Localisation polymorphism of mdg-1, copia, I and P mobile elements in genomes of *Drosophila melanogaster*, from data of inbred lines

C. Biémont and
C. Gautier

Laboratoire de Biologie des Populations, Université
Lyon 1, 69622 Villeurbanne cedex, France.

Using data from *in situ* hybridisation of the giant chromosomes from salivary glands, mdg-1, I, copia, and P mobile element polymorphism was studied in 17 highly-inbred lines of *Drosophila melanogaster* and in the expected hybrids obtained from the theoretically-crossed inbred lines. The mean copy numbers of each element on the inbred lines were close and equal to 16.8 for mdg-1, 17.5 for I, 17.8 for copia and 18.6 for P (these values included the insertions in the centromeric regions). The P element differed from the other three for variance in copy number and distribution of the number of insertion site occurrences. A low frequency of X-linked copia element insertions (as compared with frequency of insertions in the other chromosomes) was reported, suggesting that natural selection acts against insertional deleterious mutations of this element. Such a low frequency of X-linked insertions was not observed for the other three elements. The mdg-1, I and copia element copy numbers were positively correlated, thus leading to lines with high or low copy number. No correlation was observed between the P element copy number and the numbers of mdg-1, I and copia elements. Moreover, a genomic control of copy number was detected for mdg-1, I and copia, but not for P. The number of labelling sites, the number of copies, the number of heterozygous insertion sites, and various indices of genetic variability were estimated from the expected hybrids obtained from the theoretically-crossed inbred lines; the calculated numbers of labelling sites for mdg-1 and I did not differ from those obtained directly from the initial wild-derived population (we had no estimate of the number of labelling sites available in this population for copia and P). An average proportion of heterozygous sites of 0.86, 0.92, 0.90 and 0.94, and an average number of copies of 33, 35, 38 and 37 were found in the theoretical hybrids for the mdg-1, I, copia and P elements, respectively.

INTRODUCTION

The idea that mobile genetic elements are normal components of the genome of many organisms is now widely accepted. These elements may promote mutations by insertions, excisions and regulation of classical genes (Green, 1969; Bréglino *et al.*, 1980; Kidwell, 1985) and thus participate in the genetic load (Mukai and Yukuhiro, 1983; Biémont *et al.*, 1985; Yukuhiro *et al.*, 1985; Mackay, 1986; Fitzpatrick and Sved, 1986). By increasing the potential genetic variability of a population, they may favour selection on quantitative traits (Mackay, 1985). Since environmental stress can increase transcription of some of them in yeast (Paquin and Williamson, 1984; Rolfe *et al.*, 1986) and *Drosophila* (Strand and Macdonald, 1985), and may promote transposition in *Drosophila* (Junakovic *et al.*, 1986), these elements may play an important role in population adaptation. However, although data on polymorphism of such

elements in a genome have begun to accumulate, these studies are generally restricted to the X chromosomes (Montgomery and Langley, 1983; Ronsseray and Anxolabéhère, 1986; Leigh Brown and Moss, 1987). Little work has addressed the whole genome (Belyaeva *et al.*, 1984; Biémont *et al.*, 1985; Pasyukova *et al.*, 1986; Biémont, 1986; Biémont and Aouar, 1987), and the insertion sites are not always localised precisely on the chromosomes (Montgomery *et al.*, 1987). We thus have no idea of the localisation polymorphism in most of the elements so far detected, nor of the relationships between them, although such information is necessary for assumptions to be made for the construction of theoretical models (Charlesworth, 1983; Charlesworth and Charlesworth, 1983; Langley *et al.*, 1983; Charlesworth and Langley, 1986; Brookfield, 1986). Here, we analyse the localisation polymorphism of the mdg-1, I, copia and P elements, on all the chromosome arms of 17 *Drosophila melanogaster* inbred lines. From the

data, we elaborate the relationships between the elements themselves, and we estimate the expected degree of heterozygosity, and other indices of genetic diversity, of the individuals in the initial mass-mated population from which the lines derived.

MATERIALS AND METHODS

The Drosophila population

A laboratory population of *Drosophila melanogaster* was established from about 50 flies captured in Azerbaïdjan (USSR) in the end of 1983. The population so formed was maintained in the laboratory by mass culture at 25°C before 17 inbred lines were established in October 1984. The 17 initial pairs of flies were randomly taken from the population. The lines were then maintained by one sib pair every generation. To avoid loosing lines, 6 sib pairs were individually isolated for each line, every generation; for a given line, the sib pair that was considered for obtaining the following generation was taken from those that had progeny. At generation 52, each inbred line was analysed for its number and localisation of the mdg-1, I, copia and P mobile elements by *in situ* hybridisation of biotinylated DNA probes on giant chromosomes (Rigby *et al.*, 1977; Langer-Safer *et al.*, 1982).

DNA probes

We used the A fragment of mdg-1 (5.6 kb) incorporated at the *Hind*III site of the pBR 322 plasmid (Ilyin *et al.*, 1980; Tchurikov *et al.*, 1981), the I fragment (5.4 kb) included at the *Sal* site of the pAT 153 plasmid (Bucheton *et al.*, 1984), the *p* π 25.1 plasmid containing the whole of the P element (2.9 kb; O'Hare and Rubin, 1983), and the probe cDm 5002 containing the copia element (5 kb; Levis *et al.*, 1980; Dunsmuir *et al.*, 1980).

In situ hybridisation

Salivary glands of third instar larvae, reared at 18°C, were squashed in 45 per cent acetic acid, dehydrated in 96 per cent ethanol and air dried. The slides were incubated in 2×SSC solution for 30 min at 70°C, dehydrated with ethanol and air dried. They were then treated with 0.07 NaOH for 2 min, dehydrated and then air dried. A solution of biotinylated-denatured DNA probe, which contained 4×SSC, dextran and 50 per cent formamide,

was added to the preparations. Annealing was carried out overnight in a humid chamber at 37°C. The slides were then washed in successive changes of 2×SSC and PBS. The hybridisation sites were made visible by reacting the chromosomes to a biotinylated-streptavidin complex labelled with horseradish peroxidase and finally with diaminobenzidine and H₂O₂ as stains for peroxidase. The slides were then Giemsa stained. The hybridisation sites were analysed using coloured pictures obtained from a phase-contrast microscope (see Biémont and Terzian, 1986). Four to six nuclei were analysed per larva. It was observed that no polymorphism of insertion existed between different nuclei of the same salivary gland. To be sure of the insertion site location, the pictures of chromosomes from all the larvae were compared two by two. For a detailed analysis of the patterns of mdg-1 and I insertion sites in individuals from the mass-mated population, see Biémont (1986), and for the mdg-1 insertion pattern in the present inbred lines at generations before the 52nd, see Biémont and Aouar (1987).

RESULTS

Copy number of elements

Table 1 shows the number of mdg-1, I, copia and P element insertion sites found on the chromosome arms of the 17 inbred lines. The average genome had a total of 16.8 insertions for mdg-1, 17.5 for I, 17.8 for copia and 18.6 for P. There is no difference between the mean copy numbers of mdg-1, I, copia and P (the mean numbers were compared two by two by a Wilcoxon's signed rank test; Lehmann, 1975). The variances were similar for mdg-1, I and copia, but a greater variability in copy number between the lines was observed for the P element (Siegel-Tukey's test; Lehmann, 1975). This is clearly reflected in the distributions of copy number presented in fig. 1. The high variability in P copy number was due to some lines with an extreme number of elements and especially to one line (No. 7) with an unusually high number of insertions (43 insertion sites). We have tested whether this high copy number of P elements resulted from a recent transposition event within the line or expressed only the segregation over generations of sib crosses of an initially high P-copy number in the individuals from which the line was derived. Thus, we checked for polymorphism among larvae of this line 7 in further generations. No polymorphism of insertion was detected.

Table 1 Number of insertion sites of mdg-1, I, copia and P elements per chromosome arm in the 17 inbred lines

Line	CHROMOSOME ARMS											
	Mdg-1			I								
	X	2L	2R	3L	3R	Total	X	2L	2R	3L	3R	Total
2	4	0	5	3	4	16	4	0	5	5	6	20
3	1	2	4	4	5	16	4	1	4	5	4	18
4	3	1	2	3	5	14	4	2	4	2	5	17
5	4	3	4	2	2	15	1	1	8	5	6	21
6	2	3	6	2	5	18	6	2	2	4	4	18
7	2	3	8	3	1	17	4	2	3	3	3	15
8	1	3	2	4	4	14	3	1	4	4	2	14
9	3	3	6	4	5	21	4	2	5	3	7	21
10	1	4	4	3	4	16	3	2	4	5	4	18
11	3	2	2	5	6	18	2	3	5	3	7	20
12	5	0	2	3	5	15	5	1	4	5	5	20
13	2	2	6	2	3	15	3	1	1	3	5	13
14	5	3	4	5	2	19	1	2	3	4	7	17
15	3	2	3	2	5	15	6	2	2	1	5	16
16	3	3	5	1	4	16	1	2	3	5	5	16
17	3	2	3	5	5	18	3	1	2	5	6	17
18	3	1	5	3	10	22	2	2	6	4	3	17
Total	48	37	71	54	75	285	56	27	65	66	84	298
Mean	2.82	2.18	4.18	3.18	4.41	16.76	3.29	1.59	3.82	3.88	4.94	17.53
Variance	1.44	1.21	2.86	1.32	3.65	4.97	2.31	0.48	2.72	1.39	2.07	5.20

Line	Copia						P					
	X	2L	2R	3L	3R	Total	X	2L	2R	3L	3R	Total
2	2	3	3	2	6	16	4	2	4	2	1	13
3	2	5	3	3	3	16	3	4	3	2	5	17
4	1	3	2	4	6	16	9	3	2	2	5	21
5	1	7	4	1	3	16	6	1	4	4	8	23
6	4	4	3	2	4	17	2	4	3	3	7	19
7	2	5	2	4	4	17	8	6	12	7	10	43
8	0	2	1	2	7	12	6	4	1	2	5	18
9	1	7	4	2	7	21	4	4	7	4	6	25
10	0	6	1	6	9	22	6	6	4	2	3	21
11	3	6	3	5	8	25	0	3	0	5	3	11
12	2	7	1	1	6	17	3	2	5	3	4	17
13	1	8	4	3	2	18	3	3	1	2	3	12
14	4	5	3	5	2	19	6	3	1	1	7	18
15	2	4	3	6	3	18	3	2	3	3	4	15
16	1	2	4	1	8	16	2	3	5	2	2	14
17	1	6	5	6	4	22	0	3	1	3	1	8
18	0	2	2	3	8	15	6	2	4	2	7	21
Total	27	82	48	56	90	303	71	55	60	49	81	316
Mean	2.59	4.82	2.82	3.29	5.29	17.82	4.18	3.24	3.53	2.88	4.65	18.59
Variance	1.42	3.53	1.32	3.06	4.97	9.30	6.15	1.72	7.62	1.99	5.43	56.40

Moreover, during the establishment of inbred lines, many abnormalities were observed in line 7; so that at generation 27, a subpopulation had been set up for further detailed genetic analysis. At generation 60, this subpopulation showed a similar pattern of P insertions as in line 7, except for 2 sites. These results reflect the segregation in line 7 and its subline over the generations of inbreeding of a high number of P element insertions. This particular behaviour of the P elements (as opposed to mdg-1, I and copia) is also evident on the total

copy number on the chromosome arms (table 1). Indeed, while the smallest number of insertions was observed for the 2L arm for the mdg-1 and I elements, or the X and the 2R for copia, the P element showed high insertions in the X chromosome and a few in the 3L as already reported (Ronsseray and Anxolabéhère, 1986; Leigh Brown and Moss, 1987). As in other studies, the 3R chromosome arm had the highest copy number of insertions for each element (Pierce and Lucchesi, 1981; Belyaeva *et al.*, 1984; Biémont *et al.*, 1985;

Figure 1 Distributions of copy number of the mdg-1, I, copia, and P mobile elements on the genomes of the 17 inbred lines. Data for all the chromosome arms are compounded.

and the 3R 330 (27.5). These values can be compared with the cytogenetic measure of band counts in the polytene chromosomes (Lefevre, 1976). From the X chromosomes, 1012 bands (20.18 per cent) were estimated, 804 (16.04 per cent) on the 2L, 1136 (22.7 per cent) on the 2R, 884 (17.63 per cent) on the 3L and 1178 (23.49 per cent) on the 3R. These cytogenetic measures and observed insertion numbers are not apparently correlated, *i.e.* the number of element insertions was not proportional to the number of bands per chromosome arm; the discrepancy is due to the low number of insertions in the X chromosome for the copia element and an associated high copy number on the 2L. This low frequency of X-linked copia elements recalls a recent observation (Montgomery *et al.*, 1987) of a low frequency of X-linked 412 elements (a Copia-like element). Such an observation can be taken as evidence of natural selection acting against insertional deleterious mutations, since insertions on the X are submitted to more intense selection in the hemizygous males than insertions on the autosomes. However, although this hypothesis is sustained by the low frequency of X-linked copia elements reported by Strobel *et al.* (1979), it is not corroborated by the results of these same authors on the 412 element whose relative proportions of insertions among the chromosomes depend greatly on the *Drosophila* stock analysed. In addition, this hypothesis does not agree with the behaviour of our three other elements, especially that of the P element, which showed a high X-linked copy number. Note that instead of searching for low X-linked copy number to test a selection hypothesis, it would be wise to try to explain the quite low values of the 2L-linked I and the 3L-linked P copy numbers! Sampling error may be responsible for such variations which, as shown by previous studies on the initial population, can change also with generation time in the laboratory (Biémont *et al.*, 1985; Biémont, 1986; Biémont and Aouar, 1987).

Genomic control of copy number

To test for the possible regulation in mobile element copy number within the genome, we have considered for each element the variance of the number of copies per inbred line, and per chromosome arm. If there is no genomic control, the variable, number of copies in each chromosome arm, must be statistically independent. This implies: (a) that the observed variance of the total copy number equals the sum of the variances of copy numbers in each arm, (b) that permutations

Biémont, 1986; Ronsseray and Anxolabéhère, 1986; Biémont and Aouar, 1987; Leigh Brown and Moss, 1987).

If we consider the total number of insertions, all elements compounded, the X and the 2L had identical copy numbers, 202 (16.8 per cent) and 201 (16.7 per cent), respectively; the 2R showed up 244 insertions (20.3 per cent), the 3L 225 (18.7

among the lines of the copy numbers on chromosome arms do not modify in the same direction the variance of total copy number. Table 2 shows that for mdg-1, I and copia the observed variance is inferior to the sum of the variances for each arm. This indicates that a compensatory mechanism might exist between chromosome arms, leading to a total number of copies more constant than expected. A test using 1000 sets of permutations shows that this structure is highly significant for mdg-1 and has a critical level of about 10 per cent for I and copia. A complete opposite feature is observed for the P element for which the observed variance is superior to the sum of variances and always higher than those simulated. This latter result is mainly due to two extreme lines with 43 and 8 P copy numbers. However, if we eliminate these two lines, then the calculated and observed variances become only equal. These results suggest genomic control of copy number for the mdg-1, I and copia elements but not for P.

A simple way to reveal a genomic control is to compare the variance of the total number of copies of an element with the sum of the variances of the number of copies in each arm. The ratio of these values is close to 1 if no relation exists between numbers of copies among the chromosome arms; it is greater than 1 if these copy numbers are positively correlated; it is less than 1 if a genomic control exists. Although no statistical test can be associated with this approach, the values of the ratio of the variances (table 2) confirm the conclusion of the simulation.

Correlation between the element copy numbers

The above analysis done on each element gives no information on the relationships between the elements themselves. Table 3 presents the correlation

matrix for the four elements in the whole genome. The main tendency is that mdg-1, I and copia are positively correlated whereas no association is visible with the P element. This tendency is confirmed by analysing the standardised copy number of elements in the lines by Principal Component Analysis (fig. 2). The 17 lines are represented in a 4-dimensional space with the copy number as coordinates; these lines and the four initial axes are then projected in the plan that gives the less distorted 2-dimensional representation. The position of the mdg-1, I, copia and P axes gives a graphical representation of the above correlation matrix. Correlation between the elements mdg-1, I and copia determines the first component of the analysis; this component is mainly linked to the number of mdg-1 + I + copia insertions with no effect of the P elements; these P elements contribute only to the scatter of the lines on the second component. We thus eliminated the P element copy number from our analysis, and considered only the association between mdg-1, I and copia. The Friedman's test on the data from these three elements shows that some lines had a uniformly high copy number of mdg-1, I and copia insertions, while other lines had a low copy number of these three elements ($\chi^2 = 25.5$; $P = 0.06$), the number of P insertions being distributed independently of these three elements. The not-so-low value of the probability associated with the correlation between the mdg-1, I and copia copy numbers implies more experiments for the validity of this correlation to be ascertained.

Distribution patterns of the mobile elements in the genome

Figs. 3 to 6 illustrate the distribution patterns of mdg-1, I, copia and P along the chromosomes for all the lines. In total, 285 insertions were detected

Table 2 Observed variances in copy number per line for each element, variances calculated after having permuted the inbred lines, and sums of variances of copy number in each chromosome arm. Fsup: frequency of simulated (1000 sets of permutations) variance values superior or equal to that observed; Finf: frequency of simulated variance values inferior or equal to that observed. P*: the two lines with the lowest (8) and highest (43) P copy number are excluded from the calculations. TV/SV is the ratio of the variance of the total number of copies in a line (TV) (see table 1) over the sum of the variances of the copy number in each arm (SV)

	Mdg-1	I	Mobile element Copia	P	P*
Observed variance	5.00	5.19	9.32	56.36	15.82
Mean of 1000 estimated variances	10.47	9.04	14.39	23.44	14.54
Sum of variances per arm	10.48	8.97	14.30	22.91	14.53
Fsup	0.97	0.91	0.88	0.00	0.36
Finf	0.03	0.08	0.12	1.00	0.64
TV/SV	0.47	0.58	0.65	2.46	1.09

Figure 2 Principal Components Analysis of the standardized copy number of mdg-1, I, copia and P elements in the 17 inbred lines. The numbers correspond to the amount of mdg-1 + I + copia insertions for each line. Position of the mdg-1, I, copia, and P initial axes, which are projected on the plan, gives a graphical representation of the correlation matrix between the element copy numbers.

on 102 cytologically-distinguishable sites for mdg-1, 298 insertions on 132 sites for I, 303 insertions on 140 sites for copia, and 316 insertions on 180 sites for P. The P elements thus had the highest number of insertion sites. This is particularly evident on the X chromosomes, where the insertion sites were scattered throughout the chromosome. Yet, as in previous studies, the chromosomal region 14-17 had the lowest number of insertions (Ronsseray and Anxolabéhère, 1986). Hence, in agreement with those studies on the X chromosomes (Montgomery and Langley, 1983; Ronsseray and Anxolabéhère, 1986) and with

analyses on the whole genome (Belyaeva *et al.*, 1984; Biémont *et al.*, 1985; Biémont, 1986; Biémont and Aouar, 1987), there are some regions of the chromosomes in which more insertions were found than in others. Moreover, the high scatter of P insertions along the chromosomes reveals that the P elements had a high polymorphism in the initial population; only the band 93F appeared as a hot spot of insertion. The patterns are quite different for mdg-1, I and to a lesser extent copia, in which many bands with high frequency of insertions were observed.

Visual inspection of the distributions of the four types of elements clearly suggests that each element had its preferential regions of insertion in the population under study (see also Belyaeva *et al.*, 1984; Ronsseray and Anxolabéhère, 1986). A region that shows a high frequency of insertion of an element may also show the complete absence of another element (see the region 95-100 with high insertions of I elements and a few insertions of P elements). However, no exclusion can be concluded from our data.

Table 3 Correlation matrix for the mdg-1, I, copia, and P copy number among the inbred lines

	Mdg-1	Elements I	Copia
I	0.26		
Copia	0.36	0.34	
P	0.13	-0.04	-0.22

Figure 3 Location on the chromosomes of *mdg-1* mobile element insertions found in the 17 inbred lines.

Figure 5 Location on the chromosomes of *copia* mobile element insertions found in the 17 inbred lines.

Figure 4 Location on the chromosomes of *I* mobile element insertions found in the 17 inbred lines.

Figure 6 Location on the chromosomes of *P* mobile element insertions found in the 17 inbred lines.

One interesting point is the comparison of the patterns of localisation established from our inbred lines with the patterns obtained directly on individuals from the initial population from which the lines were derived. This comparison can be done for the I and mdg-1 elements whose polymorphism has been analysed in this initial population (Biémont, 1986). The patterns have many similarities, thus suggesting that an analysis with inbred lines gives the same kind of information as an analysis on individuals from the initial mass-mated population. However, it is striking to observe a quasi absence of I insertions on the major part of the 2L chromosome arm from the data with inbred lines, in contrast to what was observed on the 2L chromosome arm from the initial population (Biémont, 1986). This reflects either the segregation of hybridisation points, which were highly heterozygous on this chromosome arm in the mass-mated population, or only the result of sampling error.

Table 4 resumes the distribution of the number of occurrences of each element (the frequency spectra) in the 17 inbred lines. Yet, the elements mdg-1, I and copia appeared to be similarly distributed ($\chi^2 = 11.9$; df = 12; $P = 0.46$) but differed from the distribution of the P elements which present high number of unique insertion sites ($\chi^2 = 33.5$; df = 18; $P = 0.01$ for all the data of the contingency table).

Influence of inversions on copy number

In a recent work (Biémont and Aouar, 1987) we have reported that one line (No. 18) had the inversion 86DE-94F, and that genomes heterozygous in this line for this inversion had more mdg-1 insertions on the 3R chromosome arm than genomes without the inversion. In the present study, where the four elements mdg-1, I, copia and P were analysed in the lines at the 52nd generation of sib matings, we still observed this phenomenon for mdg-1, but also for copia (4 additional insertions were seen in the genomes heterozygous for the

inversion). However, no change in pattern of localisation was observed for I and P elements.

Measurements of genetic variability indexes in the reconstituted mass-mated population

Since homozygous inbred lines may be considered as a random sample of gametes from individuals of the initial population, we analysed the expected insertion patterns of the $n(n-1)/2$ hybrids obtained from theoretically crossing the n inbred lines. In first approximation, these hybrids may represent individuals from the initial mass-mated population. We have previously shown that consideration of the hybrids without crosses between the lines is the equivalent of taking two gametes in a binomial fashion from the ensemble of gametes (Biémont and Gautier, 1987).

We then calculated the number of labelling sites (NS), the number of copies in the genomes (NC) and the number of heterozygous insertions (NH). We also calculated other indices of genetic variability as defined in Biémont and Gautier (1987):

$PH = (1/r)(2\sum f_i(1-f_i))$: the probability that a random site is heterozygous in a random zygote; where r is the total number of sites in the ensemble of lines, and f_i is the frequency of site i .

$H = \sum (f_i)^2$: the mean number of sites at which the two chromosome sets of a zygote are occupied (Langley *et al.*, 1983).

$A = (\sum (f_i)^2) / \sum f_i = H / \sum f_i = 1 - (NH/NC)$: the probability that a random copy of a genome finds a homologous copy of the same chromosomal site of another genome of the population (Ohta, 1985, 1986).

All values of these variables and indices calculated on the whole genome are resumed in table 5; the data on each chromosome arm are not given but we again found the same tendencies as revealed with the inbred lines: a high number of labelling sites in the 3R chromosome arm for all the elements; many insertions in the X chromosome for the P element; low number of insertions for the

Table 4 Distributions on the whole genome of the number of occurrences (the frequency spectra) of the mdg-1, I, copia, and P elements. For example, there are 17 sites in which 3 lines have a mdg-1 insertion

Element	Number of occurrences												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Mdg-1	45	14	17	8	5	6	1	2	0	2	0	2	0
I	69	25	19	6	4	3	0	3	0	2	0	0	1
Copia	69	34	15	10	2	3	2	1	0	3	0	0	0
P	114	40	12	4	3	5	1	0	1	1	0	0	0

2L arm for mdg-1 and I. The same remarks can be drawn for the number of copies and the number of heterozygous sites; however, the proportions of heterozygous sites (NH/NS per individual) were homogeneous between the 2L, 2R, 3L and 3R chromosome arms, whatever the element considered; this proportion was lower in the X chromosomes for mdg-1, copia and P elements but not for I. When the whole genome is concerned the proportion of heterozygous sites is similar for each mobile element and is around 0.90 (ranging from 0.86 to 0.94, see table 5). Overall, for a given element, 33–38 copies of the element in the diploid genome corresponded to an average number of labelling sites around 30–35; among these sites, 25–33 were heterozygous. The number of labelling sites estimated on these hybrids for mdg-1 (29.35) and I (31.69) are close to the values reported on individuals from the wild-derived initial population (29.46 for mdg-1, 30.71 for I; Biémont, 1986; Biémont and Aouar, 1987). We have no estimate of the number of labelling sites available in this population for copia and P.

All the information obtained from the inbred line and the hybrid values are summarised by the P_H , A and H indices (see table 5). For example, P_H , i.e. the probability that a random site is heterozygous in a random zygote, is low for the P elements as a result of a high number of copies of these elements in sites highly scattered over the chromosomes; P_H is low because of the high homozygosity for element absence, which results from insertions seen only in one or two lines. Langley's index H is not really useful if considered alone, but taken with NH, it gives complete information on the genetic variability. Note, however, that an estimate of the number of homozygous insertion sites can be obtained directly from NS, NC and NH values. Ohta's A index gives information similar to that of the Het values. Indeed, the higher the degree of heterozygosity, the smaller

the probability that a random copy finds a homologous copy in another genome. Hence, knowing NS, NC and NH is in itself sufficient to have a good idea of the genetic variability of a genome. Yet, from all these data, it appears that the highest variability is observed with the P element, and the lowest with mdg-1.

DISCUSSION

Our results reveal that an important polymorphism of insertions exists for the mdg-1, I, copia and P elements. This confirms previous observations of other *Drosophila* populations on the same and other mobile elements (Pierce and Lucchesi, 1981; Montgomery and Langley, 1983; Belyaeva *et al.*, 1984; Ronsseray and Anxolabéhère, 1986). Moreover, this suggests that the I and P elements, which are involved in the I-R and P-M systems of hybrid dysgenesis, respectively (Brégliao *et al.*, 1980; Kidwell, 1985), may behave in a genome in ways similar to that of other transposable elements. This is particularly evident for the I element, which behaved like mdg-1 and copia. The P element showed, however, many specific features. Indeed, for many characteristics (distributions of occurrences, variability in copy number among lines, genomic control of copy number), the three elements mdg-1, I and copia were similar, but they differed from the P element. This difference in behaviour between the I and P elements reinforces the idea that the two I-R and P-M systems of hybrid dysgenesis are independent (Kidwell, 1979, 1985).

The similar behaviour of the two elements mdg-1 and copia is not surprising since they both belong to a category of transposons that have structures similar to those of retroviral proviruses, and they show a similar profile of developmental transcription (Parkhurst and Corces, 1987). These two elements differ, however, from the I element, which

Table 5 Genetic variability indexes on the whole genome, calculated on the simulated hybrids between the inbred lines. NS: labelled site number; NC: number of copies ($NC = 2NS - NH$); NS* and NC*: values including the number of insertions in the centromeric regions; NH: heterozygous site number; Het: proportion of heterozygous sites ($Het = NH/NS$); P_H : probability of heterozygous sites among all the sites observed; A : allelism as defined by Ohta (1985–1986); H : Langley's index (Langley *et al.* 1983)

Elements	Genetic variability indexes							P_H	A	H
	NS	NS*	NC	NC*	NH	Het				
Mdg-1	26.97	29.35	30.71	33.40	23.24	0.86	0.24	0.29	4.42	
I	28.12	31.69	30.24	35.06	26.12	0.92	0.21	0.19	2.88	
Copia	31.29	33.96	34.47	37.70	28.11	0.90	0.21	0.23	3.99	
P	33.17	34.92	35.36	37.20	31.35	0.94	0.18	0.16	2.83	

lacks terminal repeats, and is more like the LINE sequences in human DNA. We suggest that a criterion of discrimination between the elements is the manner in which they transpose. Indeed, DNA sequence analyses of *mdg-1*, *I* and *copia* strongly indicate that they transpose by reverse transcription (Arkhipova *et al.*, 1986; Fawcett *et al.*, 1986; Emori *et al.*, 1985), while the *P* element involves a transposase (Spradling and Rubin, 1982; Engels, 1984). It remains to be seen why the presence of a reverse transcriptase leads to the genomic control of *mdg-1*, *I* and *copia* copy numbers with all these numbers being correlated within a line, whereas the presence of a transposase leads to an independent repartition of the *P* element on the chromosomes and in the lines. It may be that some specific sequences are common to elements regulated in a similar fashion. Hence, information on the temporal transcription of each element during development, and identification of the DNA sequences involved in the regulation of transcription, should help clarify the nature of the processes that control the copy number in genomes of individuals. Note that the autosomal *P* elements may include both complete and defective elements (Engels, 1984). It is then possible that our inbred lines did not regulate their overall *P* factor copy number but used instead a mechanism inactivating some of these elements. The number of active *P* elements could then be correlated with the copy numbers of the other mobile elements.

An intriguing observation is that the variability in *P* copy number among the inbred lines is large, with values ranging from 8 to 43. At first sight, this can reveal the high polymorphism existing in individuals of the initial mass-mated population, the patterns in the inbred lines being only the consequence of segregation over generations of inbreeding of the heterozygous insertion sites (note that the average proportion of heterozygous insertion sites is equal to 0.94). That such a process may exist is undeniable though it suggests that the line No. 7, with its 43 insertions, originates from individuals with a number of *P* labelling sites as great as 90, a number much higher than has yet been found for the *P* element in a natural population. But, the absence of polymorphism within this line, as well as the high concordance in insertion patterns with its subline established 25 generations ago, strongly suggests that the 43 sites result from such segregation. Of course, as reported previously, this does not exclude a copy-number-dependent regulation of transposition or excision for some elements in some lines (Kidwell, 1985; Biémont and Aouar, 1987). Indeed, the relation-

ship between the number of copies of *mdg-1*, *I* and *copia* indicates that either the numbers of copies of these three elements were regulated simultaneously during the establishment of the inbred lines, and that no compensation existed between the elements, or that the *mdg-1*, *I* and *copia* copy numbers were already correlated in the initial individuals from which the lines were derived. This gives new insight in the observation of Rivin *et al.* (1986) that various cloned repeated sequences vary independently between inbred lines of maize. These authors did not find lines with uniformly high or low copy numbers. However, according to the great divergence in behaviour between *P* and the three other elements in our *Drosophila* inbred lines, these results in maize may be due to the great variety of repeated sequences that can behave differently. Perhaps repeated sequences, which share common peculiarities in sequences or transposition mechanism, could have been found to be correlated.

Moreover, a recent observation done at the 60th generation of inbreeding (unpublished results) reveals that the transposable element copy number of the lines depends on some "characteristics" of these lines. Indeed, working recently with the line No. 8 on *mdg-1* insertion pattern that is stable from the 15th generation, we were astonished to see changes in the insertion pattern of the *P* element. At the 52nd generation studied before, this line had 14 *mdg-1* and 18 *P*; at the 60th generation, although the pattern for *mdg-1* had not changed, the line had lost 4 *P* sites, and hence had 14 *P* copies. It is hard to admit that after 60 generations of sib crosses, a residual heterozygosity still exists, especially in this line No. 8 in which no change in *mdg-1* insertion pattern was seen since the 15th generation (Biémont and Aouar, 1987). Such observations favour better the idea that *P* elements move in some highly inbred lines, thus leading to new heterozygous sites that subsequently segregate. Such changes in *P* insertion patterns could also explain the wide scatter of the *P* elements along the chromosomes, and the quasi absence of hot spots of insertion. Hence, the ways different families of mobile element are regulated in inbred lines is one of the major questions to be solved in the near future. Consequently, more observations on the polymorphism of many different families of movable elements are necessary before a comprehensive explanation can be proposed for the ways these elements are maintained in natural populations. Indeed, our data on *mdg-1*, *I* and *P* elements do not sustain a simple model of selection

against mutational effects of mobile element insertions when proportion of insertions on the X chromosomes is taken as a test, although such model seems at first sight adequate to account for the distribution of the copia element copy number. The differences in behaviour between P and the elements mdg-1, I and copia strongly suggest that the mechanisms involved may be complex and may vary according to the element families.

Acknowledgements We thank C Arnault, A. Bucheton, P. Fouillet, J. M. Legay, A. Pelisson and D. Pontier for their various help and comments, A. Aouar for the establishment of the inbred lines, and A. Heizmann for her valuable technical assistance. This work was supported by the Centre National de la Recherche Scientifique (Laboratory No. 243) and the Association pour la Recherche sur le Cancer.

REFERENCES

- ARKHIPOVA, I. R., MAZO, A. M., CHERKASOVA, V. A., GORELOVA, T. V., SCHUPPE, N. G. AND ILYIN, Y. V. 1986. The steps of reverse transcription of *Drosophila* mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. *Cell*, **44**, 555-563.
- BELYAEVA E.Sp., ANANIEV, E. V. AND GVOZDEV, V. A. 1984. Distribution of mobile dispersed genes (mdg-1 and mdg-3) in the chromosomes of *Drosophila melanogaster*. *Chromosoma*, **90**, 16-19.
- BIEMONT, C. 1986. Polymorphism of the Mdg-1 and I mobile elements in *Drosophila melanogaster*. *Chromosoma*, **93**, 393-397.
- BIEMONT, C. AND 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.
- BIEMONT, C., BELYAEVA, E. Sp., PASYUKOVA, E. G. AND KOGAN, G. 1985. Mobile gene localization and viability in *Drosophila melanogaster*. *Experientia*, **41**, 1474-1476.
- BIEMONT, C. AND GAUTIER, C. 1987. Mdg-1 mobile element heterozygosity in *Drosophila melanogaster*. *Heredity*, **58**, 167-172.
- BIEMONT, C. AND TERZIAN, C. 1986. Regulation in the number of Mdg-1 mobile elements in inbred *Drosophila melanogaster*. *Genetica*, **71**, 161-165.
- BREGLIANO, J. C., PICARD, G., BUCHETON, A., PELISSON, A., LAVIGE, J. M. AND L'HERITIER, P. 1980. Hybrid dysgenesis in *Drosophila melanogaster*. *Science*, **207**, 606-611.
- BROOKFIELD, J. F. Y. 1986. The population biology of transposable elements. *Phil. Trans. R. Soc. Lond.*, **312**, 217-226.
- BUCHETON, A., PARO, R., SANG, H. M., PELISSON, A. AND FINNEGAN, D. J. 1984. The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. *Cell*, **38**, 153-163.
- CHARLESWORTH, D. 1983. Models of the evolution of some genetic systems. *Proc. R. Soc. London*, **219**, 265-279.
- CHARLESWORTH, B. AND CHARLESWORTH, D. 1983. The population dynamics of transposable elements. *Genet. Res.*, **42**, 1-27.
- CHARLESWORTH, B. AND LANGLEY, C. H. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics* **112**, 359-383.
- CHARLESWORTH, B., LANGLEY, C. H. AND STEPHAN, W. 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics*, **112**, 947-962.
- DUNSMUIR, P., BROREIN, W. J., SIMON, M. A. AND RUBIN, G. M. 1980. Insertion of the *Drosophila* transposable element Copia generates a 5 base pair duplication. *Cell*, **21**, 575-579.
- EMORI, Y., SHIBA, T., KANAYA, S., INOUE, S., YUKI, S. AND SAIGO, K. 1985. The nucleotide sequences of Copia-related RNA in *Drosophila* virus-like particles. *Nature*, **315**, 773-776.
- ENGELS, W. R. 1984. A trans-acting product needed for P factor transposition in *Drosophila*. *Sciences*, **226**, 1194-1196.
- FAWCETT, D. H., LISTER, C. K., KELLET, E. AND FINNEGAN, D. J. 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. *Cell*, **47**, 1007-1015.
- FITZPATRICK, B. J. AND SVED, J. A. 1986. High levels of fitness modifiers induced by hybrid dysgenesis in *Drosophila melanogaster*. *Genet. Res.*, **48**, 89-94.
- GREEN, M. M. 1969. Controlling element mediated transposition of the white gene in *Drosophila melanogaster*. *Genetics*, **61**, 423-428.
- ILYIN, Y. V., CHMELIAUSKAITE, V. G. AND GEORGIEV, G. P. 1980. Double-stranded sequences in RNA of *Drosophila melanogaster*: relation to mobile dispersed genes. *Nucleic Acid Res.*, **8**, 3439-3457.
- JUNAKOVIC, N., DI FRANCO, C., BASSANTI, P. AND PALUMBO, G. 1986. Transposition of Copia-like nomadic elements can be induced by heat shock. *J. Mol. Evol.*, **24**, 89-93.
- KIDWELL, M. G. 1979. Hybrid dysgenesis in *Drosophila melanogaster*: the relationships between the P-M and I-R interaction systems. *Genet. Res.*, **33**, 205-217.
- KIDWELL, M. G. 1985. Hybrid dysgenesis in *Drosophila melanogaster*: nature and inheritance of P element regulation. *Genetics*, **111**, 337-350.
- LANGER-SAFER, P. R., LEVINE, M. AND WARD, D. C. 1982. Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci.*, **79**, 4381-4385.
- LANGLEY, C. H., BROOKFIELD, J. F. Y. AND KAPLAN, N. 1983. Transposable elements in Mendelian populations. I. A theory. *Genetics*, **104**, 457-471.
- LEFEVRE, G. 1976. A photographic representation and interpretation of the polytene chromosomes of *Drosophila melanogaster* salivary glands. ASHBURNER, M. and E. NOVITSKI (eds). In *The Genetics and Biology of Drosophila*, Vol. 1a, Academic Press, Orlando.
- LEHMANN, E. L. 1975. *Nonparametrics: Statistical Methods Based on Ranks*. McGraw-Hill.
- LEIGH BROWN, A. J. AND MOSS, J. E. 1987. Transposition of the I element and Copia in a natural population of *Drosophila melanogaster*. *Genet. Res.*, **49**, 121-128.
- LEVIS, R., DUNSMUIR, P. AND RUBIN, G. M. 1980. Terminal repeats of the *Drosophila* transposable element Copia: nucleotide sequence and genomic organisation. *Cell*, **21**, 581-579.
- MACKAY, T. F. C. 1985. Transposable element-induced response to artificial selection in *Drosophila melanogaster*. *Genetics*, **111**, 351-374.
- MACKAY, T. F. C. 1986. Transposable element-induced fitness mutations in *Drosophila melanogaster*. *Genet. Res.*, **48**, 77-87.
- MONTGOMERY, E. A. AND LANGLEY, C. H. 1983. Transposable elements in Mendelian populations. II. Distribution of copia-like elements in natural populations. *Genetics*, **104**, 473-483.

- MONTGOMERY, E. A., CHARLESWORTH, B. AND 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*. *Genet. Res.*, **49**, 31-41.
- MUKAI, T. AND YUKUHIRO, K. 1983. An extremely high rate of deleterious viability mutations in *Drosophila* possibly caused by transposons in non-coding regions. *Proc. Jap. Acad. Sci.*, **59**, 316-319.
- O'HARE, K. AND 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.
- OHTA, T. 1985. A model of duplicative transposition and gene conversion for repetitive DNA families. *Genetics*, **110**, 513-524.
- OHTA, T. 1986. Population genetics of an expanding family of mobile genetic elements. *Genetics*, **113**, 145-159.
- PAQUIN, C. E. AND WILLIAMSON, V. M. 1984. Temperature effects on the rate of Ty transposition. *Science*, **226**, 53-55.
- PARKHURST, S. M. AND CORCES, V. G. 1987. Developmental expression of *Drosophila melanogaster* retrovirus-like transposable elements. *EMBO*, **6**, 419-424.
- PASYUKOVA, E. G., BELYAEVA, E. Sp., KOGAN, G. L., KAIDANOV, L. Z. AND GVOZDEV, V. A. 1986. Concerted transpositions of mobile genetic elements coupled with fitness changes in *Drosophila melanogaster*. *Mol. Biol. Evol.*, **3**, 299-312.
- PIERCE, D. A. AND LUCCHESI, J. C. 1981. Analysis of a dispersed repetitive DNA sequence in isogenic lines of *Drosophila melanogaster*. *Chromosoma*, **82**, 471-492.
- RIGBY, P. W., DIECKMANN, M., RHODES, C. AND BERG, P. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.*, **113**, 237-251.
- RIVIN, C. J., CULLIS, C. A. AND WALBOT, V. 1986. Evaluating quantitative variation in the genome of *Zea mays*. *Genetics*, **113**, 1009-1019.
- ROLFE, M., SPANOS, A. AND BANKS, G. 1986. Induction of yeast Ty element transcription by ultraviolet light. *Nature*, **319**, 339-340.
- RONSSERAY, S. AND ANXOLABEHERE, D. 1986. Chromosomal distribution of P and I transposable elements in a natural population of *Drosophila melanogaster*. *Chromosoma*, **94**, 433-440.
- SPRADLING, A. C. AND RUBIN, G. M. 1982. Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Sciences*, **218**, 341-347.
- STRAND, D. J. AND McDONALD, J. F. 1985. Copia is transcriptionally responsive to environmental stress. *Nucleic Ac. Res.*, **13**, 4401-4410.
- STROBEL, E., DUNSMUIR, P. AND RUBIN, G. M. 1979. Polymorphisms 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. AND GEORGIEV, G. P. 1981. General properties of mobile dispersed genetic elements in *Drosophila*. *Cold Spring Harbor Sympos. Quant. Biol.*, **45**, 655-665.
- YUKUHIRO, K., HARADA, K. AND MUKAI, T. 1985. Viability mutations induced by the P elements in *Drosophila melanogaster*. *Jpn. J. Genet.*, **60**, 531-537.