and all elements pooled) site density versus chromosomal location*

	Chromosome arms			
Element	X	2L	3L	3R
copia	−0·88*	−0·94**	0·4	−0·6
mdg1	−0·39	−0·37	−0·10	0·00
mdg3	0·69	−0·83*	0·5	0·8
gypsy	—	—	−0·71	0·36
P	−0·32	−0·77*	−0·90*	0·50
all TEs	−0·57	−1***	−0·20	−0·40

The values associated with the pooled elements correspond to the curves represented in Fig. 2 (right part). We did not consider the data on the 2R arm because the regions analysed by Lindsley & Sandler (1977) were too small for statistical analysis to be reasonable. * $P < 0{\cdot}05$, ** $P < 0{\cdot}01$, *** $P < 0{\cdot}001$.

correlation coefficient of the TE site densities versus the localization on the chromosomes of the regions concerned. The 2L arm shows a clear negative correlation with all elements. There is no such evident tendency with the chromosome 3 and especially the 3L arm which shows this TE density decrease only for the P element. In this analysis we considered only the regions of the chromosomes as defined in Lindsley & Sandler (1977); the different sizes of their regions, which can even be reduced to a single chromosomal band, rendered the statistical analysis difficult and may explain the discrepancies in results observed between chromosomes and between TEs. We have thus reconsidered the analysis by dividing each chromosome arm into 10 parts with similar DNA content, independently of the Lindsley and Sandler study. The diagrams in the right part of Fig. 2 show unambiguously the decrease in TE density from the distal to the proximal parts of the chromosome arms, with a drastic increase near the proximal region for the X chromosome. The 3L arm does not follow this picture. Note that the diagrams on the left and right parts of the Fig. 2 should not be compared directly since the regions defined following Lindsley & Sandler (1977) considered in the left part are different to the regions considered in the right part which concerns the whole chromosome arms (except the regions 20, 40, 41, 80, 81). For example, the sharp decrease of TE density observed in the extreme tip of the X and 2R (see Fig. 2, left part) disappears when a larger portion of the chromosome is considered (Fig. 2, right part). Considering the number of copies instead of the number of sites (one site may have many copies distributed over the whole set of genomes analysed) gives the same results (data not shown).

(ii) *Copy numbers in genomes and site occupancy frequencies*

The total numbers of insertion sites on entire genomes differ greatly, ranging from 1·35 for gypsy to 23·80 for the P element in haploid gametes (Table 2). 3R appears to possess generally the highest number of insertions, and 2L and the X the smallest, with some minor deviations from this picture according to the element considered. As seen in Table 2, the variances in copy number in the haploid genomes are close to the mean values, and the haploid copy number does not deviate from a Poisson distribution as tested by chi-square tests.

The mean occupancy frequencies (total number of elements in sample per total number of occupied sites) for all the transposable elements are summarized in Table 3. For all the TEs the mean occupancy frequency is greater than one (except for the copia insertions on the X chromosome for which the

Table 2. *Tests of the Poisson distribution on each chromosome arm and on the entire genome of gametes for the copia, mdg1, mdg3, gypsy, and P transposable elements, on samples of gametes from the Valence natural population*

	Elements									
	copia		mdg1		mdg3		gypsy		P	
Chromosome arms	m	V_n	m	V_n	m	V_n	m	V_n	m	V_n
Observed values										
X	1·40	0·45	1·50	1·81	1·71	1·09	0·16	0·13	4·92	5·40
2L	2·05	1·63	3·50	5·50	1·58	1·39	0·10	0·09	3·53	2·14
2R	1·95	1·10	3·83	5·10	1·89	3·87	0·14	0·12	5·00	4·12
3L	3·25	2·83	3·18	1·87	1·40	1·19	0·49	0·45	4·78	1·48
3R	2·15	1·50	5·38	9·81	1·89	1·19	0·47	0·33	5·20	4·48
Entire haploid genome	10·75	9·67	17·67	20·73	8·49	8·67	1·35	1·03	23·80	16·07
d.f.	2		2		2		4		2	
χ^2	$0{\cdot}14^{ns}$		$0{\cdot}06^{ns}$		$1{\cdot}52^{ns}$		$0{\cdot}86^{ns}$		$0{\cdot}13^{ns}$	

m: mean number of insertion sites; V_n: variance of number of insertions sites. ns: not significant.

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

Table 3. *Mean occupancy frequencies (total number of elements in sample per total number of occupied sites) of the copia, mdg1, mdg3, and P elements on the chromosome arms*

Chromosome arm	Mean occupancy frequency	Variance of occupancy frequency
copia		
X	1·00	0·00
2L	1·64	3·26
2R	1·95	1·95
3L	1·78	0·92
3R	1·48	0·60
Total	1·62	1·39
mdg1		
X	1·33	0·36
2L	1·70	1·02
2R	1·72	1·11
3L	1·57	0·81
3R	2·29	2·85
Total	1·80	1·49
mdg3		
X	1·41	0·82
2L	1·67	0·90
2R	1·58	1·44
3L	1·32	0·28
3R	1·65	1·05
Total	1·54	0·92
gypsy		
X	2·25	3·58
2L	1·67	0·33
2R	1·4	1·44
3L	4·17	17·37
3R	2·3	7·56
Total	2·43	7·18
P		
X	1·90	3·76
2L	1·46	0·75
2R	1·98	2·89
3L	1·85	1·73
3R	2·17	2·98
Total	1·88	2·38

frequency is exactly 1·00), the highest value being observed for gypsy with an average of 2·43 elements per site on the whole genome (4·17 on the 3L arm). These values are in the same range as those reported in other studies for elements like copia, I, 297, 412, P (see Mackay, 1989; Charlesworth & Langley, 1991). The only exception is for gypsy which presents a small number of insertion sites with a high incidence of occupancy in some sites. Given the coarse resolution of the *in situ* hybridization technique, we cannot of course entirely eliminate the hypothesis that the multiple occupancy of some sites reflects, in fact, different insertions that are too close to be distinguished in the salivary gland chromosomes. Such a hypothesis cannot satisfactorily apply to sites seen in many individuals, as, for example, the sites 61C and 98A for gypsy, or the 1A region for the P element for which we have many reasons to suspect the insertion of only one or at most two P elements (Ronsseray, Lehmann & Anxolabéhère, 1991). Even if we suppose that all the gypsy insertions on the 61C site are in fact in different DNA sequences, we still have to explain why this 61C chromosomal region has so many insertions while the other regions of the chromosome are almost empty.

(iii) *Tests of models of containment of TE copy number*

Possible forces opposing the transpositional increase in copy number include the regulation of the rate of transposition with increasing copy number, selection against excessive insertional mutations, and induction of deleterious chromosome rearrangements by ectopic exchange. Under the hypothesis that element copy number is regulated by copy number dependent transpositions (the transposition rate per element declines with the number of copies) or excision independent of chromosomal location, an equiproportionality between the X and the autosomes is expected in the *Drosophila* genome. Alternatively, if elements are eliminated by natural selection against mutational effects of their insertion into the chromosomes, then the number of insertions on the X chromosome must be less than the number of insertions on the autosomes. Montgomery, Charlesworth & Langley (1987) and Langley *et al.* (1988) have modelled these neutral and negative selection hypotheses and predicted a frequency of elements on the X of 0·13 with selection against deleterious insertions, and of 0·20 if insertions are neutral and the copy number is stabilized by a balance between regulated transpositions and excisions. As seen in Table 4, the proportion of transposable element insertions on the X chromosomes varies greatly from one element to another and the accepted model of containment of TE copy number depends on the element, as shown by the chi-square test values. The mdg3 and P elements follow the neutral model while copia perfectly fits the selective model (the neutral model cannot be rejected, however, since the chi-square test value is not significant). The selective model may also apply to gypsy but the smaller number of insertions for this element does not allow the chi-square value to be significant for the neutral model (see Table 4). For mdg1, however, the situation is more complex: the number of insertions on the X is so small that even the selective model is rejected.

An alternative to this latter selected model is that selection could act against deleterious chromosome mutations produced by ectopic exchange between insertions of TEs in non-homologous regions of the chromosomes. Two unequal exchange models have been proposed (Langley *et al.* 1988). The first model postulates that all inserted elements can recombine only with any element located elsewhere in the

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

Table 4. *Tests of the proportions of transposable elements on X chromosomes compared to autosomes of* Drosophila melanogaster. *Chi-square tests of neutral and negatively selective containment of copy number and of two models of recombination UC1 and UC2 (see text)*

Element	Total Nb. of sites	Proportion on X†	Models			
			Neutral	Selected	UC1	UC2
copia	201	0·068	0·11	0·069	0·087	0·099
			(3·50)	(0·00)	(0·76)	(1·93)
mdg1	359	0·058	0·14	0·091	0·113	0·128
			(20·86)***	(4·49)*	(10·54)***	(15·42)***
mdg3	281	0·146	0·148	0·094	0·117	0·132
			(0·009)	(8·89)**	(2·30)	(0·46)
gypsy	68	0·117	0·20	0·13	0·16	0·18
			(0·18)	(0·14)	(0·90)	(1·80)
P	373	0·147	0·139	0·088	0·109	0·124
			(0·210)	(16·31)***	(5·44)*	(1·82)

† Because we have data on some diploid genomes, the sample size is different for the X chromosomes and the autosomes. So, to take this into account, the predictions of the proportions of sites found on the X chromosome were modified in the following way: if i represents the sample size for the X chromosome, and j the sample size for the autosomes, and y is the proportion of sites expected on the X, then the proportion of sites that we expect on the X becomes $y.i/(y.i+(1-y)j)$. The table gives the expected y value corresponding to the models; the chi-squared values are in parentheses. * $P < 0{\cdot}05$; ** $P < 0{\cdot}01$; *** $P < 0{\cdot}001$.

genome; under this hypothesis the proportion of insertions on the X chromosome is expected to be 0·16. A second model in which elements can recombine with other elements located in the same chromosomal region leads to a proportion of insertions on the X chromosome equal to 0·18. These two unequal exchange models were tested by chi-square tests. As expected in view of the results obtained with the neutral and selected models (see Table 4), the elements that appeared to be selected follow also the first ectopic exchange model UC1 (that gives the lowest number of insertions on the X chromosome). The 'neutral' elements are more in agreement with the second model (UC2, Table 4) for which the expected proportion of insertions on the X is 0·18, a value close to the 0·20 given by the neutral model. Because of the low power of the chi-square test and especially its sensitivity to sample size, it is difficult to compare reliably models that give close expectations, and thus the ectopic exchange models cannot be adequately tested in face of the more drastic neutral and selected hypotheses. However, another way of testing the ectopic exchange hypothesis is to note that insertions should accumulate in chromosomal regions of reduced crossing over such as the tip and base of the X chromosome and the base of the autosomes. As seen from Figure 2 and Table 1 our results do not show an accumulation of TE insertion sites on such regions, except maybe for the tip of the 3R arm although for this arm there is more a gradual decrease of TE density from the tip to the base than a particular high density at the tip. Additionally, the copia element, which appears to obey the selection model, does not show any evidence of accumulation of insertions in regions of reduced crossing over, but instead the copia site density follows the exchange coefficient value which decreases from the distal part of the chromosomes to the base and which is low at the tip of some chromosomes.

(iv) *Estimation of model parameters and expected occupancy profiles*

In the theoretical models of Charlesworth & Charlesworth (1983) and Langley, Brookfield & Kaplan (1983), the distribution of element frequencies across sites depends on n, the average copy number, and a parameter β (in the model of Charlesworth & Charlesworth, 1983) or θ (in the model of Langley, Brookfield & Kaplan, 1983) equal to $4N_e\nu$, with ν the rate of deletion per element per generation and N_e the effective population size. These two models assume that the total number of occupable sites is infinite. With a finite total number of potential sites, the distribution depends on a parameter $\alpha = 4N_e\mu$ (with μ the rate of transposition of an element to a site) which is the probability of transposition to a new site relative to the total number of occupable sites. The model parameters reported in Table 5 were estimated by different methods (see the Appendix). The values of β–θ vary from 5 for P to 26 for mdg3 with only a few variations among chromosome arms (we looked with chi-square tests for any significant evidence that the individual chromosomes are inconsistent with the mean value; we found that there is no significant evidence for variation in θ across the chromosomes), except for the gypsy element where the two sites with multiple occupations in the 3L and 3R arms greatly decrease the estimations of θ. Indeed, a low value of θ may reflect a small number of sites with multiple occupations, while a high value may reflect low occupancy frequencies, as is particularly evident for

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

Table 5. *Estimates of the parameters (θ, α and β) of the probability distribution of element frequencies for the copia, mdg1, mdg3, gypsy, and P elements from the Valence natural population. α and β are estimated simultaneously according to the equation of Charlesworth & Charlesworth (1983). $\theta_A, \theta_B, \theta_C$, are the same parameters as β but estimated in three different ways following the model of Kaplan & Brookfield (1983) in which α is 0 (see the Appendix for details)*

Element	Chromosome	α	β	θ_A	θ_B	θ_C
copia	X	—	—	∞	—	∞
	2L	—	—	9·08	—	11·75
	2R	—	—	6·45	—	7·75
	3L	—	—	7·92	—	9·91
	3R	—	—	18·41	—	31·45
	entire genome	0·00	9·79	—	9·79	—
mdg1	X	—	—	11·45	—	18·69
	2L	—	—	10·16	—	13·18
	2R	—	—	9·84	—	12·37
	3L	—	—	11·83	—	16·03
	3R	—	—	5·96	—	6·85
	mean†	—	—	9·85	—	13·42
	entire genome	0·60	13·48	—	8·8	—
mdg3	X	—	—	19·88	—	31·23
	2L	—	—	17·39	—	22·64
	2R	—	—	19·12	—	25·57
	3L	—	—	31·78	—	52·10
	3R	—	—	17·32	—	22·53
	mean†	—	—	21·10	—	30·81
	entire genome	0·44	26·33	—	19·5	—
gypsy	X	—	—	12·28	—	13·95
	2L	—	—	24·29	—	31·38
	2R	—	—	37·27	—	55·59
	3L	—	—	4·57	—	4·79
	3R	—	—	12·61	—	14·38
	mean†	—	—	18·20	—	24·02
	entire genome	0·35	24·03	10·96	18·61	12·29
P	X	—	—	3·57	—	4·28
	2L	—	—	10·45	—	14·91
	2R	—	—	5·71	—	6·85
	3L	—	—	6·15	—	7·48
	3R	—	—	5·22	—	6·05
	mean†	—	—	6·22	—	7·91
	entire genome	0	5·85	—	5·90	—

† mean of the θ estimates calculated on each chromosome arm.

mdg3. The θ value for copia (around 10) is lower than estimates obtained from other populations ($\theta = 17$ in the population analysed by Leigh-Brown & Moss, 1987; $\theta = 48{\cdot}3$ in a Raleigh population analysed by Kaplan & Brookfield, 1983), as also was our P value ($\theta = 6$) when compared to the estimate obtained from the data of Ronsseray & Anxolabéhère (1986) ($\theta =$ 16·6, see Mackay, 1989). The θ value is higher for the mdg3 element which is characterized by a low insertion site frequency. However, although there is great uncertainty in the estimations of the θ parameter, which are sensitive to sample size, the values of Table 5 are clearly higher than 1, the value expected if drift had been the main force explaining the containment of TE copy numbers. The α values are, however, not different from zero.

Table 6 shows the observed and expected occupancy distributions obtained with the parameter estimates of Table 5, for two models of TE copy number containment in which: (a) the total number of occupable sites is infinite, α is taken as zero, and $\beta (= \theta)$ estimated by the method A (see the Appendix), (b) α and β are estimated jointly by the method of Charlesworth & Charlesworth (1983). The fit between observation and prediction is very close for all the elements with the two models, implying that the infinite alleles model fits the data well and that estimating α and β simultaneously does not greatly improve the test (except for the gypsy element for which we started with two degrees of freedom with three classes, and then optimized two explanatory variables, α and β).

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

Table 6. *Analysis of the distributions of site occupancy frequencies of copia, mdg1, mdg3, gypsy, and P over the entire genome of individuals from the Valence natural population of* D. melanogaster. (*a*) *Expected distribution when* $\alpha = 0$, *and* β *is estimated by method A following the model of Kaplan & Brookfield (1983)*: (*b*) *expected distribution when* α *and* β *are estimated jointly by the method of Charlesworth & Charlesworth (1983)*

No. of occurrences	1	2	3	4	5	6	7	8	9	10	11	12	χ^2
copia													
No. of sites observed with given occupancy	83	25	5	5	3	0	0	0	1	0	0	0	
No. of sites expected													
(a)	84·66	20·87	8·49	3·94	1·98	1·02	0·52	0·28	0·13	0·07	0·02	0·00	2·42
(b)	81·13	23·98	9·34	4·03	1·81	0·82	0·37	0·16	0·07	0·02	0·01	0·00	2·50
mdg1													
No. of sites observed with given occupancy	126	46	29	9	6	1	1	1	0	0	0	0	
No. of sites expected													
(a)	134·61	44·01	19·26	9·51	5·03	2·78	1·58	0·92	0·54	0·32	0·19	0·11	7·66
(b)	124·18	51·23	23·26	10·81	5·02	2·30	1·02	0·44	0·18	0·07	0·03	0·01	2·28
mdg3													
No. of sites observed with given occupancy	122	43	7	5	4	0	1	0	0	0	0	0	
No. of sites expected													
(a)	126·47	34·00	12·32	5·05	2·22	1·02	0·47	0·21	0·11	0·05	0·02	0·00	4·92
(b)	122·13	37·62	13·47	5·09	1·96	0·76	0·29	0·11	0·04	0·01	0·01	0·00	4·24
gypsy													
No. of sites observed with given occupancy	15	6	0	2	2	0	0	0	0	1	0	1	
No. of sites expected													
(a)	14·44	5·15	2·54	1·47	0·93	0·62	0·44	0·32	0·23	0·18	0·14	0·11	0·32
(b)	19·43	7·78	3·71	1·90	1·00	0·54	0·29	0·16	0·08	0·05	0·02	0·01	1·82
P													
No. of sites observed with given occupancy	124	43	21	6	8	2	2	3	0	0	1	0	
No. of sites expected													
(a)	121·70	43·31	20·47	10·80	6·01	3·41	1·95	1·09	0·60	0·32	0·16	0·08	2·87
(b)	121·43	44·04	20·92	10·96	5·99	3·33	1·85	1·02	0·55	0·29	0·14	0·07	3·07

We used a chi^2 statistic to test the goodness of fit of the observed occupancy frequencies to the one expected according to the calculated α or $\beta(\theta)$ values. * $P < 0{\cdot}05$.

4. Discussion

The maintenance of transposable element copies in genomes raises the question of the nature of the forces involved in natural populations. Because of the low amount of data available on complete genomes, we do not so far have an objective and clear view of which model has the greatest chance of explaining copy number containment. It appears, indeed, that different elements may undergo different forces, but to be sure about which theoretical expectations have to be used for a given element we need a large amount of data on many different populations. Our analysis of the localization of five transposable elements on all the chromosome arms of genomes from a natural *Drosophila melanogaster* population is thus an attempt to gain information on various population dynamic parameters of TEs for complementary models to be tested.

(i) *Copy number distribution and occupancy frequencies*

If the element frequencies are low with no variance between different sites and no linkage disequilibrium,

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

it is expected that the number of TE copies (n) of a given element in a genome across individuals will follow a binomial distribution. If n is much smaller than the number of chromosomal sites into which TEs are capable of inserting, this reduces to a Poisson distribution (Charlesworth & Charlesworth, 1983; Langley, Brookfield & Kaplan, 1983). All our five elements studied are apparently in agreement with this Poisson hypothesis. We must, however, recognize that by pooling the classes with low frequency, the resulting low number of classes reduces the ability of the chi-square test to detect deviation from the distribution under test. Indeed, the hypothesis of low occupancy frequency and low variance in element frequencies between sites is not verified for all the elements, but this does not make any difference for the chi-square test. After all, an heterogeneity in occupancy frequency across sites on a chromosome arm is expected to lead to low variance in copy number (Charlesworth, 1985; Eanes *et al.* 1988).

The high occupancy frequency observed with the gypsy element could at first sight reveal a homozygosity by descent resulting from a certain degree of inbreeding in the population, or drift in previous generations. Such a hypothesis is to be rejected because no such strong accumulation of element insertions as with gypsy was observed with the other elements (inbreeding should affect the whole genome), although some specific sites showed high occupancy frequencies such as the 1A site; but this site is known to contain a particular P element with specific characteristics and it is observed in most populations studied (Ajioka & Eanes, 1989; Biémont *et al.* 1990*b*; Ronsseray, Lehmann & Anxolabéhère, 1991). The high occupancy frequencies observed for the gypsy element contradicts, moreover, the assumption of Charlesworth & Langley (1991) that low copy number families rarely show occupancies higher than two. The sites 61C and 98A with high occupancy frequencies of the gypsy element may be hot spots of insertions, perhaps as a result of their particular structure or DNA sequences, since it is well documented that gypsy insertion is site specific (Freund & Meselson, 1984). Note that the estimate of occupancy frequencies in the whole genome cannot be accurately approached in works dealing only with X chromosomes. Indeed, as seen in Fig. 1 and Table 1, the frequency of site occupancy is small on the X chromosome for all the elements, with a mean value even equal to 1 for the copia element.

(ii) *Copy number containment*

In the studies of Charlesworth & Langley (1989) and Charlesworth & Lapid (1989) only the roo element out of 12 TEs analysed showed an accumulation of insertion sites in the proximal regions of the chromosomes (a clear effect for the X but a small one for the autosomes) but not in the distal regions, these two regions being characterized, however, by reduced rates of crossing over. It is known, however, that the copia-like elements accumulate in centromeric regions, especially in the β-heterochromatin known to be a nest of old insertions of different types of immobilized transposable elements (Daniels & Strausbaugh, 1986; Lansman *et al.* 1987; Miklos *et al.* 1988; Vaury, Bucheton & Pélisson, 1989). So, when we remove the chromosomal regions 20, 40, 41, 80 and 81 from the analysis of the distribution of TE site density along the chromosomes, we find no evident accumulation of TE insertions in proximal regions of the autosomes where crossing-over is reduced (see Lindsley & Sandler, 1977); we instead observe the opposite result. It is thus not at all clear whether the accumulation in β-heterochromatin is directly associated with crossing over, or is a feature of β-heterochromatin itself.

The tendency for the TE insertion site density to decrease from the distal part of the chromosome arms to the base, following the values of the exchange coefficient, may mean that a high recombination rate may favour insertion of TEs even if the insertion sites are selected against because of chromosome rearrangements due to ectopic exchanges according to the model and observations of Goldberg *et al.* (1983), Langley *et al.* (1988) and Montgomery *et al.* (1991). It may be that the genome is more accessible to TE insertions in regions of intense recombination, or that some TEs take advantage of the recombination machinery for their own insertion. It remains, however, to explain why this is particularly evident for the 2L chromosome arm and for the copia element, but not at all for the 3L arm, as reported in other studies (Langley *et al.* 1988; Charlesworth, Lapid & Canada, 1992*a*, *b*).

Our results thus do not support the idea of a major selection effect against dominant lethal chromosome mutations produced by ectopic exchange between insertions of TEs over non-homologous regions of the chromosomes, since such a hypothesis predicts the accumulation of TEs in regions of reduced recombination. Observations of an accumulation of the *mys* element on the X and Y chromosomes of deer mice and other mammals (Baker & Wichman, 1990), and the acquisition of many repeated sequences in the newly heterochromatized neo-Y of *D. miranda* (Steinemann, 1982) are, however, in agreement with the ectopic exchange hypothesis, but the absence of *mys* copies in the autosomal heterochromatin which has no recombination, and the accumulation of copies in the highly recombinogenic pseudoautosomal segments, disagree with this hypothesis (Hale, 1992). These latter results may rather indicate constraints on TE accumulation due to the structural nature of some chromosomal regions. The Langley model (Langley *et al.* (1988) is, however, supported by the excess abundance of P elements in minority inversions of an African *D. melanogaster* population (Eanes, Wesley &

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

Charlesworth, 1992), and by a similar tendency for mdg1 (Biémont & Aouar, 1987) and for an ensemble of 9 various elements in other populations (Charlesworth, Lapid & Canada, 1992); although such results strongly need to be confirmed by detailed and complete analyses on many other elements. The fact, however, that there is no accumulation of TEs at the extremity of the tip of the X chromosome in *Drosophila* where crossing over is strongly reduced, raises serious doubts about the generality and even the validity of the ectopic exchange model for explaining maintenance of TE copy number.

The hypothesis of ectopic exchange is an alternative to other hypotheses of TE copy number containment which postulate that transposable elements are maintained in populations as a result of a balance between the transpositional increase in copy number and either regulation of the rate of transposition with increasing copy number or selection against insertional mutations. In fact, it appears that the mdg3 and P elements follow the neutral model of no selected sites, copia and gypsy fit the selected model, and for mdg1 the number of insertions on the X is so small that even the selected model is rejected. From these results it is evident that no one model can account for the regulation in copy number and a mixture of different phenomena could be involved. Even in the case of copia in which in all published studies the selective model holds for explaining the maintenance of the number of copies (see Biémont, 1992) other hypotheses can still be considered, e.g. a low number of insertions on the X chromosome because of less specific insertions on this chromosome, as can be also inferred for mdg1. Hence, even a systematic observation made on different populations cannot be used as absolute evidence for the model under test.

As in most populations studied, we find that the number of P copies on the X chromosome is slightly higher than expected, although this does not show up in the statistical test when only one population is considered. O'Hare *et al.* (1992) consider that such a result is compatible with the model of Engels *et al.* (1990) (see also Stephen & Langley, 1992) in which a homologous template is required for repair after excision of a P element. Since the X is hemizygous in males, a lower frequency of excision from the X compared to the autosomes is expected. Hence, the frequency of insertions among chromosomes may reflect the transposition or excision mechanisms more than the effect of selection or genomic recombination.

The observation in the present study that TE density of some elements globally decreases as rate of recombination decreases from the distal part of the chromosome arms to the base could be interpreted under the hitch-hiking hypothesis. This hypothesis postulates that regions of low recombination are 'hit' by selective sweep more often than regions of high recombination (Begun & Aquadro, 1992); the associated reduction in the effective population size for the chromosomal region concerned thus leads to a reduction in the proportion of sites at which TEs segregate (Charlesworth, Lapid & Canada, 1992*a*, *b*). A decreasing TE site density with a decreasing rate of recombination is thus expected. This is exactly what is clearly observed for the copia element and the TE site density on the 2L chromosome arm. Moreover, the absence of accumulation of TE sites in the tip of the X chromosome is also postulated (and is observed), since this distal region was found to have depleted nucleotide site variability due to hitch-hiking (Begun & Aquadro, 1991, 1992). The difficulty with this hypothesis is that the low number of sites in regions of reduced crossing over should be associated with a high number of copies per site (which is not observed), and more TE insertion sites on the X than on the autosomes should be observed because the X chromosomes have more opportunity to participate in meiotic exchange, since two-thirds of the X chromosomes are present in females and only one-half of the autosomes. It thus remains that the decrease of TE site density along the chromosome arms could only be more a structural characteristic of the chromosomes rather than the result of an association with rate of exchanges. The particular behaviour of the tip of some chromosome arms, could, for example, be analysed in relation with the amplification of subtelomeric specific repeat sequences (Danilevkaya & Lapta, 1991; Biessmann *et al.* 1992; Karpen & Spradling, 1992).

An alternative to the model of selection is regulation of the rate of transposition in response to the number of elements per genome. Experiments with laboratory strains seem to show a great excess of the rate of transposition over excision, and so such regulation seems unlikely to be an important force, since excision and transposition rates are expected to be approximately equal at equilibrium under this model. The problem with the transposition and excision rates is that in many experiments the populations studied were not at equilibrium for copy number but in a process of gaining new sites without any detectable excisions. Such an absence of equilibrium is sometimes advanced for eliminating P element data in estimating TE biology parameters (Charlesworth & Langley, 1991). However, this precaution should also hold for other elements for which many features in natural populations are not known. For example, Jenkins & Copeland (1985) have reported an enhanced transposition of certain murine ecotropic retroviruses under conditions similar to the ones encountered in hybrid dysgenesis, and recent observations in *Drosophila* show that retrotransposons can be mobilized in some interstrain crosses (Pasyukova *et al.* 1988; Mével-Ninio *et al.*; Georgiev *et al.* 1990) or even without any evident causes (Gerasimova *et al.* 1984*a*; Gerasimova, Mizrokhi & Georgiev, 1984; Biémont, Aouar & Arnault, 1987; Kim & Belyaeva, 1991; Pasyukova & Nuzhdin, 1992). Moreover, the idea that the loss of elements from their chromosomal locations is very

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

rare comes from the analysis of mutations that were detected visually and that resulted from a TE insertion. These mutations do not concern the whole set of insertions on the chromosomes, contrary to what is seen in studies dealing with location of all chromosomal sites by *in situ* hybridization and which show, at least for some elements, equal rates of transposition and excision (Ising & Block, 1981; Biémont & Aouar, 1987; Eggleston, Johnson-Schlitz & Engels, 1988; Biémont *et al.* 1990*b*; Harada, Yukuhiro & Mukai, 1990). We are clearly in need of reliable estimations of rates of TE movement in controlled populations, especially since these transposition and excision parameters determine the equilibrium state in copy number according to the various above models. Although more data on natural populations from different *Drosophila* species are required, experiments aimed to calculate rates of movement under various conditions, to follow the way copy number increases or is regulated in lines with a high or low initial number, and to determine selection values of TE insertions in order to test a possible advantage as recently postulated in yeast (Wilke & Adams, 1992), will still be of great use in understanding the way TEs are maintained in populations.

(iii) *Estimates of the model parameters* α *and* β

The values of α and β estimated in our population for the five elements are similar to, but less variable than, the values reported in various other studies in natural populations concerning either mainly the X chromosomes (Charlesworth & Langley, 1991) or the autosomes analysed individually (Charlesworth, Lapid & Canada, 1992*a*, *b*). Although in our study the X and the autosomes give in most cases the same estimates, the values, as seen for example with copia, can range from 6·45 in the autosomes to infinity when only the X chromosomes are concerned. We must thus be cautious not to draw inferences about the whole genome from only one chromosome. The close estimates of the θ parameter may reflect the fact that the elements were analysed in the same population (and in the same sample for some of them), and the variation between chromosome arms for the gypsy element perfectly expresses the difference in site occupancy frequencies.

The high estimates of β (around 6–30), which measures the joint effects of drift, selection and excision, clearly indicate that drift is rather ineffective in relation to the other forces controlling element frequencies, since a value close to one is expected for element frequency to be explained by drift. However, small values of β may also result from high occupancy through multiple insertions. A high frequency of multiple insertions may thus be the right explanation of the low values of β (around 1) observed in *D. algonquin* and *D. affinis* (Hey, 1989) in which many sites showed occupancies of 50% or more. The values of β depend also on the value of the rate of transposition, and differences in β estimates may reflect differences in transposition rates. The high β value found for the mdg3 element may thus reflect that this element may have a higher intrinsic transposition rate than the other elements. A wider knowledge of TE insertion polymorphism in various populations and in different species is then necessary for the resolution of the fundamental question of the amount and nature of genetic variation in natural populations (Begun & Aquadro, 1992) and on the forces that maintain it.

We thank C. Arnault, C. Bazin, P. Capy, N. Junakovic, A. Kim, C. Loevenbruck, C. Terzian, and the referees, for discussion, constructive comments on the manuscript, and various help, S. Ronsseray for *Drosophila* stocks, and N. Chaminade for technical assistance. This work was supported by the Centre National de la Recherche Scientifique (URA 243, UPR 2411), the Association pour la Recherche sur le Cancer (contract 6251 to C.B.), the Hasselblad Foundation, a EEC grant for M.P.G.G., a CNRS fellowship for E.G.P.

Appendix

The α and β parameters were estimated simultaneously according to the equation of Charlesworth & Charlesworth (1983) assuming that the probability distribution of element frequencies fits a beta distribution. The pair of α and β estimates were then those which minimized the chi-squared value used to compare the observed and estimated distributions. θ (the same parameter as β when α equals zero) was estimated following the model of Langley, Brookfield & Kaplan (1983) in three different ways:

In method A we assumed that the frequency distribution of TE sites follows the infinite alleles distribution of Kimura & Crow (1964). This states that the number of alleles in the frequency range from x to $x+\delta x$ is given by:

$$\theta_A x^{-1}(1-x)^{\theta_A-1}\,\delta x.$$

In the case of TEs we can use the $x^{-1}(1-x)^{\theta_A-1}$ term to arrive at a discrete distribution of expected numbers of sites for all possible frequencies from $1/j$ to 1, where j is the number of chromosomes sampled. Thus, for any j and θ_A, we calculate the value corresponding to i chromosomes occupied out of a possible j. Thus, the expected value for any i, which we call N(i) is:

$$(j/i)(1-i/j)^{\theta_A-1},$$

the discrete distribution gives a mean frequency of the sites given by:

$$\sum_{i=1}^{j}(\mathrm{N}(i).i)\Big/\sum_{i=1}^{j}\mathrm{N}(i).$$

This mean frequency of sites increases with decreasing θ_A, and can be compared with the mean site frequency

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

in the real data. θ_A is changed until these two values are the same.

Method B is similar to method A but the observed and expected distributions were compared using a chi-square test, and we changed the value of θ_B until the chi-squared value was minimized. The problem with this method is that the expected values for most of the frequency classes are quite small (less than one) thus making the chi-square test illegitimate. Thus, we pooled the values for various classes. This pooling can be done in a number of different ways giving each time a different estimate of the θ_B value which minimizes the chi-squared value. To take into account the different number of chromosome arms (especially the difference in number of X chromosomes and autosomes), we calculated for a given θ_B value the expected frequency spectrum for a sample of j chromosomes, where j is the number in the sample of the particular chromosome arm being considered. The calculation was repeated for each chromosome arm, using the different j values for these arms, and we combined the expected distributions, weighting the individual chromosomes according to the mean number of elements seen in the sample on each chromosome. We thus obtained a predicted frequency spectrum across all the chromosomes, which we compared to the observed distribution with a chi-square test. This method has the advantage of allowing a confidence limit on θ_B to be calculated. These confidence limits for θ_B are then 4·0 and 13·8 for copia, 6·8 and 10·9 for mdg1, 11·3 and 24·3 for mdg3, 2·6 and 37·9 for gypsy, and 4·5 and 7·5 for P.

Method C is the infinite alleles distribution method used in Kaplan & Brookfield (1983) in which the expected number of alleles, k, seen in a sample of a given size, is given, according to Ewens (1979), by:

$$\sum_{i=0}^{j-1} \theta_C/(\theta_C+i).$$

We thus calculated the value of θ_C that made equal the values of k and the reciprocal of the mean observed site frequency. Some simulations (Kaplan & Brookfield, 1983) have shown, however, that the θ_C values calculated by this method are considerably larger than true ones, especially when θ_C is large, and that the estimated values are less accurate than those produced by methods A and B.

With methods A and C, the variation in the number of chromosomes made it impossible to estimate θ_A and θ_C for the entire genome. So we calculated the average value over the chromosome arms for each element. It appears that the mean θ_A values are close to the θ_B estimates, while the mean θ_C (the parameter of the infinite alleles distribution) values are close to the β values when estimated simultaneously with α, and this is true for all elements (this calculation was not done for copia since the estimates of θ_A and θ_C for the X chromosome are infinite). However, for gypsy, for which the same number of chromosome arms was analysed, θ_A and θ_C were calculated for the entire genome directly. The values obtained are lower than the mean of the individual chromosome arm values and thus than the estimates of θ_B and β. It seems that the mean of the individual chromosome arm θ values takes better into account the sites with high insertion frequencies (which decrease the value of θ) than the θ estimates calculated directly on the entire genome.

References

Ajioka, J. W. & Eanes, W. F. (1989). The accumulation of P-elements on the tip of the X chromosome in populations of *Drosophila melanogaster*. *Genetical Research* **53**, 1–6.

Ajioka, J. W. & Hartl, D. L. (1989). Population dynamics of transposable elements. In *Mobile DNA* (ed. D. E. Berg and M. M. Howe), pp. 939–958. Washington, D.C.: American Society for Microbiology.

Ananiev, E. V., Barsky, V. E., Ilyin, Y. V. & Ryzic, M. V. (1984). The arrangement of transposable elements in the polytene chromosomes of *Drosophila melanogaster*. *Chromosoma* **90**, 366–377.

Baker, R. J. & Wichman, H. A. (1990). Retrotransposon MYS is concentrated on the sex chromosomes: implications for copy number containment. *Evolution* **44**, 2083–2088.

Bayev, A. A., Jr., Lyubomirskaya, N. V., Dzhumagalies, E. B., Ananiev, E. V., Amiantova, I. G. & Ilyin, Y. V. (1984). Structural organisation of transposable element gypsy from *Drosophila melanogaster* and a nucleotide sequence of its long terminal repeats. *Nucleic Acids Research* **12**, 3707–3723.

Begun, D. J. & Aquadro, C. F. (1991). Molecular population genetics of the distal portion of the X chromosome in Drosophila: evidence for genetic hitchhiking of the *yellow-achaete* region. *Genetics* **129**, 1147–1158.

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

Belyaeva, E. S., Ananiev, E. V. & Gvozdev, V. A. (1984). Distribution of mobile dispersed genes (mdg-1 and mdg-3) in the chromosomes of *Drosophila melanogaster*. *Chromosoma* **90**, 16–19.

Biémont, C. (1986). Polymorphism of the mdg-1 and I mobile elements in *Drosophila melanogaster*. *Chromosoma* **93**, 393–397.

Biémont, C. (1992). Population genetics of transposable elements. A drosophila point of view. *Genetica* **86**, 67–84.

Biémont, C. & Aouar, A. (1987). Copy-number dependent transpositions and excisions of the mdg-1 mobile element in inbred lines of *Drosophila melanogaster*. *Heredity* **58**, 39–47.

Biémont, C., Aouar, A. & Arnault, C. (1987). Genome reshuffling of the copia element in a *Drosophila melanogaster* inbred line. *Nature* **329**, 742–744.

Biémont, C., Arnault, C., Heizmann, A. & Ronsseray, S. (1990*a*). Massive changes in genomic locations of P elements in an inbred line of *Drosophila melanogaster*. *Naturwissenschaften* **77**, 485–488.

Biémont, C. & Gautier, C. (1988). Localisation polymorphism of mdg-1, copia, I and P mobile elements in genomes of *Drosophila melanogaster*, from data of inbred lines. *Heredity* **60**, 335–346.

Biémont, C., Gautier, C. & Heizmann, A. (1988). Independent regulation of mobile element copy number in *Drosophila melanogaster* inbred lines. *Chromosoma* **96**, 291–294.

Biémont, C., Ronsseray, S., Anxolabéhère, D., Izaabel, H. & Gautier, C. (1990*b*). Localization of P elements, copy

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

number regulation, and cytotype determination in *Drosophila melanogaster*. *Genetical Research* **56**, 3–14.

Biesmann, H., Champion, L. E., O'Hair, M., Ikenaga, K., Kasravi, B. & Mason, J. M. (1992). Frequent transpositions of *Drosophila melanogaster* Het-A transposable elements to receding chromosome ends. *EMBO Journal* **11**, 4459–4469.

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

Brookfield, J. F. Y. (1986). The population biology of transposable elements. *Philosophical Transactions of the Royal Society of London* B **312**, 217–226.

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 population. In *Plant Transposable Elements* (ed. O. Nelson), pp. 189–212. New York, London: Plenum Press.

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

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

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

Charlesworth, B. & Langley, C. H. (1991). Population genetics of transposable elements in Drosophila. In *Evolution at the Molecular Level* (ed. R. K. Selander, A. G. Clark and T. S. Whittam). Sunderland: Sinauer Associates Inc.

Charlesworth, B. & Lapid, A. (1989). A study of 10 transposable elements on X chromosomes from a population of *Drosophila melanogaster*. *Genetical Research* **54**, 113–125.

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

Charlesworth, B., Lapid, A. & Canada, D. (1992*b*). 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.

Daniels, S. B. & Strausbaugh, L. D. (1986). The distribution of P element sequences in Drosophila: the *willistoni* and *saltans* species group. *Journal of Molecular Evolution* **23**, 138–148.

Danilevkaya, O. N. & Lapta, G. E. (1991). Long telomeres in the polytene chromosomes of *Drosophila melanogaster* are associated with amplification of subtelomeric repeat sequences. *Génétique, Sélection, Evolution* **23**, 15–24.

Dunsmuir, P., Brorein, W. J., Simon, M. A. & Rubin, G. M. (1980). Insertion of the *Drosophila* transposable element copia generates a 5 base pair duplication. *Cell* **21**, 575–579.

Eanes, W. F., Wesley, C., Hey, J., Houle, D. & Ajioka, J. (1988). The fitness consequences of P element insertion in *Drosophila melanogaster*. *Genetical Research* **52**, 17–26.

Eanes, W. F., Wesley, C. & Charlesworth, B. (1992). Accumulation of P elements in minority inversions in natural populations of *Drosophila melanogaster*. *Genetical Research* **59**, 1–9.

Echalier, G. (1989). Drosophila retrotransposons; interactions with genome. *Advances in Virus Research* **36**, 33–105.

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

Engels. W. R., Johnson-Schlitz, D. M., Eggleston, W. B. & Sved, J. (1990). High frequency P element loss in Drosophila is homolog dependent. *Cell* **62**, 515–525.

Ewens, W. J. (1979). *Mathematical Population Genetics*. Berlin: Springer-Verlag.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends in Genetics* **5**, 103–107.

Freund, R. & Meselson, M. (1984). Long terminal repeat nucleotide sequence and specific insertion of the gypsy transposon. *Proceedings of the National Academy of Sciences, USA* **81**, 4462–4464.

Georgiev, P. G., Kiselev, S. L., Simonova, O. B. & Gerasimova. T. I. (1990). A novel transposition system in *Drosophila melanogaster* depending on the stalker mobile genetic element. *EMBO Journal* **9**, 2037–2044.

Gerasimova, T. I., Matyunina, L. V., Ilyin, Y. V. & Georgiev, G. P. (1984). Simultaneous transposition of different mobile elements. Relation to multiple mutagenesis in *Drosophila melanogaster*. *Molecular and General Genetics* **194**, 517–522.

Gerasimova, T. I., Mizrokhi, L. J. & Georgiev, G. P. (1984). Transposition bursts in genetically unstable *Drosophila melanogaster*. *Nature* **309**, 714–716.

Goldberg, M. L., Sheen, J.-Y., Gehring, W. & 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.

Hale, D. W. (1992). Sex chromosomes, heterochromatin, and retrotransposon accumulation in deer mice. *Evolution* **46**, 1955–1958.

Harada, K., Yukuhiro, K. & Mukai, T. (1990). Transposition rates of movable genetic elements in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA* **87**, 3248–3252.

Hey, J. (1989). The transposable portion of the genome of *Drosophila algonquin* is very different from that in *Drosophila melanogaster*. *Molecular Biology and Evolution* **6**, 66–79.

Ilyin, Y. V., Chmeliauskaite, V. G. & Georgiev, G. P. (1980). Double-stranded sequences in RNA of *Drosophila melanogaster*: relation to mobile dispersed genes. *Nucleic Acids Research* **8**, 3439–3457.

Ising, B. & Block, K. (1981). Derivation-dependent distribution of insertion sites for a Drosophila transposon. *Cold Spring Harbor Symposia on Quantitative Biology* **45**, 527–544.

Jenkins, N. A. & Copeland, N. G. (1985). High frequency germ-line acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. *Cell* **43**, 811–819.

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

Kaplan, N. L., Darden, T. & Langley, C. H. (1985). Evolution and extinction of transposable elements in Mendelian populations. *Genetics* **109**, 459–480.

Karpen, G. H. & Spradling, A. C. (1992). Analysis of subtelomeric heterochromatin in the Drosophila minichromosome *Dp1187* by single P element insertional mutagenesis. *Genetics* **132**, 737–753.

Kidwell, M. G. (1983). Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA* **80**, 1655–1659.

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press

Kim, A. I. & Belyaeva, E. S. (1991). Transposition of mobile elements gypsy (mdg4) and hobo in germ-lime and somatic cells of a genetically unstable mutator strain of *Drosophila melanogaster*. *Molecular and General Genetics* **229**, 437–444.

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Langley, C. H., Brookfield, J. F. Y. & Kaplan, N. (1983). Transposable elements in Mendelian populations. I. A theory. *Genetics* **104**, 457–472.

Langley, C. H., Montgomery, E. A., Hudson, R., Kaplan, N. & Charlesworth, B. (1988). On the role of unequal exchange in the containment of transposable element copy number. *Genetical Research* **52**, 223–236.

Lansman, R. A., Shade, R. O., Grigliatti, T. A. & Brock, H. W. (1987). Evolution of P transposable elements: sequences of *Drosophila nebulosa* P elements. *Proceedings of the National Academy of Sciences, USA* **84**, 6491–6495.

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

Levis, R., Dunsmuir, P. & Rubin, G. M. (1980). Terminal repeats of the Drosophila transposable element copia: nucleotide sequence and genomic organisation. *Cell* **21**, 579–581.

Lindsley, D. L. & Sandler, L. (1977). The genetic analysis of meiosis in female *Drosophila melanogaster*. *Philosophical Transactions of the Royal Society of London* B **277**, 295–312.

Lyubomirskaya, N. V., Arkhipova, I. R., Ilyin, Y. V. & Kim, A. I. (1990). Molecular analysis of the gypsy (mdg4) retrotransposon in two strains differing by genetic instability. *Molecular and General Genetics* **223**, 305–309.

Mackay, T. F. C. (1989). Transposable elements and fitness in *Drosophila melanogaster*. *Genome* **31**, 284–295.

Mevel-Ninio, M., Mariol, M. C. & Gans, M. (1989). Mobilization of the gypsy and copia retrotransposons in *Drosophila melanogaster* induces reversion of the ovo^D dominant female-sterile mutations: molecular analysis of revertant alleles. *EMBO Journal* **8**, 1549–1558.

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 Sciences, USA* **85**, 2051–2055.

Modolell, J., Bender, W. & Meselson, M. (1983). *Drosophila melanogaster* mutations suppressible by the suppressor of Hairy-wing are insertions of a 7·3-kilobase mobile element. *Proceedings of the National Academy of Sciences, USA* **80**, 1678–1682.

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

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

O'Hare, K., Driver, A., McGrath, S. & Johnson-Schlitz, D. M. (1992). Distribution and structure of cloned P elements from the *Drosophila melanogaster* P strain π_2. *Genetical Research* **60**, 33–41.

O'Hare, K. & Rubin, G. M. (1983). Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* **34**, 25–35.

Pasyukova, E. G., Belyaeva, E. S., Ilyinskaya, L. E. & Gvozdev, V. A. (1988). Outcross-dependent transpositions of copia-like mobile genetic elements in chromosomes of an inbred *Drosophila melanogaster* stock. *Molecular and General Genetics* **212**, 281–286.

Pasyukova, E. G. & Nuzhdin, S. V. (1992). Mobilization of retrotransposon copia in genome. *Genetika* **28**, 5–18.

Peifer, M. & Bender, W. (1988). Sequences of the gypsy transposon of Drosophila necessary for its effects on adjacent genes. *Proceedings of the National Academy of Sciences, USA* **85**, 9650–9654.

Ronsseray, S. & Anxolabéhère, D. (1986). Chromosomal distribution of P and I transposable elements in a natural population of *Drosophila melanogaster*. *Chromosoma* **94**, 433–440.

Ronsseray, S., Lehmann, M. & Anxolabéhère, D. (1989). Distribution of P and I mobile elements copy number in *Drosophila melanogaster* populations. *Chromosoma* **98**, 207–214.

Ronsseray, S., Lehmann, M. & Anxolabéhère, D. (1991). The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome. *Genetics* **129**, 501–512.

Scheinker, V. S., Lozovskaya, E. R., Bishop, J. G., Corces, V. G. & Evgen'ev, M. B. (1990). A long terminal repeat-containing retrotransposon is mobilized during hybrid dysgenesis in *Drosophila virilis*. *Proceedings of the National Academy of Sciences, USA* **87**, 9615–9619.

Steinemann, M. (1982). Multiple sex chromosomes in *Drosophila miranda*: a system to study the degeneration of a chromosome. *Chromosoma* **86**, 59–76.

Stephan, W. & Langley, C. H. (1992). Evolutionary consequences of DNA mismatch inhibited repair opportunity. *Genetics* **132**, 567–574.

Strobel, E., Dunsmuir, P. & Rubin, G. M. (1979). Polymorphism in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. *Cell* **17**, 429–439.

Tchurikov, N. A., Ilyin, Y. V., Skryabin, K. G., Ananiev, E. V., Bayev, A. A., Krayev, A. S., Zelentsova, E. S., Kulguskin, V. V., Lyubomirskaya, N. V. & Georgiev, G. P. (1981). General properties of mobile dispersed genetic elements in Drosophila. *Cold Spring Harbor Symposia on Quantitative Biology* **45**, 655–665.

Vaury, C., Bucheton, A. & Pélisson, A. (1989). The β heterochromatic sequences flanking the I elements are themselves defective transposable elements. *Chromosoma* **98**, 215–224.

Wilke, C. M. & Adams, J. (1992). Fitness effects of Ty transposition in *Saccharomyces cerevisiae*. *Genetics* **131**, 31–42.

Yamaguchi, O., Yamazaki, T., Saigo, K., Mukai, T. & Robertson, A. (1987). Distribution of three transposable elements, P, 297, and copia in natural populations of *Drosophila melanogaster*. *Japanese Journal of Genetics* **62**, 205–216.

https://doi.org/10.1017/S0016672300032353 Published online by Cambridge University Press