Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 7097–7102, July 1996
Evolution

Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation

(long terminal repeat retrotransposon/transposable element/evolution/gene expression)

LILYA V. MATYUNINA, I. KING JORDAN, AND JOHN F. MCDONALD*

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

Communicated by John C. Avise, University of Georgia, Athens, GA, March 28, 1996 (received for review January 23, 1996)

ABSTRACT **Significant differences in levels of *copia* [*Drosophila* long terminal repeat (LTR) retrotransposon] expression exist among six species representing the *Drosophila melanogaster* species complex (*D. melanogaster*, *Drosophila mauritiana*, *Drosophila simulans*, *Drosophila sechellia*, *Drosophila yakuba*, and *Drosophila erecta*) and a more distantly related species (*Drosophila willistoni*). These differences in expression are correlated with major size variation mapping to putative regulatory regions of the *copia* 5′ LTR and adjacent untranslated leader region (ULR). Sequence analysis indicates that these size variants were derived from a series of regional duplication events. The ability of the *copia* LTR–ULR size variants to drive expression of a bacterial chloramphenicol acetyltransferase reporter gene was tested in each of the seven species. The results indicate that both element-encoded (cis) and host–genome-encoded (trans) genetic differences are responsible for the variability in *copia* expression within and between *Drosophila* species. This finding indicates that models purporting to explain the dynamics and distribution of retrotransposons in natural populations must consider the potential impact of both element-encoded and host–genome-encoded regulatory variation to be valid. We propose that interelement selection among retrotransposons may provide a molecular drive mechanism for the evolution of eukaryotic enhancers which can be subsequently distributed throughout the genome by retrotransposition.**

Transposable elements (TEs) are a primary source of spontaneous mutations that cause major phenotypic effects (1, 2) and are hypothesized to be of evolutionary significance (3–6). Consistent with this hypothesis, recent analyses have demonstrated that TEs are responsible for a number of fixed differences among species in chromosome structure and patterns of gene expression (7–15). Despite the growing body of evidence that TEs are of evolutionary importance, relatively little is currently known about the biological mechanisms that may contribute to the establishment and maintenance of TEs in natural populations and species. One important question yet to be answered is the extent to which intra- and interspecific variability in TE activity (expression/transposition rates) is due to element-encoded (cis) versus host-encoded (trans) variation. Until this question is satisfactorily answered, an accurate understanding of the evolutionary dynamics of TEs in natural populations and species may be impossible (16). Another issue that requires knowledge of the magnitude and extent of element–host interactions in order to be fully resolved is the horizontal transfer of TEs between species (17, 18). If a TE is vectored into a new species in a horizontal fashion, its ultimate establishment in the new genome will depend upon the extent to which its ability to replicate is regulated by the host (19).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

For the past several years, our laboratory has been engaged in an analysis of naturally occurring structural and functional variation in the *Drosophila* long terminal repeat (LTR) retrotransposon *copia*. We previously reported that *copia* transcript levels in adults vary widely among natural populations of *Drosophila melanogaster*, and that this variation may be due to both element encoded and host encoded genetic variation (20, 21). We also reported that *copia* transcripts are not detectable in adults of the *Drosophila* sibling species *Drosophila simulans* and *Drosophila mauritiana*, and that this lack of activity is apparently not attributable to the confinement of *copia* elements in constitutive heterochromatin.

In this paper, we extend our interspecific survey of *copia* RNA levels to the larval stage of development and show that among six species representing the *D. melanogaster* species complex (*D. melanogaster*, *D. mauritiana*, *D. simulans*, *Drosophila sechellia*, *Drosophila yakuba*, and *Drosophila erecta*) and a more distantly related species (*Drosophila willistoni*), *copia* expression is detectable in only *D. melanogaster*. In addition, we report that three major size variants mapping to putative regulatory regions of the *copia* 5′ LTR and adjacent untranslated leader region (ULR) are distributed in variable frequencies among all of the seven species examined in this study. Sequence analysis suggests that these size variants derived from a series of regional duplication events. The ability of these *copia* LTR–ULR variants to drive expression of a bacterial chloramphenicol acetyltransferase (CAT) reporter gene was tested in each of the seven species. The results indicate that both element-encoded (cis) and host–genome-encoded (trans) genetic differences are responsible for the variability in *copia* expression within and between *Drosophila* species.

MATERIALS AND METHODS

***Drosophila* Strains.** The *D. melanogaster* Iquitos (Peru) and Loua (Congo) isofemale strains have been described (20). Strains of the other *Drosophila* species used in the study were obtained from either the National *Drosophila* Stock Center (Bloomington, IN) (*D. sechellia*, no. 3590; *D. erecta*, no. 1013) or the National *Drosophila* Species Stock Center (Bowling Green, OH) (*D. simulans*, no. 0251.4; *D. mauritiana*, no. 02410; *D. yakuba*, no. 0261.0; *D. willistoni*, no. 0811.2).

RNA Preparation and Northern Analyses. Total RNA was isolated from 1st and 2nd instar larvae by using the RNAeasy Kit (Qiagen, Chatsworth, CA). RNA was electrophoresed through 1% agarose-formaldehyde gels, transferred to nitrocellulose filters, and hybridized as described (20). The *copia* genomic clone Dm5002 was used as probe (22). The integrity

Abbreviations: TE, transposable element; LTR, long terminal repeat; ULR, untranslated leader sequence; CAT, chloramphenicol acetyltransferase.

Data deposition: The sequence reported in this paper has been deposited in the GenBank data base [accession nos. U60291 (double gap), U60292 (leader gap), and U60293 (full length)].

*To whom reprint requests should be addressed.

of the RNA samples was determined by rehybridization of filters with a *Drosophia* β-tubulin probe (23). Autoradiographs were densitometrically scanned on a Beckman DU-8 spectrophotometer using the GELSCAN program.

Restriction and Sequence Analyses. Genomic DNA was prepared from adult flies representing the eight strains examined in this study as described (24). The primers shown in Fig. 1 were used to PCR amplify (25) *copia* LTR–ULRs from each of the eight DNA samples under the following conditions: 1× PCR buffer (Fisher), 200 μM dNTPs (Pharmacia), 3 mM $MgCl_2$, 0.5 μM of each primer, 2.5 units *Taq* polymerase (Fisher), and 100 ng of genomic DNA. Thermal cycle conditions were as follows: 5 min at 94°C followed by 30 cycles of 1 min at 94°C, 1 min at 50°C, and 1 min at 72°C followed by a final 5 min at 72°C. The resulting eight pools of *copia* LTR–ULRs were subcloned into pCRII by using the TA Cloning Kit (Invitrogen). Individual subclones containing LTR–ULR variants present in each of the eight genomes surveyed were PCR amplified as described above, and the *copia* LTR–ULR products were digested with *Apa*I (Promega), which cuts just 3′ to the end of the *copia* LTR (Figs. 1 and 2). The resulting two LTR- and ULR-containing fragments were fractionated through 2% agarose (Sigma) gels. Variability in fragment sizes indicated the presence or absence of gaps in the LTR and/or ULR (Fig. 2). Sequence analysis was subsequently performed to determine the position of gaps in these naturally occurring *copia* LTR–ULR size variants. Automated sequencing was done at the University of Georgia Molecular Genetics Instrumentation Facility. *Copia* LTR–ULR subclones were sequenced in both directions by using T7 and SP6 primers, respectively. The resulting chromatograms were aligned and compared by using the SEQED program (Applied Biosystems) to resolve ambiguities. Multiple sequences were aligned using the Pileup program of the Wisconsin Genetics Computer Group computer package.

Reporter Plasmids Construction. Using the primers shown in Fig. 1, *copia* LTR–ULRs were PCR amplified from two previously characterized *D. melanogaster copia* clones (Iquitos, Peru-IP2; Loua, Congo-LC9) and a *D. simulans* (SIM-5) clone (21). These *copia* clones were selected because they were determined to contained the three major *copia* LTR–ULR size variants segregating in *Drosophila* species. The *D. simulans* element contains a gap in the LTR and the ULR (double-gap variant); the Loua, Congo element contains the same ULR gap but no gap in the LTR (ULR-gap variant); the Iquitos, Peru element contains no gaps in the LTR or ULR (full-length variant). The three distinct PCR-amplified *copia* LTR–ULR size variants were separately cloned into the *Sal*I–*Xba*I sites of the pCAT basic vector (Promega). A plasmid lacking the entire leader region (LTR–no ULR) was constructed from the full-length Iquitos LTR–leader–CAT plasmid by removing the *Apa*I–*Xba*I fragment. The *hsp70*–CAT plasmid was constructed by placing the CAT gene from the pCAT basic vector (*Xba*I–*Bam*HI fragment) under the control of the *hsp70* promoter contained in the pCaSpeR-hs plasmid (26). The basic pCAT vector, which lacks any promoter, was used as our negative control plasmid (p0CAT).

Embryo Injections and CAT Assays. Reporter plasmid DNA was isolated by using the Plasmid Mini Kit (Qiagen) and

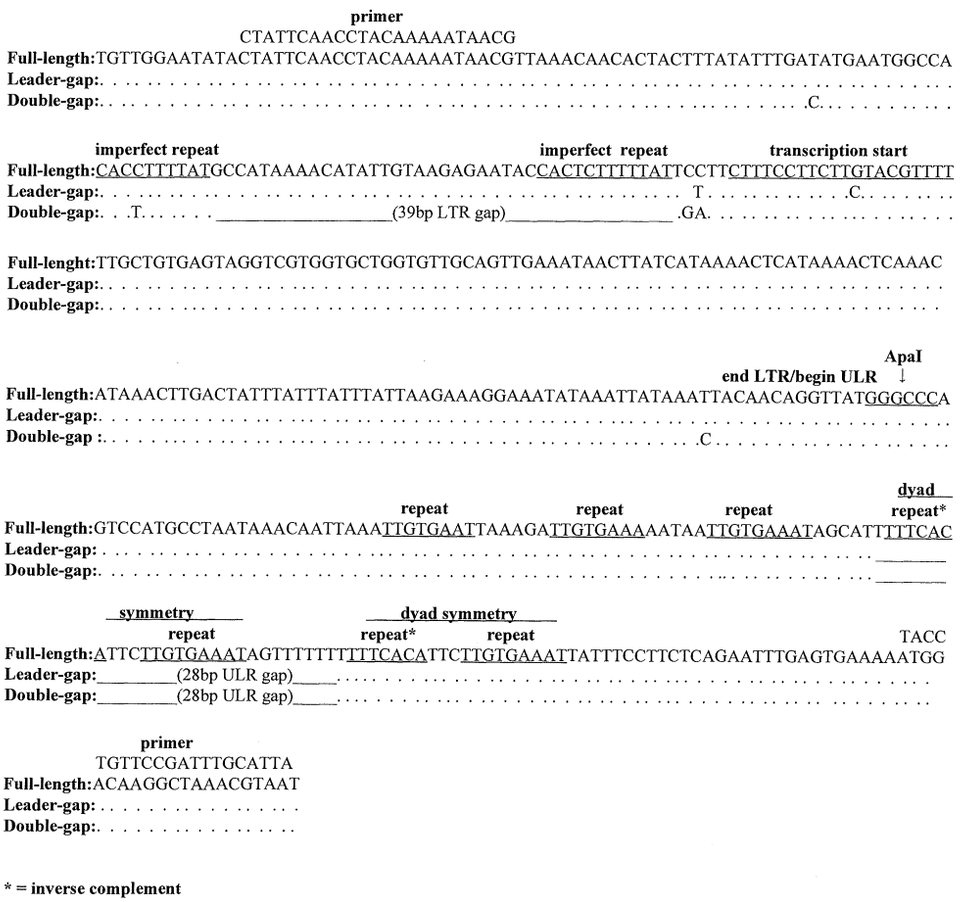

FIG. 1. Sequences of the LTR–ULR region of three *copia* elements representing the major size variants segregating in natural populations of *Drosophila* that were fused to the bacterial CAT reporter gene for transient expression assays reported in this paper. The primers used in PCR amplification of the region are shown. Also indicated are imperfect repeats and the putative transcriptional start site located within the LTR, the LTR–ULR boundry, the adjacent *Apa*I site, and the seven repeating sequence motifs located within the ULR.

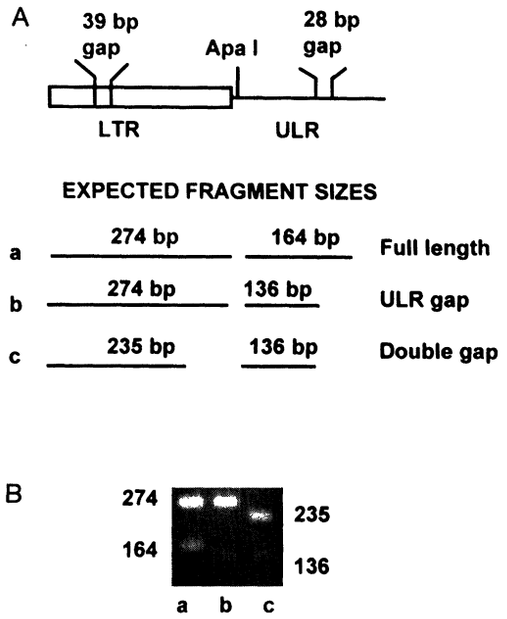

FIG. 2. (*A*) Representation of *copia* LTR–ULR showing the position of major gaps in naturally occurring elements. Also shown are the expected fragment sizes of PCR amplified LTR–ULR size variants after digestion with *Apa*I. (*B*) Restriction pattern resulting from an *Apa*I digestion of PCR amplified LTR–ULRs from naturally occurring *Drosophila* species. Lanes: a, full-length variant from *D. melanogaster* (Iquitos, Peru); b, ULR gap variant from *D. simulans*; c, double gap variant from *D. mauritiana*. The sequences of these variants are shown in Fig. 1.

adjusted to a concentration of 300 μg/ml in injection buffer (5 mM KCl/0.1 mM phosphate buffer, pH 6.8). Dechorionated 1-h embryos were injected with the plasmid DNA under oil as described (27). The DNA sample (4.6 nl) was injected per embryo by using a positive displacement electronic nanoliter injector (World Precision Instruments, Sarasota, FL). After injection, slides containing the embryos were placed in a humidity chamber at 18°C. Active 1st and 2nd instar larvae were harvested 48 hr after injection and individually homogenized, and the supernatant assayed for CAT activity by liquid scintillation counting of CAT reaction products by using the CAT Enzyme Assay System (Promega). The embryos injected with the *hsp70*–CAT plasmid were treated as described above except that at 24 hr after injection they were shifted from 18°C to 37°C for a period of 30 min and then returned to 18°C until harvested at 48 hr postinjection as described above. CAT activity was independently assayed for 10 larvae (one assay per larvae) representing each of the eight *Drosophila* strains for three *copia* LTR–ULR–CAT plasmids and the *hsp70*–CAT (positive control) plasmid. Five to 10 larvae per strain were assayed for the p0CAT (negative control) plasmid. Means and standard deviations were computed for each plasmid or strain. Two-tailed *t* tests were used to determine whether or not the average differences between individual plasmid or strains were of significance. ANOVAs were carried out to determine the overall significance of between strain variation for each reporter plasmid tested.

RESULTS

Larval *copia* RNA Levels Are Variable Among *Drosophila* Species. Csink and McDonald (20) previously reported that steady-state levels of *copia* RNA in *D. melanogaster* adults vary significantly among natural populations. In addition, it was reported that no *copia* transcripts are detectable in *D. simulans* or *D. mauritiana* adults. The results of our survey of *copia* RNA levels in 1st and 2nd instar larvae are consistent with these earlier findings (Fig. 3). Although *copia* transcripts are present in larvae of both the Iquitos and Loua *D. melanogaster* strains, levels are >30-fold higher in the former than the latter. We detect no *copia* transcripts in *D. simulans* or *D. mauritiana* larvae or in any of the other tested members of the *D. melanogaster* species group (*D. sechellia, D. yakuba*, and *D. erecta*). Likewise, no *copia* transcripts were detected in *D. willistoni* larvae.

Size Variation Mapping to the LTR and ULR of Naturally Occurring *copia* Elements Was Likely Generated by Duplication. Csink and McDonald (21) previously reported that *D. melanogaster copia* elements differ from those of *D. simulans* and *D. mauritiana* by size variation mapping to the 5′ LTR and ULR. Relative to the Iquitos *D. melanogaster copia* elements, all of the *D. simulans* and *D. mauritiana* elements sequenced in this previous study were found to have a 39-bp gap within the 5′ LTR located just 5′ to the transcriptional start site. In addition, all nine of the *D. mauritiana* elements and two of three of the *D. simulans copia* elements sequenced were found to have a 28-bp gap within the 5′ ULR. This 28-bp gap in the leader sequence was also found in two out of nine of the *copia* elements isolated from the Loua *D. melanogaster* population and was observed to encompass a region of dyad symetry homologous to the core sequence of the simian virus 40 enhancer.

We find that one or more of these same *copia* LTR–leader size variants are present within the genomes of all eight of the strains analyzed in this study. Genomic DNA isolated from each strain was PCR amplified by using the primers shown in Fig. 1. The PCR products were subcloned and the *copia* LTR–ULR variants characterized by restriction mapping and sequence analysis (Figs. 1 and 2). A summary of our results is presented in Table 1. The detailed results of the sequence analysis of these and other strains will be published elsewhere (I.K.J. and J.F.M., in preparation).

Consistent with the previous findings of Csink and McDonald (21), we find that the full-length LTR–ULR variant (no gaps) is the only class of *copia* element present within the *D. melanogaster* Iquitos strain. This full-length variant is one of three LTR–ULR size variants found within the *D. simulans*, *D. erecta*, and *D. willistoni* strains. All *copia* elements present within the *D. mauritiana* strain have both the 39-bp LTR gap and the 28-bp leader region gap. The double-gap and the ULR-gap are the only variants found within the *D. sechellia* strain, whereas the Loua and *D. yakuba* strains have only the full-length and ULR-gap variants (Table 1).

Comparison of the sequences of the gap and full-length variants suggests that the latter originally arose from the former by regional duplications. For example, the region of dyad symetry within the ULR that is homologous to the simian virus 40 enhancer is repeated twice in the full-length variant

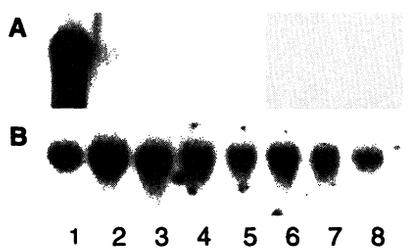

FIG. 3. (*A*) Northern blot of total RNA isolated from 1st and 2nd instar larvae representing various *Drosophila* species and geographic populations. Blots were hybridized against a full-length *copia* probe (Dm5002) (22). Lanes: 1, *D. melanogaster* (Iquitos, Peru); 2, *D. melanogaster* (Loua, Congo); 3, *D. simulans*; 4, *D. mauritiana*; 5, *D. erecta*; 6, *D. yakuba*; 7, *D. sechellia*; 8, *D. willistoni*. (*B*) Same Northern blot as in *A*, which has been rehybridized against a *Drosophila* β-tubulin probe (23).

Table 1. Distribution of *copia* LTR–ULR size variants among the eight *Drosophila* strains examined in this study

Strain	*Copia* variants		
	Full-length	ULR-gap	double-gap
D. melanogaster (Iquitos)	+	−	−
D. melanogaster (Loua)	+	+	−
D. mauritiana	−	−	+
D. simulans	+	+	+
D. sechellia	−	+	+
D. yakuba	+	+	−
D. erecta	+	+	+
D. willistoni	+	+	+

+, Presence of variant; −, absence of variant.

but only once in the ULR-gap variant (Fig. 1). These repeated motifs are bordered by runs of T's, which may have facilitated regional mispairing between these homologous regions, leading to unequal crossing over and consequent duplication. Homologous recombination events are known to occur between the two RNA strands packaged within retroviral particles (28). Similar opportunities for both equal and unequal recombination events may also exist within LTR retrotransposon particles. Sequences immediately adjacent to the region missing in the LTR of the double-gap variant are duplicated within the LTR of the full-length variant. This suggests that the full-length variant might have also derived from a gap variant by regional duplication.

Naturally Occurring *copia* LTR–ULR Size Variation Is of Functional Significance. The variation in levels of *copia* expression that exists within and between *Drosophila* species may be caused by genetic variation between *copia* elements and/or between host chromosomal genes that transregulate *copia* expression. To determine if the naturally occurring *copia* LTR–ULR region size variation may be of functional significance, we tested the relative ability of the full-length, the double-gap, and the ULR-gap variants to drive expression of a bacterial CAT reporter gene in Iquitos *D. melanogaster* larvae. The results presented in Table 2 indicate that: (*i*) all three of the naturally occurring variants (full-length, ULR-gap, and double-gap) are able to drive CAT expression at levels significantly higher than the negative control in all strains tested ($P < 0.0025$); (*ii*) the full-length variant is able to drive CAT expression at levels significantly higher than either the ULR-gap (*t* test, $P < 0.001$) or the double-gap ($P < 0.001$) variants in all strains tested; and (*iii*) the level of CAT activity associated with the ULR-gap variant is significantly higher than that associated with the double-gap variants in the Iquitos and *D. simulans* strains ($P < 0.0001$). From these results we conclude that the *copia* LTR–ULR size variation that is segregating in *Drosophila* populations and species is of functional significance.

To further investigate the role of the ULR in *copia* expression, we tested the activity of a laboratory-constructed *copia* LTR–CAT variant in which the entire ULR was deleted. The results (Table 2) demonstrate that removal of the entire ULR results in levels of expression not significantly different at the 95% confidence level from those associated with the naturally occurring ULR-gap variant in five out of the seven strains tested (Iquitos, Loua, *D. mauritiana*, *D. yakuba*, and *D. erecta*). The level of CAT activity associated with the ULR-gap variant was 33% higher ($P < 0.0001$) than that associated with the LTR (no ULR) construct in the *D. simulans* strain, whereas the opposite relationship (39% lower) was found to hold in the *D. sechellia* strain.

***copia* LTR–ULR CAT Expression Is Variable Among *Drosophila* Populations and Species.** The above results indicate that regulatory sequences that influence *copia* expression are located within the element's 5′ LTR and adjacent ULR. Consistent with this finding, putative enhancer sequences localized to the *copia* ULR have been shown to bind *Drosophila* nuclear proteins indicating that this region may be important for host–element regulation (refs. 29 and 30; S. Wilson and J.F.M., unpublished data). To determine whether transregulatory variation may be contributing to interspecific differences in levels of *copia* expression, we tested the ability of the three *copia* LTR–ULR variants described above to drive expression of a bacterial CAT reporter gene in the eight strains previously described.

Analysis of the results presented in Table 2 demonstrate that the ability of both the full-length (ANOVA, $F = 16.56$, $P < 0.0001$) and the ULR-gap (ANOVA, $F = 51.12$, $P < 0.0001$) variants to drive CAT expression is significantly different among all strains tested. The full-length variant is most active in the Iquitos *D. melanogaster* strain. The activity of this same construct is reduced by more than 40% in the Loua *D. melanogaster* strain and by ≈70% in strains representing other members of the *D. melanogaster* species group (*D. simulans*, *D. sechellia*, *D. yakuba*, and *D. erecta*). The *D. willistoni* strain is able to support expression of the full-length variant at only 14% of the level at which it is expressed in the Iquitos *D. melanogaster* strain. In contrast, no significant differences in expression between strains was associated with the double-gap variant.

As is the case with the full-length variant, the ULR-gap variant is expressed at highest levels in the Iquitos strain. Levels of ULR-gap driven expression were significantly lower in all of the other strains tested ($P < 0.001$).

The pattern of expression supported by the *copia* LTR alone (no ULR) is similar to that of the ULR-gap variant in that expression in the Iquitos strain is significantly higher than all

Table 2. Units of CAT activity for various *copia*–CAT and *hsp70*–CAT reporter constructs expressed in 1st and 2nd instar larvae of the *Drosophila* species studied

Species/strains	Full-length	ULR-gap	Double-gap	LTR (no ULR)	*hsp70*	p0CAT (negative control)
D. melanogaster (Iquitos)	1.30 ± 0.55	0.43 ± 0.06	0.22 ± 0.06	0.37 ± 0.20	11.36 ± 0.52	0.10 ± 0.01
D. melanogaster (Loua)	0.73 ± 0.26	0.16 ± 0.02	0.18 ± 0.01	0.19 ± 0.05	11.17 ± 0.15	0.08 ± 0.01
D. mauritiana	0.46 ± 0.15	0.16 ± 0.03	0.19 ± 0.02	0.17 ± 0.03	11.22 ± 0.55	0.07 ± 0.01
D. simulans	0.49 ± 0.17	0.27 ± 0.02	0.17 ± 0.01	0.20 ± 0.02	11.15 ± 0.55	0.11 ± 0.01
D. sechellia	0.44 ± 0.11	0.19 ± 0.03	0.19 ± 0.01	0.26 ± 0.02	12.07 ± 0.36	0.10 ± 0.01
D. yakuba	0.43 ± 0.16	0.17 ± 0.03	0.16 ± 0.01	0.20 ± 0.02	11.45 ± 0.32	0.10 ± 0.01
D. erecta	0.43 ± 0.12	0.21 ± 0.07	0.18 ± 0.03	0.19 ± 0.01	12.05 ± 0.16	0.11 ± 0.01
D. willistoni	0.18 ± 0.01	ND	ND	ND	10.66 ± 0.23	0.08 ± 0.01

Data are expressed as mean ± SD of CAT activity (1 unit = 1 nmol of acetyl groups from acetyl coenzyme A to chloramphenicol in 1 min at 37°C per larvae). The four *copia*-CAT constructs include full-length, ULR-gap, double-gap, and LTR-no ULR constructs. ND, not determined.

the other strains tested ($P < 0.0001$). In contrast, the double-gap variant was associated with a consistently low level of expression in all of the strains tested.

Collectively, our results suggest that host-encoded *copia* trans-regulatory variation exists both within and between *Drosophila* species and that at least some of the sequences which mediate these trans-regulatory effects are located within regions of the *copia* LTR–ULR that are variable in natural populations.

***hsp70*–CAT Expression Is Uniformly High Among All the *Drosophila* Strains Tested.** The variation among *Drosophila* species to support *copia* LTR–ULR expression may be the result of regulatory differences specific for *copia* or the result of nonspecific regulatory differences between species. To address this question, we tested the ability of the *D. melanogaster hsp70* promoter to drive CAT expression in the same strains in which the *copia* LTR–ULR–CAT constructs were tested. The results presented in Table 2 demonstrate that *hsp70*–CAT activity is uniformly high in all strains. From these data we conclude that the differences in *copia* LTR–ULR expression detected among the eight strains examined in this study are neither artifactual nor the result of nonspecific regulatory variation.

DISCUSSION

Retrotransposons are dependent upon host transcriptional and processing functions for execution of their replication and transposition cycles (31). For this reason, host–element interactions have been postulated to play particularly important roles in retrotransposon evolution (3–6). Transcriptional initiation is a rate-limiting step in the transposition of LTR retrotransposons and involves interactions between host-encoded RNA polymerase and associated trans-regulatory proteins with complementary cis-regulatory protein binding sites located within the retrotransposon's 5′ LTR and adjacent ULR (32–35). In this study, we utilized the *Drosophila copia* LTR retrotransposon to investigate the extent to which naturally occurring variability in retrotransposon expression is attributable to host-encoded versus element-encoded genetic variation.

Our results indicate that both element-encoded and host-encoded genetic variation contribute to the differences in levels of *copia* expression that exist within and between *Drosophila* species. For example, the *D. mauritiana* strain contains only the double-gap variant, whereas the *D. sechellia* strain contains the double-gap as well as the ULR-gap variant. Our expression assays indicate that both of these gap variants are cis-defective and cannot support *copia* expression in these species. Although the transcriptionally defective double-gap and ULR-gap variants are present in the other strains in which *copia* transcripts could not be detected (*D. simulans*, *D. yakuba*, *D. erecta*, and *D. willistoni*), the fact that these strains also carry the potentially functional full-length variant suggests that additional trans-acting factors affecting *copia* expression may be at work. Our expression assays are consistent with this prediction in that they demonstrate that the ability of the full-length variant to drive reporter gene expression is consistently low in all the species-strains tested except *D. melanogaster*. This finding suggests that host-encoded regulatory effects may be contributing to the low levels or lack of *copia* transcripts in the majority of strains examined in this study. It should also be noted that chromosome-mediated regulatory influences, such as position or chromatin associated effects, go undetected in transient expression assays (36). Indeed, it has recently been shown that regulatory sequences within the ULR of the *Drosophila gypsy* LTR retrotransposon can have a significant effect on chromatin structure (37). Whether or not regulatory sequences within the *copia* ULR may have a similar effect remains to be determined. However, the fact that we observe (extra-chromosomal) *copia* LTR–ULR-driven CAT expression in strains where we are unable to detect *copia* transcripts suggests that chromatin-mediated regulatory differences may be an additional contributing factor to naturally occurring variation in *copia* expression.

It has long been speculated that interelement selection may be a significant factor in shaping the character and abundance of TEs within genomes (38, 39). Previous studies conducted with yeast LTR retrotransposons have demonstrated that the relative transcriptional activity of these elements is proportional to their relative transposition frequency (40). In so far as this same relationship holds for *Drosophila* LTR retrotransposons, our results indicate that the full-length variant should be selectively favored over the double-gap and ULR-gap variants present within *Drosophila* genomes. However, the potential selective advantage associated with the full-length variant can only be realized if the competing variants are present within a genome which supports *copia* expression. Consistent with the interelement selection model, the full-length variant is fixed within the *D. melanogaster* Iquitos strain, which our expression assays demonstrate is supportive of *copia* expression. The frequency and function of *copia* variants will need to be characterized in more *Drosophila* genomes before the interelement selection model can be fully evaluated. Nevertheless, our present results clearly indicate that any model purporting to explain the dynamics and distribution of retrotransposons in natural populations must consider the potential impact of both element-encoded and host–genome-encoded regulatory variation in order to be valid.

Our results demonstrate that sequences critical to *copia* expression are contained within the element's 5′ ULR. Consistent with earlier expression assays conducted in *D. melanogaster* cell lines (30), we find that deletion of the entire untranslated leader region results in a dramatic reduction in *copia* expression (e.g., ≈70% reduction within the Iquitos, Peru genetic background). Interestingly, a similar low level of expression is associated with the naturally occurring ULR-gap variant, in which only a 28-bp region within the ULR is missing. One interpretation of this result is that all of the sequences critical to the ULR regulatory effect are localized within the 28-bp region missing in the ULR-gap variant. However, two observations argue against this interpretation. First, previous expression assays conducted in *D. melanogaster* and *Drosophila hydei* cell lines have identified putative regulatory sequences within the *copia* ULR that lie outside the 28-bp region missing in the ULR-gap variant (29, 30). Second, the putative regulatory sequences that lie within the region missing in the ULR-gap variant are not unique to this region but are repeated elsewhere in the ULR. Indeed, it is for this reason that we consider it likely that the full-length variant originally derived from the gap variants by a series of regional duplications (see *Results*).

It has been previously noted that the *copia* ULR contains seven copies of a conserved 9-bp motif [(T)TGTGAA(A)(T)] dispersed throughout the ULR (21). Two (inverted repeat) copies of this motif are contained within the region that is missing in the ULR-gap variant. Short repeating sequence motifs are a common feature of eukaryotic enhancer regions, and the relative regulatory strength of enhancers has often been found to correlate with the number of repeating motifs they contain (41–43). Thus, a second possible explanation of the low level of expression associated with the ULR-gap variant is that it contains two fewer copies of the repeating motif than does the full-length variant. According to this hypothesis, deletion of any two copies of the repeating motif within the ULR might be expected to have a similar negative effect on *copia* expression. Experiments are currently under way to test this hypothesis. However, it is interesting at this point to speculate that regional duplications within the ULRs of *copia* and other LTR retrotransposons may be favored by interelement selection and thereby provide a molecular drive mechanism for the evolution of enhancers that may be subsequently distributed throughout the genome by transposition (4, 5).

We thank Dr. Daniel Promislow (University of Georgia) for help with the statistical analysis of our data. This work was supported by National Science Foundation Grant DEB-9307220 to J.F.M.

1. Lambert, M. E., McDonald, J. F. & Weinstein, I. B., eds. (1988) *Eukaryotic Transposable Elements as Mutagenic Agents* (Cold Spring Harbor Lab. Press, Plainview, NY).
2. Berg, D. & Howe, M., eds. (1989) *Mobile DNA* (Am. Soc. for Microbiol., Washington, DC).
3. McDonald. J. F., ed. (1993) *Transposable Elements and Evolution* (Kluwer, Dordrecht, The Netherlands).
4. McDonald, J. F. (1995) *Trends Ecol. Evol.* **10,** 123–126.
5. McDonald, J. F. (1990) *BioScience* **40,** 183–191.
6. McDonald, J. F. (1993) *Curr. Opin. Genet. Dev.* **3,** 855–864.
7. Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S. & Marchetti, E. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3804–3808.
8. Cirera, S., Martin-Campos, J. M., Segarra, C. & Aguade, M. (1995) *Genetics* **139,** 321–326.
9. White, S. E., Habera, L. F. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11792–11796.
10. Flavell, A. J., Pearce, S. R. & Kumar, A. (1994) *Curr. Opin. Genet. Dev.* **4,** 838–844.
11. Mattick, J. S. (1994) *Curr. Opin. Genet. Dev.* **4.** 823–831.
12. Robins, D. M. & Samuelson, L. C. (1993) in *Transposable Elements and Evolution,* ed. McDonald, J. F. (Kluwer, Dordrecht, The Netherlands), pp. 5–15.
13. Lyttle, T. W. & Haymer, D. S. (1993) in *Transposable Elements and Evolution*, ed. McDonald, J. F. (Kluwer, Dordrecht, The Netherlands), pp. 201–214.
14. Purugganan, M. (1993) *Trends Ecol. Evol.* **8,** 239–243.
15. Mason, J. M. & Biessman, H. (1995) *Trends Genet.* **11,** 58–62.
16. Biemont, C. (1993) in *Transposable Elements and Evolution*, ed. McDonald, J. F. (Kluwer, Dordrecht, The Netherlands), pp. 74–91.
17. Clark, J. B., W. P. Maddison & Kidwell, M. G. (1994) *Mol. Biol. Evol.* **11,** 40–50.
18. Robertson, H. M. & Lampe, D. J. (1995) *Annu. Rev. Entomol.* **40,** 333–357.
19. Kidwell, M. G. (1994) *J. Hered.* **85,** 339–346.
20. Csink, A. K. & McDonald, J. F. (1990) *Genetics* **126,** 375–385.
21. Csink, A. K. & McDonald, J. F. (1995) *Mol. Biol. Evol.* **12,** 83–93.
22. Levis, R., Dunsmuir, P. & Rubin, G. M. (1980) *Cell* **21,** 581–588.
23. Bialogan, S., Faulkenburg, D. & Renkawitz-Pohl, D. (1985) *EMBO J.* **3,** 2543–2548.
24. Bender. W., Spirrer, D. & Hogness, D. S. (1983) *J. Mol. Biol.* **168**, 17–33.
25. Innis, M. A. & Gelfand, D. H. (1990) in *PCR Protocols*, eds. Innis, M. A, Gelfand, D. H., Sninsky, J. J. & White, T. J. (Academic, New York), pp. 3–12.
26. Thummel, C. S. & Pirrotta, V. (1992) *Drosophila Inf. Ser.* **71,** 150.
27. Spradling, A. C. & Rubin, G. M. (1982) *Science* **218,** 341–347.
28. Temin, H. (1991) *Trends Genet.* **7,** 71–74.
29. Cavarec, L. & Heidmann, T. (1993) *Nucleic Acids Res.* **21,** 5041–5049.
30. Cavarec, L., Jensen, S. & Heidmann, T. (1994) *Biochem. Biophys. Res. Commun.* **203,** 393–399.
31. Boeke, J. D. & Corces, V. G. (1989) *Annu. Rev. Microbiol.* **43,** 403–434.
32. Sneddon, A. & Flavell, A. J. (1989) *Nucleic Acids Res.* **17,** 4025–4035.
33. Mazo, A. M., Mizrokhi, L. J., Karavanov, A. A., Sedkov, Y. A., Krichevskaja, A. A. & Ilyin, Y. V. (1989) *EMBO J.* **8,** 903–911.
34. Bryant, L. A., Brierly, C., Flavell, A. J. & Sinclair, J. H. (1991) *Nucleic Acids Res.* **19,** 5533- 5536.
35. Corces, V. G. & Geyer, P. K. (1991) *Trends Genet.* **7,** 86–90.
36. Pennie, W. D., Hager, G. L. & Smith, C. L. (1995) *Mol. Cell. Biol.* **15,** 2125–2134.
37. Gerasimova, T. I., Gdula, D. A., Gerasimova, D. V, Simonova, O. & Corces, V. G. (1995) *Cell* **82,** 587–597.
38. Orgel, L. E. & Crick. F. H. C. (1980) *Nature (London)* **284,** 604–607.
39. Doolittle, W. F. & Sapienza, C. (1980) *Nature (London)* **284,** 601–603.
40. Boeke, J. D., Garfinkle, D. J., Styles, C. A. & Fink, G. R. (1985) *Cell* **40,** 491–500.
41. Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987) *Science* **236,** 1237–1244.
42. Muller, M. M., Gerster, T. & Schaffner, W. (1988) *Eur. J. Biochem.* **176,** 485–495.
43. Tjian, R. & Maniatis, T. (1994) *Cell* **77,** 5–8.