copia Expression Is Variable Among Natural Populations of Drosophila

Amy K. Csink and John F. McDonald

Department of Genetics, University of Georgia, Athens, Georgia 30602

Manuscript received June 26, 1989
Accepted for publication June 18, 1990

ABSTRACT

A survey of *copia* (retroviral-like element) expression in flies representing 37 populations worldwide of *Drosophila melanogaster*, *Drosophila simulans* and *Drosophila mauritiana* demonstrates that, although *copia* elements are present in all three species, *copia*-encoded transcripts are detectable only in *D. melanogaster*. Levels of *copia* transcripts vary nearly 100-fold among flies representing geographically diverse populations of *D. melanogaster* and this variation is not correlated with variability in *copia* copy number. Analysis of transcript levels in interpopulation hybrids demonstrates that much of this variability may be attributable to the action of *trans*-acting controls. The geographic and phylogenetic pattern of *copia* expression suggests that moderate to high levels of *copia* expression may be a relatively recent evolutionary acquisition. The potential evolutionary significance of these findings is discussed.

THE largest and most widely distributed group of eukaryotic transposons consists of elements whose structure and mode of replication parallel those of mammalian retroviruses (BALTIMORE 1985). Retroviral-like transposable elements (RLEs) have been identified within the genomes of vertebrates (*e.g.*, CALLAHAN *et al.* 1985; BENVENISTE 1985; ANDERSON *et al.* 1979), invertebrates (*e.g.*, FINNEGAN and FAWCETT 1986), plants (JOHNS, MOTTINGER and FREELING 1985) and yeast (ROEDER and FINK 1983) and it is likely that they are constituents of most, if not all, eukaryotic genomes.

RLEs possess a number of qualities which suggest that they may be particularly important in the process of adaptive evolution (*e.g.* MCDONALD 1990, 1989; STAVENHAGEN and ROBINS 1988; FINNEGAN 1989; MCDONALD *et al.* 1987; Temin 1982; STEELE 1979). For example, RLEs are known to be a significant source of spontaneous mutation (LAMBERT, MCDONALD and WEINSTEIN 1988). In addition, there is a growing body of evidence that rates of RLE transposition may be subject to host-mediated controls, including those involved in the eukaryotic stress response systems (*e.g.*, MCDONALD *et al.* 1987; MCCLANAHAN and MCENTEE 1986; STRAND and MCDONALD 1985). Perhaps most significant from the evolutionary perspective, however, is the fact that RLE insertion mutants often display novel regulatory phenotypes which are subject to phenotypic suppression by chromosomal genes (KUBLI 1986). These unique properties of RLEs suggest that they may play an important role in the evolution of new regulatory and developmental networks (MCDONALD 1990).

As interesting and provocative as the unique properties of RLEs may be, their true evolutionary potential can only be evaluated after we have acquired greater knowledge of their frequency and function in natural populations. Towards this end we have initiated an extensive analysis of RLE expression in natural populations of Drosophila. In this paper we report the results of a survey of variation of *copia* copy number and transcript levels which exists among 38 geographically diverse populations of *Drosophila melanogaster* and its sibling species *Drosophila simulans* and *Drosophila mauritiana*. The results demonstrate that significant intra- and interspecific variation in *copia* transcript levels exists among natural populations in the *melanogaster* subgroup that cannot be accounted for by variability in *copia* element copy number. In addition, we present evidence that much of the variability in *copia* expression existing among natural populations is attributable to the action of trans-acting regulatory controls.

MATERIALS AND METHODS

Drosophila strains and production of interpopulation hybrids: Strains listed in Table 1 were established from 15–50 gravid females collected in the wild (collection dates and locations are presented in Table 1). Their descendants were maintained as mass laboratory cultures at 22° on cornmeal-molasses medium. Strain 30 (Napa, California) is a homozygous stock which has been described previously (ANDERSON and MCDONALD 1983).

The parents used to produce the interpopulation hybrids were derived from iso-female lines established from the mass cultures representing four geographically isolated populations of *D. melanogaster*. The hybrid crosses were made by placing five virgin females and five males in individual vials containing standard Drosophila medium. Six replicate sets of crosses were made for each interpopulation hybrid.

DNA preparation, restriction and blot analysis: Genomic DNA was isolated from approximately 1000 adult flies by a previously described method (BINGHAM, LEVIS and RUBIN 1981). Genomic DNA was digested with the restriction enzyme *Hin*dIII according to the supplier's instructions

TABLE 1

Variation in *copia* relative copy number and transcript levels (5 kb) among natural populations of Drosophila

Population locale	Date collected	Relative amount of *copia* DNA (full)	DNA (internal)	RNA
D. melanogaster				
1) Australia, Melbourne	1983	0.57	0.54	0.33
2) Tahiti	1983	0.47	0.51	0.23
3) S. Africa, Capetown	1984	0.46	0.39	0.26
4) S. Africa, Johannesburg	1984	0.64	0.63	0.16
5) Congo, Brazzaville	1982	0.49	0.58	0.11
6) Congo, Brazzaville (Primus)	1985	0.87	0.65	0.20
7) Congo, Brazzaville (Kronenbourg)	1985	0.81	0.63	0.15
8) Congo, Loua	1986	0.60	0.56	0.03
9) Congo, Loukanga	1987	0.74	0.66	0.07
10) Congo, Dimonika	1983	0.62	0.56	0.10
11) Burundi, Bujumbura	1984	0.57	0.55	0.25
12) Tanzania, Nyenges	1984	0.47	0.63	0.17
13) Benin, Cotonou	1982	0.45	0.45	0.17
14) Ivory Coast, Tai Forest	1981	0.66	0.60	0.22
15) Ivory Coast, Lamto	1984	0.43	0.35	0.08
16) Dahomey	1972	0.70	0.85	0.67
17) Sierra Leone, Freetown	1981	0.70	0.63	0.81
18) Tunisia, Nasrallah	1984	0.42	0.51	0.23
19) France, Paris	1984	0.57	0.63	0.17
20) USSR, Dushnabe	1981	0.63	0.70	0.32
21) USSR, Dilizhan	1966	0.67	0.82	0.29
22) USSR, Krasanodak	1980	0.70	0.81	0.11
23) Argentina, Sato	1980	0.97	1.00	0.20
24) Chile, Santiago	1957	0.48	0.36	0.16
25) Peru, Tacna	1980	0.43	0.46	0.26
26) Peru, Lima	1982	0.55	0.54	1.00
27) Peru, Iquitos	1982	0.56	0.53	0.94
28) Guadeloupe	1983	0.52	0.54	0.14
29) Georgia, Athens	1986	0.81	0.61	0.25
30) California, Napa	1975	1.00	0.62	0.42
D. simulans				
31) New Calidonia	1983	0.31	0.33	0.00
32) Seychelle Islands	1983	0.25	0.32	0.00
33) Congo, Brazzaville	1982	0.23	0.29	0.00
34) French Guiana	1981	0.33	0.35	0.00
35) Tunisia, Nasrallah	1984	0.31	0.33	0.00
36) France, Antibes	1983	0.28	0.29	0.00
D. mauritiana				
38) Maurituis	1983	0.42	0.43	0.00

Each value is an average of the three densitometry readings and has been divided into the maximum value in each column in order to facilitate reading. The F values for the ANOVAs (see MATERIALS AND METHODS) for each of the three data sets is as follows: DNA (full length probe), $F = 14.729$, d.f. = 36, DNA (internal probe), $F = 11.653$, d.f. = 36, RNA, $F = 2.890$, d.f. = 29. See MATERIALS AND METHODS for details.

(New England Biolabs). Digested DNA samples were fractionated through 0.8% agarose (Sigma) gels, and transferred to nylon filters (Amersham-Hybond) according to the method of SOUTHERN (1975) (Figure 1A). Relative *copia* copy number was determined by a modification of the slot blot procedure of RIVEN, CULLIS and WALBOT (1986). Genomic DNA from each population was serially diluted twofold (1, 0.5 and 0.25 μg) and transferred to nitrocellulose using a slot-blot apparatus (Figure 1B).

Filters were prehybridized in 5× SSC, 5× Denhardt's solution, 50% formamide, 200 mg/ml denatured salmon sperm DNA, 50 mM Na_2PO_4, and 10% dextran sulfate for 12 hr at 42°. After prehybridization, the ^{32}P-labeled (FEINBERG and VOGELSTEIN 1984) *copia* probe was added. Two *copia* probes were used in this study. The "full-length" probe was the genomic clone Dm5002 previously described by LEVIS, DUNSMUIR and RUBIN (1980). An "internal" probe was the *Apa*I-*Hpa*I fragment of Dm5002 encompassing *copia* coding sequences but not the LTRs (Figure 1C). Hybridization was allowed to proceed for 16 to 20 hr at 42°. Washes were carried out at 60°. Slot-blot autoradiographs were quantitated by densitometry (see below).

RNA preparation and Northern analysis: Total cellular RNA was isolated from 50–200 adult flies (7–10 days post-eclosion) representing each population, according to STRAND and MCDONALD (1985). RNA was electrophoresed through 1% agarose-formaldehyde gels and transferred to nitrocellulose filters according to THOMAS (1980). Hybridization procedures were as described above except that all washes were carried out at 42°. Replicate gels were run on

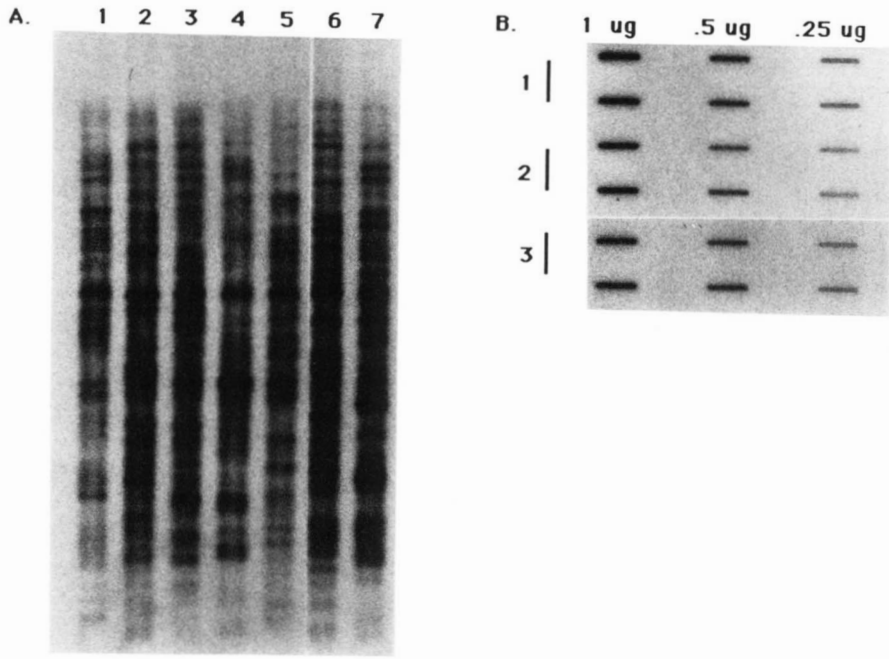

FIGURE 1.—(A) Southern blot of *Hin*dIII-digested genomic DNA isolated from natural populations of *D. melanogaster* and hybridized against a *copia* probe. Lane designations: 1, Capetown, South Africa; 2, Johannesburg, South Africa; 3, Nyenges, Tanzania; 4, Brazzaville, Congo; 5, Loua, Congo; 6, Kronenbourg, Congo; 7, Dimonika, Congo. (B) Example of slot-blots used to quantitate relative *copia* copy number among natural populations of Drosophila. Shown are 2× serially diluted replicate DNA samples isolated from the following populations and hybridized against a *copia* probe: 1, Kronenbourg, South Africa; 2, Dimonika, Congo; 3, Lamto, Ivory Coast. (C) Diagram of the *copia* element found in the plasmid Dm5002 showing the internal *Apa*I-*Hpa*I fragment used to probe slot blots in order to obtain an estimate of full-length *copia* elements in the genomes.

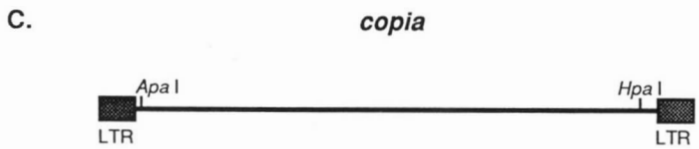

TABLE 2

Analysis of variance table for the relative amount of *copia* RNA as determined from the densitometric scans of the northern blot autoradiograms (see MATERIALS AND METHODS)

Source	Degrees of freedom	Sum of squares	Mean squares	*F* value	*P*
Strain	29	3.2115	0.1107	2.890	<0.001
Replicate	2	1.6003	0.8001	20.890	<0.001
Error	58	2.2189	0.0383		
Total	89	7.0307			

three separately purified RNA samples per population. Integrity of RNA samples was determined by the quality of the ethidium bromide-stained ribosomal bands and/or by rehybridization of filters with a Drosophila β-tubulin probe (BIALOJAN, FAULKENBURG and RENKAWITZ-POHL 1985). Quantitation of autoradiographs was by densitometry as described below.

Quantitation and statistical analysis: *Northern blots*: Three replicate RNA samples from each of the strains representing 30 geographically diverse *D. melanogaster* populations were run on 3 separate 30 lane denaturing gels (1 sample from each gel). Autoradiographs were densitometrically scanned on a Beckman DU-8 spectrophotometer using the Gel Scan program. The 3 replicate densitometry readings were used in an analysis of variance (ANOVA) employing a randomized block design (Table 2). For comparative purposes, Table 1 presents the average of the 3 RNA values obtained for each strain normalized to that of strain #26 (Lima, Peru) which contains the highest level of *copia* transcripts of all the populations surveyed. The relative RNA values for the *D. simulans* and *D. mauritiana* populations were obtained by running 2 gels with 2 samples from strains representing each geographical location. In addition, a *D. melanogaster* sample (strain #30, Napa, California) was included on each gel as an internal control.

DNA slot-blots: DNA samples were extracted from a pool of at least 1000 individuals per strain. Two replicate sets of these serially diluted DNA samples (1, 0.5 and 0.25 μg) were transferred to nitrocellulose using a slot-blot apparatus (Schleicher and Schuell, Keene, New Hampshire) (Figure 1B). Interstrain variation was analyzed by ANOVA using a randomized block design.

For comparative purposes, Table 1 presents the average DNA values normalized to the strain having the highest *copia* DNA content (column 1 presents values obtained by hybridization to a full-length *copia* probe and normalized to the Napa, California, sample; column 2 presents values obtained by hybridization to an internal *copia* probe and normalized to the Sato, Argentina, sample).

Correlations: The correlations between DNA:DNA and RNA:DNA were calculated based on least squares estimate using the values in Table 1.

Interpopulation hybrids: The data sets obtained from the four interpopulation crosses were subjected to ANOVA as described above. The Scheffe's F-test was used to determine (1) if RNA levels were significantly different between males and females and (2) if there was a significant difference in RNA levels in the F_1 progeny depending upon the direction of the cross (*i.e.*, depending upon whether a given population provided the male or female parent).

Under a codominant model of inheritance, the expected mean RNA level in the F_1 progeny will be mid-way between the parental means (*i.e.* (P1 + P2)/2) which we refer to as the mid-parent value. The mid-parent variances were estimated from the observed parental variances according to

the methods of HOGG and CRAIG (1970). The mid-parent variances were employed in *t*-tests to determine if the progeny values were within the expected range of each of the mid-parents.

The StatView Program of Abacus Concepts, Inc. (Berkeley, California) was used for all statistical computations.

***In situ* hybridizations:** The *in situ* hybridizations followed the basic procedure of PARDUE and GALL (1975), as modified by LANGER-SAFER, LEVINE and WARD (1988) using biotinylated DNA (Bethesda Research Labs) to probe polytene chromosomes. The probe used was the *Apa*I-*Hpa*I fragment from Dm5002 (Figure 1C). A complex of avidin and biotinylated horseradish peroxidase was bound to the biotinylated DNA, and a dye-coupled reaction with peroxidase substrate gave a black band of precipitation at the site of hybridization.

RESULTS

***Copia* DNA copy number varies by 2–3-fold among geographically distinct populations of *D. melanogaster*:** Southern blots of *Hin*dIII-digested DNA isolated from flies representing each of the populations examined in our survey and hybridized against a *copia* probe display the multiple banding pattern typical of dispersed middle repetitive elements (Figure 1A). Relative copy number of these dispersed elements was determined by a quantitative slot-blot procedure using DNA samples isolated from each of the mixed cultures representing 30 geographically distinct populations of *D. melanogaster* (Figure 1B). The normalized mean values are presented in Table 1. An ANOVA of the data set demonstrates that there is significant variation between populations (F = 14.73, d.f. = 36, $P < 0.001$). The overall magnitude of variability in relative *copia* copy number among the populations surveyed is also remarkably low. The majority of the normalized values cluster between 0.4 and 0.7 with only a few outlying values from 0.8 to 1.0. In general, these data are consistent with earlier findings that *copia* copy number among independently established strains of *D. melanogaster* is relatively constant (DOWSETT and YOUNG 1982).

copia sequences may exist within the Drosophila genome as full length (5 kb) elements or as independent long terminal repeats (LTRs). Free LTRs have been detected within a number of eukaryotic genomes and are believed to be the remnants of excision events resulting from recombination between a retroviral element's 5′ and 3′ LTRs (*e.g.*, ROEDER and FINK 1983; FINNEGAN and FAWCETT 1985). Since these free LTRs may, in principle, exist in substantial numbers within eukaryotic genomes, we decided to determine if our relative measure of *copia* copy number among populations would be significantly altered if we repeated the quantitations using a *copia* probe which did not contain sequences homologous to the *copia* LTR. Toward this end, we isolated and subcloned a fragment from the internal (non LTR-containing) region of the *copia* element (*Apa*I-*Hpa*I) contained within the plasmid Dm5002 (DUNSMUIR *et al.* 1980) (Figure 1C). With this probe, we were able to estimate the relative number of full-length *copia* elements within the genomes surveyed.

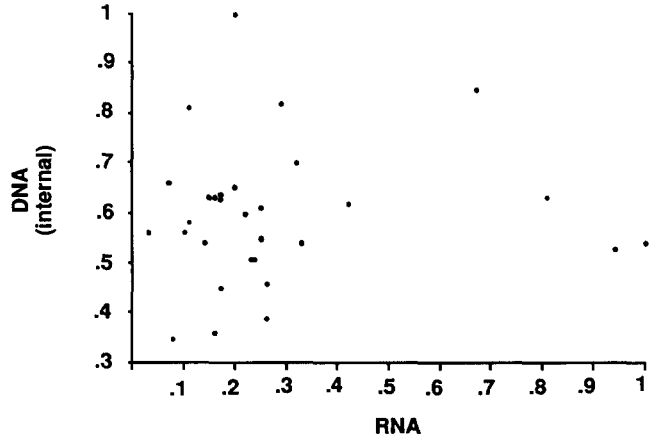

FIGURE 2.—Plot of relative *copia* copy number (*i.e.*, relative amount of *copia* DNA as determined by slot-blot hybridization using the internal *copia* probe) *vs.* relative abundance of 5-kb *copia* transcripts. Values are found in Table 1. $r = 0.072$.

Relative *copia* copy number rankings, determined by using the full-length and the internal *copia* probes, are presented in Table 1. The overall correlation between the two sets of values is highly significant (r = 0.763, $P < 0.001$) indicating that either there are relatively few free *copia* LTRs within the *melanogaster* genomes (which is consistent with earlier results; see LEWIS, DUNSMUIR and RUBIN 1980) and/or that the number of free LTRs present is positively correlated with the number of full length *copia* elements.

Levels of *copia* RNA vary approximately 100-fold among geographically distinct populations of *D. melanogaster* and are not correlated with *copia* copy number variation: To estimate transcriptional variation for *copia* in natural populations of Drosophila, we measured the levels of full length (5 kb) *copia* transcripts present in flies representing the populations described above (Figure 3). Three replicate samples of total RNA independently isolated from adult flies representing each population were loaded on denaturing agarose gels, transferred to nitrocellulose filters and hybridized with a full length *copia* probe. The abundance of 5kb *copia* transcripts was quantitated by densitometric scanning of the resulting autoradiographs (Table 1).

An ANOVA performed on the complete data set demonstrated that the between-population variation was again highly significant (F = 2.89, d.f. = 29, $P <$ 0.001) (Table 2). Overall variation in *copia* transcript levels between the *melanogaster* populations surveyed was substantially higher than that observed for *copia* copy number. Mean values varied nearly 100-fold (0.03–1.00). In addition, of the 7 outlier populations, 3 were associated with exceptionally high levels of

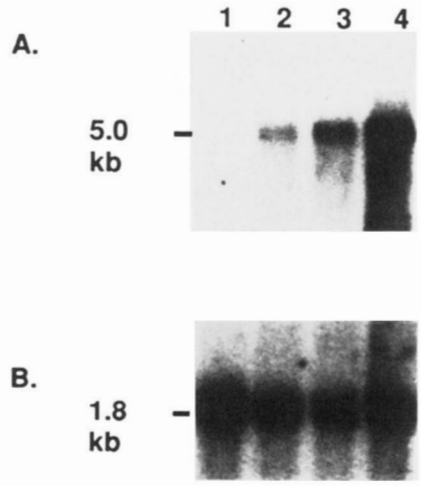

FIGURE 3.—(A) Northern blot of total RNA isolated from adult *D. melanogaster* representing various geographic populations. Blots were hybridized against a full-length *copia* probe (Dm5002). Lane designations: 1, Loua, Congo; 2, Loukanga, Congo; 3, Kronenbourg, Congo; 4, Iquitos, Peru. (B) Same Northern blot as in A which has been rehybridized against a Drosophila β-tubulin probe.

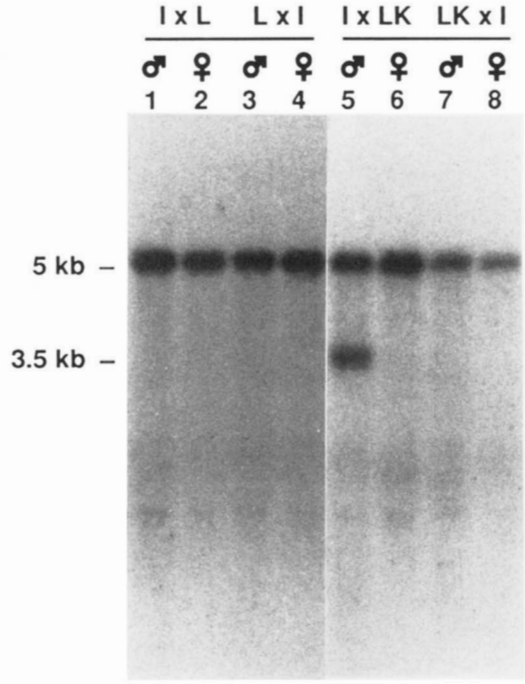

FIGURE 4.—Northern blots of total RNA isolated from adult male (♂) and female (♀) F_1 progeny from reciptocal crosses of two sets of interpopulation matings. I × L = progeny Iquitos ♀♀ × Loua ♂♂, L × I = progeny of Loua ♀♀ × Iquitos ♂♂; I × LK = progeny of Iquitos ♀♀ × Loukanga ♂♂, LK × I = progeny of Loukanga ♀♀ × Iquitos ♂♂. Blots were hybridized against a full-length *copia* probe (Dm5002).

copia transcripts (0.67–1.00) and 4 with extremely low levels (0.03–0.10). The majority of the populations surveyed, however, displayed moderate levels of *copia* RNA (0.15–0.33). Those strains displaying exceptionally high levels of *copia* transcripts represent 3 geographically diverse populations (Lima, Peru; Iquitos, Peru; Porto Novo, Benin). In contrast, all of the populations associated with very low levels of *copia* transcripts are located in central regions of Africa (Loua, Congo; Loukanga, Congo; Dimonika, Congo; Lamto, Ivory Coast).

To assess the possibility that the variation in *copia* transcript levels among populations was simply a function of the difference in the number of full length *copia* elements, the relative *copia* transcript levels representing each population was plotted against the relative number of *copia* elements present. Low correlation was observed (Figure 2, $r = 0.072$). Thus, the substantial variation in *copia* RNA abundance that exists among natural populations of *Drosophila* cannot be accounted for by variability in the number of *copia* elements present within the respective genomes.

Variability in *copia* transcript levels among *D. melanogaster* populations cannot be accounted for exclusively by *cis*-acting effects: Mechanisms that regulate transcript abundance can be divided into two categories: (1) *cis*-acting effects which map to the gene itself or are linked to the gene in question or (2) *trans*-acting effects which map to a second gene. In an initial effort to discern the genetic basis of the variation in *copia* transcript levels which is segregating in natural populations, we conducted a series of cis/trans tests between flies representing several of the populations screened in our preliminary survey. These cis/trans tests consisted of producing hybrids between individuals with different patterns of gene expression and examining the pattern of expression present in the hybrids (Figures 4 and 5).

Six replicate sets of matings were carried out for each of four interpopulation crosses investigated. The results, which are graphically presented in Figure 5, demonstrate 3 distinct patterns of inheritance which indicate that both *cis*- and *trans*-regulatory genetic variation for *copia* expression exists among natural populations of Drosophila.

The first pattern is illustrated by the Loua, Congo × Iquitos, Peru mating in which the progeny *copia* RNA levels are significantly different from either parent but lie within the expected range of the mid-parent (Figure 4, lanes 1–4; Figure 5A). This *cis*-inherited pattern of inheritance is consistent with the variability in levels of *copia* expression between flies representing the Loua and Iquitos populations being due to a difference in the number of transcriptionally active *copia* elements and/or to the presence of trans-acting regulatory differences which are inherited in an additive fashion.

A second pattern of inheritance is illustrated by the Loua × Kronenbourg, Congo (Figure 5B) and the Loukanga, Congo × Iquitos, Peru (Figure 4, lanes 5–6; Figure 5C) matings in which the progeny display *copia* RNA levels which are significantly lower than the mid-parent range, but not significantly different

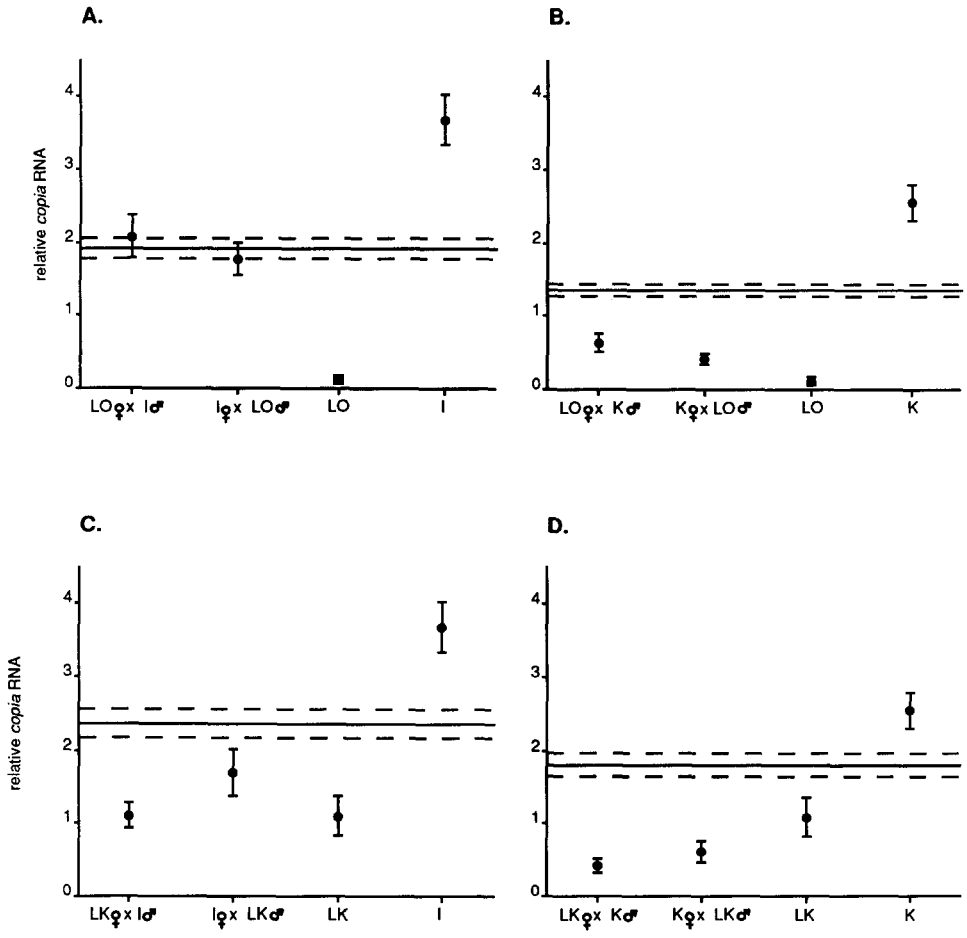

FIGURE 5.—Graphic representation of the relative levels of *copia* RNA present in parental flies representing different geographic populations of *D. melanogaster* and in interpopulation (F_1) hybrids. In each graph the F_1 progeny are (left to right) the first two points and the parents are the last two points. The vertical bars represent the 95% confidence intervals. The solid black horizontal line represents the expected mid-parent value and the dashed lines delimit the expected mid-parent 95% confidence intervals (See MATERIAL AND METHODS for details). LO = Loua, Congo; I = Iquitos, Peru; K = Kronenbourg, Congo; LK = Kronenbourg, Congo.

from the low parent value. These results indicate that the major component of the variability in *copia* expression between the Loua and Kronenbourg, and the Loukanga and Iquitos parental lines is due to a dominant trans-acting repressor gene or genes.

The third pattern of inheritance which we have observed is illustrated by the Loukanga × Kronenbourg mating in which the progeny displayed *copia* RNA levels that are not only lower than the expected mid-parent range but also significantly lower that the low parent value (Figure 5D). This pattern of inheritance suggests that the difference in *copia* expression between the Loukanga and Kronenbourg parental lines is due to some form of over-dominant trans-acting repression. Mapping experiments are currently underway to further define the nature of these *trans*-regulatory effects.

The influence of sex and the direction of interpopulational crosses on *copia* expression is strain dependent: To further explore the genetic basis of naturally occurring variability in *copia* expression, we carried out reciprocal matings for each of the four interpopulational crosses discussed in the previous section and monitored the amount and sizes of *copia* transcripts present in the male and female parents and the F_1 progeny. Our results indicate that although both sex and the direction of the cross can exert a significant influence on *copia* expression, the genetic factors mediating these effects may be complex and variable among populations.

The northern blots displayed in Figure 4 show that the abundance of 5-kb *copia* transcripts in the progeny of crosses between representatives of the Loua (Congo) × Iquitos (Peru) populations is essentially identical regardless of the sex of the flies or the direction of the cross (lanes 1–4). This finding is representative of what we observed in the F_1 progeny of all of the interpopulational crosses except one. The exception was the cross between representatives of the Loukanga (Congo) × Iquitos (Congo) populations. In three out of the six replicate sets of crosses made between these populations, we observed a significant influence of the direction of the cross on *copia* transcript levels present in the F_1 progeny (lanes 5–8). Lower levels of full-length *copia* transcripts were present in both male and female progeny of mothers from the Loukanga strain (lanes 5 and 6) than in progeny of the reciprocal cross.

In addition to this quantitative effect, we also observed a 3.5-kb *copia* transcript in the male progeny

of two out of the six replicate sets of Lokanga × Iquitos crosses (Figure 4, lane 5). In contrast to the quantitative effect described above, the presence of the 3.5-kb transcript was observed only in male progeny of crosses where the mother was taken from the Iquitos strain. Our findings are consistent with the appearance of the 3.5-kb transcript being controlled by an *X*-linked recessive factor(s) which is apparently segregating within the Iquitos isofemale line. We believe that this effect is most likely regulatory in nature (possibly a gene encoding a *trans*-processing function) because if the 3.5-kb transcript were merely the product of a defective but transcriptionally active *X*-linked *copia* element, we would expect to have detected the shortened transcript in both the male and female progeny. Moreover, we have been unable to detect the 3.5-kb transcript in flies taken from either the Loukanga or Iquitos parental (isofemale) stocks or in the progeny of the other interpopulation crosses in which the Loukanga or Iquitos flies were used as one of the parents. Collectively these two observations suggest that the appearance of the 3.5-kb band may be dependent upon a genetic factor(s) segregating in the Loukanga strain.

Although a complete understanding of the genetic basis of the above phenomena will require further study, the fact that they appeared in approximately half of the progeny of only one of the four interpopulation crosses analyzed indicates that the genetic factors involved are variable in natural populations.

***D. simulans* and *D. mauritiana* contain low to moderate numbers of *copia* elements but no detectable levels of *copia* transcripts:** In attempting to evaluate patterns of intraspecific genetic variation, it is often useful and informative to view the data within a broader phylogenetic and evolutionary context. For example, patterns of variation which are shared by closely related species are often presented as evidence that the genetic similarities existed prior to the evolutionary divergence of the species in question. In order to provide a context within which to evaluate our data on the patterns of *copia* variation among populations of *D. melanogaster*, we extended our survey to include populations of two closely related species within the *melanogaster* subgroup, *D. simulans* and *D. mauritiana*. Like *D. melanogaster*, *D. simulans* is a cosmopolitan species having a wide geographic distribution. In contrast, *D. mauritiana* is ecologically and geographically quite distinct, being limited in its distribution to the island of Mauritius off the southeast coast of Africa.

The results of our analysis, which are summarized in Table 1, demonstrate that *D. mauritiana* contains nearly as many *copia*-homologous DNA sequences as *D. melanogaster*. In contrast, *D. simulans* possesses, on average, only about half as many homologous sequences as the other two sibling species. In light of this, it is interesting that we were unable to detect any *copia*-homologous transcripts within the *D. simulans* or *D. mauritiana* flies examined.

The chromosomal distribution of *copia* elements in low transcribing *D. melanogaster* and nontranscribing *D. simulans* flies is not restricted to constitutive heterochromation: In general, transcriptionally active genes are not found in heterochromatic regions of the genome (LEWIN 1987). Previous studies have shown that mobile genetic elements which are inserted within heterochromatic regions of the Drosophila genome are often structurally defective and transcriptionally silent (*e.g.*, BUCHETON *et al.* 1984; LANSMAN *et al.* 1985). In order to determine if low levels or lack of *copia* transcripts in natural populations of Drosophila might be associated with the sequestering of elements in centric heterochromatin, we conducted an *in situ* analysis of polytene chromosomes in larvae representing the Dimonika Congo *D. melanogaster* and Seychelles *D. simulans* populations. The results, presented in Figure 6, demonstrate that 21 to 24 *copia* elements are present in the genomes of the *D. melanogaster* larvae examined while approximately half that number of elements are present within the *D. simulans* genome surveyed. In both species, however, the chromosomal distribution of *copia* elements was not limited to centric heterochromatin, but positioned randomly throughout the genome. The random distribution of *copia* elements have been previously observed in Drosophila strains where *copia* elements are transcriptionally active (YOUNG and SCHWARTZ 1981).

DISCUSSION

Naturally occurring variation in *copia* expression cannot be fully explained by variation in *copia* copy number: We have found that the variability in levels of *copia* transcripts present among natural populations of Drosophila is not correlated with variation in *copia* copy number. It is possible that significant numbers of the *copia* elements present within naturally occurring Drosophila genomes are structurally defective or otherwise transcriptionally inactive. Against such a scenario, any correlation which may exist between the number of active *copia* elements and *copia* RNA levels could be masked by the presence of a large number of inactive elements. One possibility is that transposable elements may become sequestered within transcriptionally silent regions of centric heterochromatin, accumulate mutations, and thus became biologically inactive (BUCHETON *et al.* 1984; LANSMAN *et al.* 1985). However, our *in situ* analysis of the chromosomal positions of *copia* elements within low transcribing *D. melanogaster* and nontranscribing *D. simulans* genomes seems inconsistent with the hypothesis that the

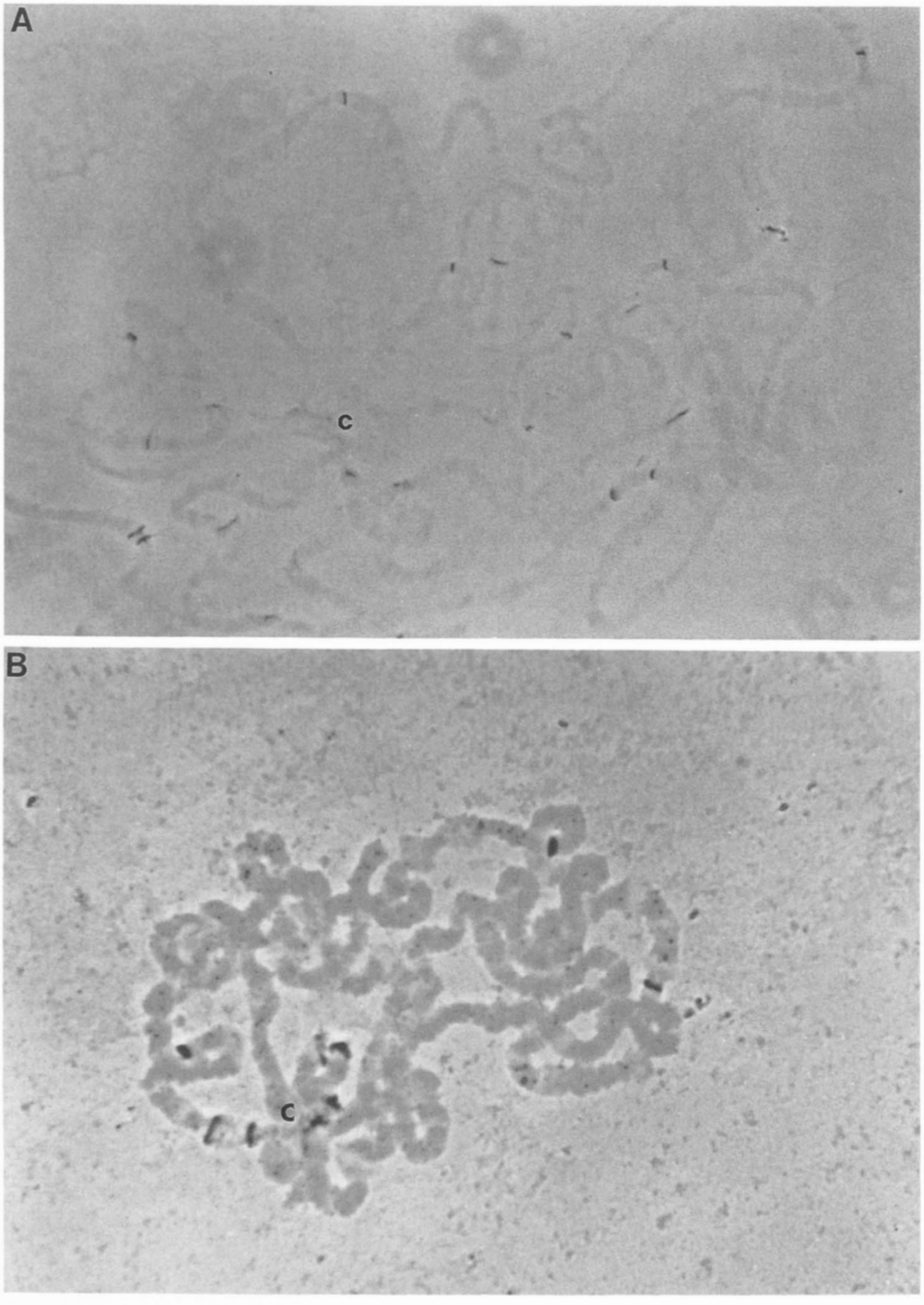

FIGURE 6.—*In situ* hybridizations of polytene chromosomes isolated from 3rd instar larvae representing the (A) Dimonika, Congo population of *D. melanogaster* and the (B) Seychelles population of *D. simulins* using the internal *copia* probe. (c = centromere).

sequestering of *copia* elements within centric heterochromatin is a major contributor to the *copia* transcriptional variation which exists among natural populations of Drosophila.

If a significant component of the transcriptional variation we have observed among populations is due to variable numbers of transcriptionally active *copia* elements, the progeny of crosses made between parents representing various populations would be expected to display *copia* RNA levels which are mid-way between those of the parents. In fact, we found that in 3 out of 4 of the interpopulations crosses examined, *copia* RNA levels of the progeny were significantly different from the expected mid-way value. In most cases, the progeny RNA levels were either more extreme or nonsignificantly different from one or other of the parents. These results suggest that much of the *copia* variation present among natural populations is the result of variation in *trans*-regulatory control(s).

Trans-regulatory effects on *copia* expression could, in theory, be element or host (chromosomal) encoded. There is, as yet, no evidence that between element regulatory control exists among RLEs in either yeast or Drosophila. By contrast, there are numerous examples in both yeast and Drosophila of host loci exerting *trans*-regulatory effects on RLE expression (WINSTON 1988; WINSTON *et al.* 1987; PARKHURST and CORCES 1986; KUBLI 1986). It seems likely, there-

fore, that the *trans*-regulatory variation in *copia* expression which exists among natural populations is due to genetic variation at host loci. Genetic mapping experiments are currently underway to test this hypothesis.

Low levels of *copia* transcripts may be the ancestral condition in *D. melanogaster*: Our results demonstrate that populations of *D. melanogaster* associated with little or no *copia* transcripts are limited in geographic distribution to remote regions of tropical Africa. Since *copia* elements are transcriptionally (STRAND and MCDONALD 1985; BRAUDE-ZOLOTAREJOVA and SCHUPPE 1987) and maybe transpositionally (JUNAKOVIC *et al.* 1986; however, see ARNAULT and BIEMONT 1989) responsive to heat shock, it could be argued that selective pressures might exist to reduce the number of biologically active *copia* elements in tropical regions of the world. However, this hypothesis appears to be inconsistent with our findings that other tropical populations of *D. melanogaster* in South America, the Caribbean, Tahiti and other regions of tropical Africa are associated with moderate to high levels of *copia* transcripts. We believe that the explanation of the low levels of *copia* RNA associated with certain tropical African populations is most likely to be historical in nature.

Based upon several lines of evidence, it is now generally accepted that *D. melanogaster* originated in tropical Africa (DAVID and CAPY 1988). The hypothesis that the presence of little or no *copia* transcripts represents the ancestral state is supported by the fact that *copia* transcripts are in extremely low abundance or absent in *D. simulans* and *D. mauritiana* populations as well. Moreover, those central African populations which do display moderate levels of *copia* transcripts are located in or near urban areas and thus may be expected to have experienced secondary introductions of nonendemic flies, by commercial transport, etc. (VOUIDIBIO *et al.* 1989).

How and why more recently evolved populations of *D. melanogaster* have acquired transcriptionally active *copia* elements is an interesting evolutionary question which cannot be presently resolved.

Naturally occurring variation in RLE expression may be of adaptive evolutionary significance: RLEs are a well recognized source of mutation (LAMBERT, MCDONALD and WEINSTEIN 1988). In Drosophila, for example, the majority of all morphologically detectable spontaneous mutations which have been examined on the molecular level are associated with the insertion of RLEs (FINNEGAN and FAWCETT 1986; SANKARANARAYANAN 1988; MCDONALD 1989).

Since RLE insertion mutants are frequently characterized by the acquisition of novel *trans*-regulatory controls, they may play an important role in the evolution of new regulatory and developmental networks (STAVENHAGEN and ROBINS 1988; MCDONALD 1989, 1990). One possible objection to the view that regulatory and developmental variants are of evolutionary significance is the fact that, in most instances, variant alleles having major phenotypic consequences should be quickly eliminated from populations by natural selection. However, since the mutant phenotypes associated with RLE insertion variants can be partially or completely suppressed by host (chromosomal) loci, this class of regulatory mutants may be shielded from natural selection and thereby be maintained in populations for extended periods of time (MCDONALD 1990). Thus, although RLE insertion variants may represent a significant long-term evolutionary potential, in the short-term they may appear to be selectively neutral or nearly neutral and of little or no evolutionary importance (MONTGOMERY, CHARLESWORTH and LANGLEY 1987; LANGLEY *et al.* 1988).

According to our model, the frequency and pattern of RLE insertion variants expected in populations depends upon the frequency and nature of suppressor alleles present (A. J. CUTICCHIA and J. F. MCDONALD, unpublished results). Although little is currently known about the frequency of suppressor alleles present in populations, the molecular mechanisms which underlie suppression are beginning to be well understood.

Genes which suppress the mutant phenotypes associated with RLE insertion variants are, in fact, *trans*-regulators of RLE expression (KUBLI 1986). Thus, the action of suppressor alleles on mutant gene activity is indirect. In a number of instances, the transcriptional activity of an inserted RLE has been shown to be inversely correlated with the transcriptional activity of the mutant chromosomal allele. For example, a naturally occurring variant allele at the *D. melanogaster* alcohol dehydrogenase (*Adh*) locus has recently been characterized in which a *copia* element has inserted 250 bp upstream of the *Adh* gene (STRAND and MCDONALD 1989). Consistent with a model of transcriptional interference, this insertion variant displays reduced *Adh* expression in tissues and at life-stages where *copia* elements are actively expressed. In tissues and life-stages where *copia* expression is significantly reduced, the variant allele's level of *Adh* expression approaches that of wild type.

According to the transcriptional interference model, chromosomal genes which decrease the transcriptional activity of an inserted RLE are expected to increase the activity of the adjacent variant allele and thereby "suppress" its mutant phenotype (KUBLI 1986). There are a number of documented examples of RLE insertion variants in both yeast and *Drosophila* whose mutant phenotypes are suppressed by this mechanism (*e.g.*, WINSTON *et al.* 1987; PARKHURST and CORCES 1986). Thus, our finding that natural

populations of Drosophila contain *trans*-repressors of *copia* expression implies that natural populations also harbor the potential to phenotypically suppress regulatory variants associated with *copia* and perhaps other RLE insertion variants.

A second possible reason why variation in *copia* expression may be of evolutionary significance is that it may contribute to variation in rates of retrotransposition (McDONALD 1989, 1990). The substantial structural homology that exists between retroviral-like transposable elements and retroviral proviruses suggests that their modes of replication may be analogous as well. Direct evidence that RLEs replicate and transpose via reverse transcription of an RNA intermediate has been acquired for *Ty* elements in yeast (BOEKE *et al.* 1985). Consistent with an RNA-dependent mode of replication, it has been demonstrated that an increased rate of *Ty* transcription results in a concomitant increase in the rate of *Ty* element transpositions (BOEKE *et al.* 1985; McCLANAHAN and McENTEE 1986). If the molecular mechanisms which appear to underlie retrotransposition events in yeast are representative of Drosophila, then naturally occurring variability in *copia* transcript levels may contribute to intraspecific variability in rates of retrotransposition. Whether or not such a mechanism may be an at least partial explanation of long standing reports of variable mutation rates among *Drosophila* populations (*e.g.*, IVES 1950; THOMPSON 1962; DUBININ 1966) is a question which remains to be answered.

The ultimate validity of these and other adaptive evolutionary scenarios involving RLEs will be better judged as more biological information becomes available on the chromosomal genes which control their expression and to what extent these genes are variable in natural populations.

We are grateful to J. DAVID and M. KIDWELL for providing many of the Drosophila strains used in this study. A. J. CUTICCHIA aided in the statistical analysis. WYATT ANDERSON and FRANCOISE LEMEUNIER provided valuable assistance in the *in situ* studies. The continuous help and encouragement of our colleages DENNIS STRAND and SUSAN VOSS is especially appreciated. A.K.C. was supported by a National Institutes of Health traineeship to the Department of Genetics, University of Georgia.

LITERATURE CITED

ANDERSON, P., M. BARBACID, S. H. TRONICK, H. CLARK and S. A. AARONSON, 1979 Evolutionary relatedness of viper and primate endogenous retroviruses. Science **204:** 318–321.

ANDERSON, S. M., and J. F. McDONALD, 1983 Biochemical and molecular analysis of naturally occurring *adh* variants in *D. melanogaster*. Proc. Natl. Acad. Sci. USA **80:** 4798–4802.

ARNAULT, C., and C. BIEMONT, 1989 Heat shocks do not mobilize mobile elements in genomes of *Drosophila melanogaster* inbred lines. J. Mol. Evol. **28:** 388–390.

BALTIMORE, D., 1985 Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell **40:** 481.

BENVENISTE, R., 1985 The contribution of retroviruses to the study of mammalian evolution, pp 359–417, in *Molecular Evolutionary Genetics*, edited by R. MACINTYRE. Plenum Press, New York.

BIALOGAN, S., D. FAULKENBURG and D. RENKAWITZ-POHL, 1985 Characterization and developmental expression of β-*tubulin* genes in *Drosophila melanogaster*. EMBO J. **3:** 2543–2548.

BINGHAM, P., R. LEVIS and G. RUBIN, 1981 Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. Cell **25:** 693–704.

BOEKE, J. D., D. J. GARFINKEL, C. A. STYLES and G. R. FINK, 1985 *Ty* elements transpose through an RNA intermediate. Cell **40:** 491–500.

BRAUDE-ZOLOTARJOVA, T., and N. G. SCHUPPE, 1987 Transient expression of hsp-CAT I and copia-CAT I hybrid genes in *D. melanogaster* and *D. virilis* cultured cells. Drosophila Inform. Ser. **66:** 33.

BUCHETON, A., A. PARO, H. SANG, A. PELISSON and D. FINNEGAN, 1984 The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. Cell **38:** 153–163.

CALLAHAN, R., I. M. CHIU, J. F. WONG, S. H. TRONICK, B. RUE, S. A. AARONSON and J. SCHLON. 1985 A new class of endogenous human retroviral genomes. Science **228:** 1208–1211.

DAVID, J. R., and P. CAPY, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

DOWSETT, A., and M. YOUNG, 1982 Differing levels of dispersed repetitive DNA among closely related species of *Drosophila*. Proc. Natl. Acad. Sci. USA **79:** 4570–4574.

DUBININ, N., 1966 *Evolution of Populations and Radiation.* Atomizdat, Moscow.

DUNSMUIR, P., W. BROREIN, M. SIMON and G. RUBIN, 1980 Insertion of the *Drosophila* transposable element *copia* generates a 5 base pair duplication. Cell **21:** 575–579.

FEINBERG, A. P., and B. VOGELSTEIN, 1984 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137:** 266–267.

FINNEGAN, D. J., 1989 Eukaryotic transposable elements and genome evolution. Trends Genet. **5:** 103–107.

FINNEGAN, D., and D. FAWCETT, 1986 Transposable elements in *Drosophila melanogaster*. Oxf. Surv. Eukaryotic Genes **3:** 1–62.

HOGG, R., and A. T. CRAIG, 1970 *Introduction to Mathematical Statistics.* Macmillian, New York.

IVES, P., 1950 The importance of mutation rate genes in evolution. Evolution **4:** 236–252.

JOHNS, M., J. MOTTINGER and M. FREELING, 1985 A low copy number *copia*-like transposon in maize. EMBO J. **4:** 1093–1099.

JUNAKOVIC, N., C. DIFRANCO, P. BARSANTI and G. PALUMBO, 1986 Transposition of *copia*-like nomadic elements can be induced by heat shock. J. Mol. Evol. **24:** 89–93.

KUBLI, E., 1986 Molecular mechanisms of suppression in *Drosophila*. Trends Genet. **2:** 204–208.

LAMBERT, M. E., J. F. McDONALD and I. B. WEINSTEIN, 1988 *Eukaryotic Transposable Elements as Mutagenic Agents.* Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

LANGER-SAFER, P., M. LEVINE and D. WARD, 1982 Immunological method for mapping genes on *Drosophila* polytene chromosomes. Proc. Natl. Acad. Sci. USA **79:** 4381–4385.

LANGLEY, C. H., E. MONTGOMERY, R. HUDSON, N. KAPLAN and B. CHARLESWORTH, 1988 On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52:** 223–235.

LANSMAN, R., S. STACEY, T. GRIGLIATTI and H. BROCK, 1985 Sequences homologous to the P mobile element of *Drosophila melanogaster* are widely distributed in the genus *Sophophora*. Nature **318:** 561–563.

LEVIS, R., P. DUNSMUIR and G. RUBIN, 1980 Terminal repeats of the *Drosophila* transposable element *copia*: nucleotide sequence and genomic organization. Cell **30:** 551–565.

LEWIN, B., 1987 *Genes.* John Wiley & Sons, New York.

MCCLANAHAN, T., and K. MCENTEE, 1986 DNA damage and heat shock dually regulate genes in *Saccharomyces cerevisiae.* Mol. Cell. Biol. **6:** 90–96.

MCDONALD, J. F., 1989 The potential evolutionary significance of retroviral-like transposable elements in peripheral populations, pp. 190–205 in *Evolutionary Biology of Transient Unstable Populations*, edited by A. FONTDEVILA. Springer-Verlag, Heidelberg.

MCDONALD, J. F., 1990 Macroevolution, gene regulation and retroviral-like elements. Bioscience **40:** 183–191.

MCDONALD, J. F., D. J. STRAND, M. E. LAMBERT and I. B. WEINSTEIN, 1987 Responsive genome: evidence and evolutionary implications, pp 239–263 in *Development as an Evolutionary Process*, edited by R. RAUFF and E. RAUFF. Alan R. Liss Press, New York.

MONTGOMERY, E., B. CHARLESWORTH and C. H. LANGLEY, 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.

PARDUE, M., and J. GALL, 1975 Nucleic acid hybridization to the DNA of cytological preparations. Methods Cell Biol. **10:** 1–16.

PARKHURST, S., and V. G. CORCES, 1986 Interactions among the *Gypsy* transposable element and the *yellow* and the *suppressor-of-hairy-wing* loci in *D. melanogaster.* Mol. Cell. Biol. **6:** 47–52.

RIVEN, C., C. CULLIS and V. WALBOT, 1986 Evolutionary quantitative variation in the genome of *Zea mays.* Genetics **113:** 1009–1019.

ROEDER, G., and G. FINK, 1983 Transposable elements in yeast, pp 300–328 in *Mobile Genetic Elements*, edited by J. SHAPIRO. Academic Press, New York.

SANKARANARAYANAN, K., 1988 Mobile genetic elements, spontaneous mutations and the assessment of genetic radiation hazards in man, pp 319–336, in *Eukaryotic Transposable Elements as Mutagenic Agents*, edited by M. E. LAMBERT, J. F. MCDONALD and I. B. WEINSTEIN. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

SOUTHERN, E., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

STAVENHAGEN, J. B., and D. M. ROBINS, 1988 An ancient provirus has imposed androgen regulation on the adjacent mouse sex-linked protein gene. Cell **55:** 247–254.

STEELE, E. J., 1979 *Somatic Selection and Adaptive Evolution.* University of Chicago Press, Chicago.

STRAND, D., and J. F. MCDONALD, 1985 *Copia* is transcriptionally responsive to environmental stress. Nucleic Acids Res. **13:** 4401–4410.

STRAND, D., and J. MCDONALD, 1989 Insertion of a *copia* element 5′ to the *Drosophila melanogaster alcohol dehydrogenase* gene (Adh) is associated with altered developmental and tissue-specific patterns of expression. Genetics **121:** 707–794.

TEMIN, H., 1982 Viruses, protoviruses, development and evolution. J. Cell. Biochem. **19:** 105–118.

THOMAS, P., 1980 Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77:** 5201–5205.

THOMPSON, P., 1962 Asynapsis and mutability in *D. melanogaster.* Genetics **47:** 337–349.

VOUIDIBIO, J., P. CAPY, D. DEFAYE, E. PLA, J. SANDRIN, A. CSINK and J. R. DAVID, 1989 Short range genetic structure of *Drosophila melanogaster* in an Afrotropical urban area and its significance. Proc. Natl. Acad. Sci. USA **86:** 8442–8446.

WINSTON, F., 1988 Genes that affect *Ty*-mediated gene expression in yeast, pp. 145–153 in *Eukaryotic Transposable Elements as Mutagenic Agents*, edited by M. LAMBERT, J. MCDONALD and I. B. WEINSTEIN. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

WINSTON, F., C. DOLARD, B. MALONE, J. CLAVE, J. KAPAKOS, P. FARABAUGH and P. MINEHART, 1987 Three genes are required for *trans*-activation of *Ty* transcription in yeast. Genetics **115:** 649–656.

YOUNG, M., and H. SCHWARTZ, 1981 Nomadic gene families in *Drosophila.* Cold Spring Harbor Sym. Quant. Biol. **45:** 629–640.

Communicating editor: M. J. SIMMONS