Genetica **100:** 3–13, 1997.

LTR retrotransposons and the evolution of eukaryotic enhancers

John F. McDonald, Lilya V. Matyunina, Susanne Wilson, I. King Jordan, Nathan J. Bowen & Wolfgang J. Miller
Department of Genetics, University of Georgia, Athens, GA 30602, USA

Accepted 22 April 1997

Key words: long terminal repeat retrotransposon, transposable element, enhancer, gene expression, *copia/Drosophila*

Abstract

Since LTR retrotransposons and retroviruses are especially prone to regional duplications and recombination events, these viral-like systems may be especially conducive to the evolution of closely spaced combinatorial regulatory motifs. Using the *Drosophila copia* LTR retrotransposon as a model, we show that a regulatory region contained within the element's untranslated leader region (ULR) consists of multiple copies of an 8 bp motif (TTGTGAAA) with similarity to the core sequence of the SV40 enhancer. Naturally occurring variation in the number of these motifs is correlated with the enhancer strength of the ULR. Our results indicate that inter-element selection may favor the evolution of more active enhancers within permissive genetic backgrounds. We propose that LTR retroelements and perhaps other retrotransposons constitute drive mechanisms for the evolution of eukaryotic enhancers which can be subsequently distributed throughout host genomes to play a role in regulatory evolution.

Introduction

Retrotransposons are the most abundant and widely distributed class of eukaryotic transposable elements (Berg & Howe, 1989). These elements are distinguished from other transposable elements by the fact that their replication involves an RNA intermediate and is reverse transcriptase dependent. Retrotransposon insertions adjacent to chromosomal genes frequently result in altered regulatory phenotypes. The molecular mechanisms that underlie these retrotransposon mediated regulatory mutations are varied and diverse (McDonald, 1995). For example, retrotransposon insertions into a gene's 5′ flanking region may affect transcriptional initiation in a temporal-specific or tissue-specific manner. Such regulatory changes can be due to the read-through of transcripts initiated in the retrotransposon promoter or to the presence of positive or negative regulatory sequences within the element. Another way in which retrotransposons may influence gene expression is through insertion induced changes in chromatin structure which may, for example, insulate a gene's promoter from enhancer sequences located distal to the site of insertion (Gerasimova et al., 1995).

Although the hypothesis that transposable elements may have a dramatic effect on regulatory evolution was first proposed by McClintock over 40 years ago (e.g., McClintock, 1951, 1956), it is only recently that experimental evidence has begun to accumulate which directly supports the hypothesis. Instances of retrotransposons contributing to the evolution of chromosome gene regulation have recently been described in vertebrates (e.g., Robins & Samuelson, 1992), *Drosophila* (e.g., Miller et al., 1995) and plants (e.g., White, Habera & Wessler, 1994). Because it now seems likely that the regulatory evolution of at least some chromosomal genes has been influenced by retrotransposon insertions, the question arises as to what factors may be influencing the evolution of retrotransposon regulatory sequences in the first place.

For the past several years, our laboratory has been studying the evolution of enhancer-like regions within the *Drosophila copia* LTR retrotransposon. LTR retrotransposons are closely related to infectious retroviruses and encode genes homologous to the retroviral *gag*

a) SV 40 early genes:

...TCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCA...
-100 -60

b) rabbit β-globin:

...CCCAGACCTCACCCTG CAGAGCCACACCCTG GTGTTGGCCAATCTACACACGGGGTAGGGA...
-100 -60

Figure 1. Eukaryotic enhancers are characterized by short repeating motifs as exemplified in the SV 40 and rabbit globin gene enhancers. The numbering of sequences is relative to the CAP site (+1).

and *pol* genes flanked by long terminal direct repeats (LTRs). Because of the similarity between retroviruses and LTR retrotransposons with regard to their structure and regulatory controls, we will group them in this paper under the collective heading 'LTR retroelements'. The *cis*-regulatory sequences contained within LTR retroelement LTRs and adjacent untranslated leader regions (ULRs) interact with host encoded regulatory proteins to control LTR retroelement expression. Our results suggest that the molecular propensity of LTR retroelements to accumulate short regional duplications within non-coding regulatory regions coupled with the opportunity for selection on these regulatory regions within the context of evolving host genomes may combine to constitute an effective molecular drive mechanism for the evolution of eukaryotic enhancers.

LTR retroelements contain enhancer-like regions within their LTRs and ULRs

Enhancers are *cis*-acting sequences that increase the utilization of promoters usually in a tissue and/or developmental specific manner (Atchison, 1988). Enhancers characteristically consist of a series of short repeated sequence motifs that are often associated with regulatory protein binding domains (Serfling, Jasin & Schaffner, 1985; Maniatis, Goodbourn & Fisher, 1987; Tijan & Maniatis, 1994). For example, the well-studied enhancer of the simian virus-40 (SV-40) early genes consists of a 6 bp DNA sequence motif [CCGCCC] which is repeated six times (Figure 1a). The rabbit β-globin enhancer consists of two adjacent 14 bp sequence motifs [TGTG(G)A(A), TCCCCAG, AAGTATGCA] (Figure 1b). The repeated motifs within enhancers are usually binding sites for regulatory proteins and the strength of an enhancer (i.e., the relative effect the enhancer has on promoter initiation) is often positively correlated with the number of repeating motifs it contains (Serfling, Jasin & Schaffner, 1985).

LTR retroelement enhancer regions are usually located within the element's LTR and ULR. For example, the LTR of the human immunodeficiency virus-1 (HIV-1) contians an enhancer with three distinct 9–23 bp motifs, each of which is repeated twice. Thus far, two of these motifs have been identified as binding sites for the host regulatory proteins NF-κB and SP1 (Koken et al., 1992) (Figure 2a). An enhancer region within the mouse VL 30 LTR retrotransposon contains four distinct sequence motifs, each of which is repeated two or three times. In this case, each of the motifs are known to be binding sites for the host-encoded regulatory proteins AP-1, CREB, JUN, CarG, and NF-1 (Nilsson & Bohm, 1994) (Figure 2b). An enhancer region contained within the tobacco LTR retrotransposon TNT-1 consists of a single 26 bp motif repeated four times. The regulatory protein or proteins which may be interacting with the motifs in the TNT-1 enhancer have yet to be identified (Casacuberta, Vernhettes & Grandbastien, 1995) (Figure 2c). An enhancer within the ULR of the *Drosophila gypsy* endogenous retrovirus contains 12 copies of a 10 bp motif [(C/T)(A/G)(T/C)TGCATA] (Smith & Corces, 1995). This region is a binding site for the host regulatory protein encoded by the *suppressor of hairy wing (su(Hw))* gene (Spana, Harrison & Corces, 1988) (Figure 2d).

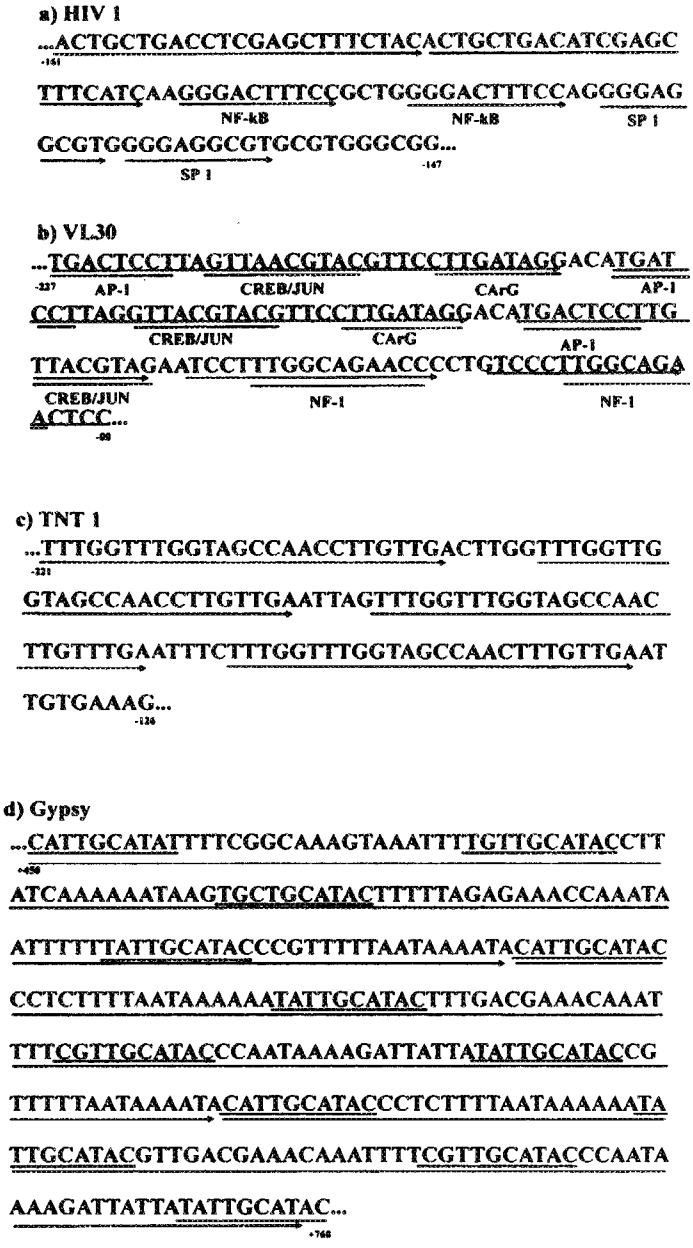

Figure 2. Functionally important sequences within LTR retroelement enhancers. Repeating sequence motifs are underlined and those known to be binding sites for host regulatory proteins are labeled. a) HIV-1: An enhancer region within the LTR of the human immunodeficiency virus-1 contains a series of short repeating sequence motifs. Many of these motifs are known to be binding sites for the host regulatory proteins NFκB and SP 1; b) VL 30: The murine VL 30 LTR retrotransposon enhancer region contains four short repeating motifs that are binding sites for host regulatory proteins AP-1, CREB, JUN, CarG and NF-1; c) TNT-1: The tobacco TNT-1 LTR retrotransposon enhancer region contains four 26 bp repeats. Although these sequences are believed to be binding sites for regulatory proteins, these proteins have not yet been identified; d) *Gypsy*: A 320 bp region within the untranslated leader region of the *Drosophila gypsy* LTR retrovirus is a binding site for the *suppressor-of-hairy-wing* (SUHW) protein. This region is an enhancer of *gypsy* expression and has also been shown to be a chromatin insulator. The region is made up of a series of four short sequence motifs that have been duplicated as a unit three times.

The characteristic pattern of repeating motifs present within LTR retroelement enhancers is a characteristic by-product of reverse transcriptase mediated replication

Examination of the LTR retrotransposon enhancers shown in Figure 2 reveals three characteristic patterns of repeats. The simplest pattern is a series of short tandem sequence motifs (aa..bb..) as present within the HIV-1, VL30, TNT-1, and *gypsy* enhancers. An intermediate pattern of complexity consisting of a series of short tandem motifs repeated as a unit two or more times (aa..bb..aa..bb..) is exemplified within the *gypsy* enhancer. An example of a more complex pattern in which two or more adjacent heterologous motifs are repeated as a group two or more times (ab..ab..) is present within the VL30 enhancer. Each one of these patterns can be generated during LTR retrotransposon replication.

Reverse transcriptase (RT) mediated LTR retroelement replication is a highly error prone process with no proofreading ability (Skalka & Goff, 1993). One common error in the reverse transcription process is the generation of short regional duplications of the 'aa..bb..' type (Burns & Temin, 1994). In addition, frequent recombination events are also known to occur between the two genomic RNA strands packaged within LTR retroelement capsids (Zhang & Temin, 1994). Regional mispairing between these two RNA templates and/or template switching errors during reverse transcription are capable of generating the more complex patterns found in many retroviral and LTR retrotransposon enhancers. LTR retroelements may also be subject to unequal ectopic recombination events between repeating elements within a genome (e.g., Olson & Temin, 1992). Such unequal DNA level exchanges could also contribute to the generation of the complex motif patterns seen in many retroviral and LTR retrotransposon enhancers.

The *Drosophila copia* LTR retroelement is a model system for the study of LTR retroelement enhancer evolution

Because LTR retroelements may be continually generating variation within their non-encoding enhancer regions, continuous opportunities may exist for natural selection to favor the evolution of adaptive enhancer configurations. This hypothesis rests upon the assumption that at leat some of the structural variability being generated within LTR retroelement enhancer regions provides a reproductive advantage upon which natural selection can act. We have begun to address this and related issues concerning the evolution of LTR retroelement enhancers within the context of the *Drosophila copia* LTR retroelement.

Copia is a good system in which to study LTR retroelement evolution for several reasons. *Copia* is an abundant and widely distributed *Drosophila* LTR retroelement (Berg & Howe, 1989). *Copia* expression has recently been shown to positively correlate with *copia* copy number and rates of retrotransposition (Pasyukova et al., 1997). Significant variation in *copia* expression exists both within and between *Drosophila* species and this variation is known to be under both *cis*- (element) and *trans*- (host) regulatory controls (Csink & McDonald, 1995; Matyunina, Jordan & McDonald, 1996). Of particular interest with regard to the issue of LTR retroelement enhancer evolution is the fact that the *copia* ULR contains a series of short sequence motifs characteristic of LTR retroelement enhancers. In the following sections, we review recent studies from our laboratory that indicate that naturally occurring variation in the number of repeating motifs within the *copia* ULR is correlated with the ability of this region to act as an enhancer. In addition, we find that the repeating motifs within the *copia* ULR are binding sites for at least two *Drosophila* regulatory proteins.

Copia *RNA levels are variable among* Drosophila *species*

Transcription is a major rate-limiting step in the retrotransposition process (Berg & Howe, 1989). Thus, naturally occurring genetic variation that influences the transcription of retrotransposons may be of evolutionary significance (McDonald, 1993). We previously reported that steady state levels of *copia* RNA in *Drosophila* adults varies significantly among *D. melanogaster* populations (Csink & McDonald, 1990). In that same study it was reported that no *copia* transcripts are detectable in *D. simulans* or *D. mauritiana* adults. We recently reported that transcript levels in larvae follow these same trends. Variation in *copia* transcript levels between *D. melanogaster* populations can vary ≈ 30-fold, whereas no transcripts are detectable in other *melanogaster* group species (*D. simulans, D. mauritiana, D. sechillia, D. yakuba, D. erecta*) nor in *D. willistoni* (Matyunina, Jordan & McDonald, 1996).

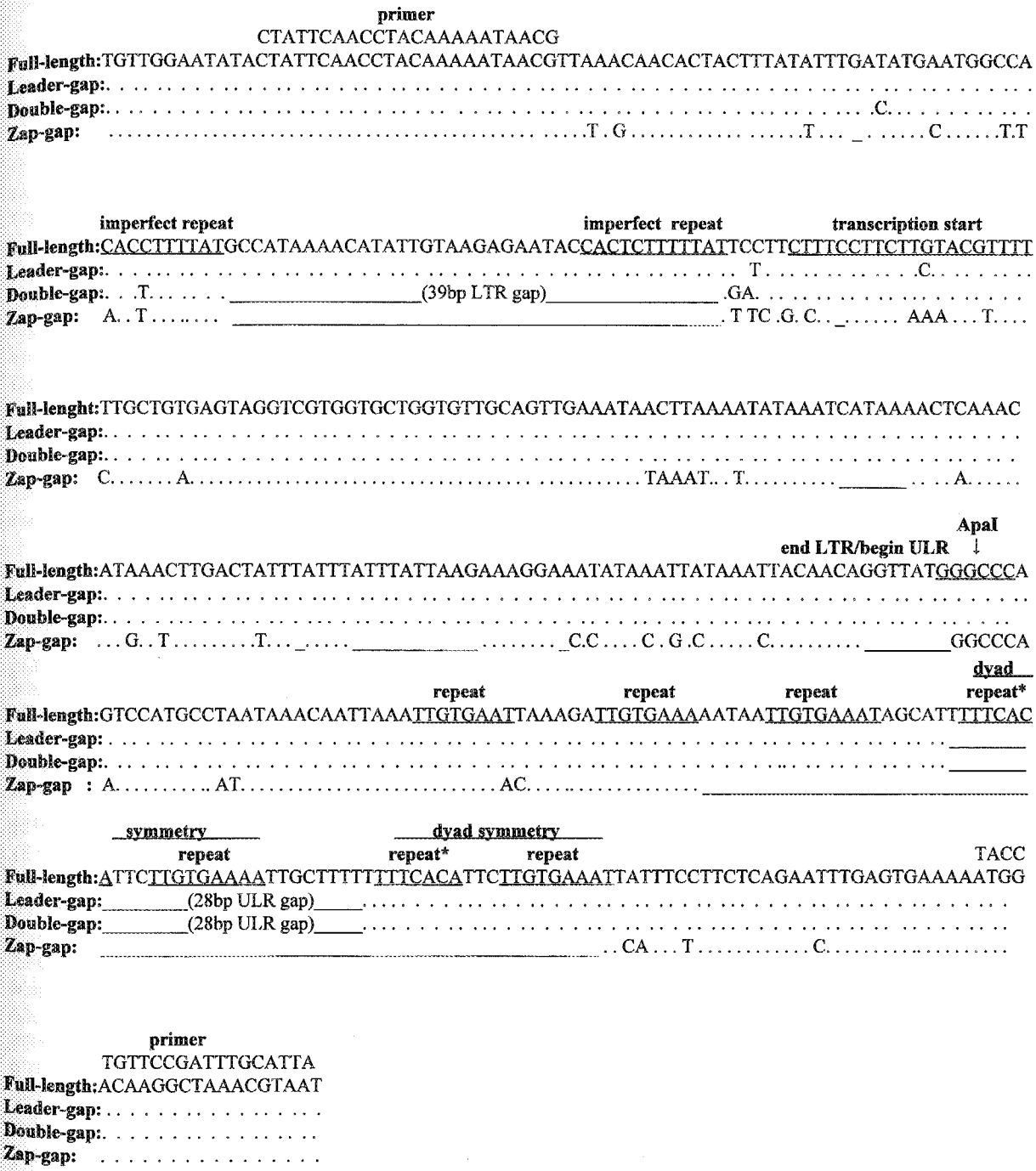

Figure 3. Naturally occurring *copia* elements are variable in the number of repeating sequence motifs present within their long terminal repeat and untranslated leader regions. The *Zaprionus tuberculatus copia* LTR-ULR has three repeats of a 9 bp motif similar to the core of the SV 40 enhancer within the ULR. Double-gap and ULR-gap variants present in many members of the *melanogaster* species group have five copies of this motif, whereas full-length *copia* elements, which are also found in many *melanogaster* and *willistoni* group species, contain seven copies of the motif.

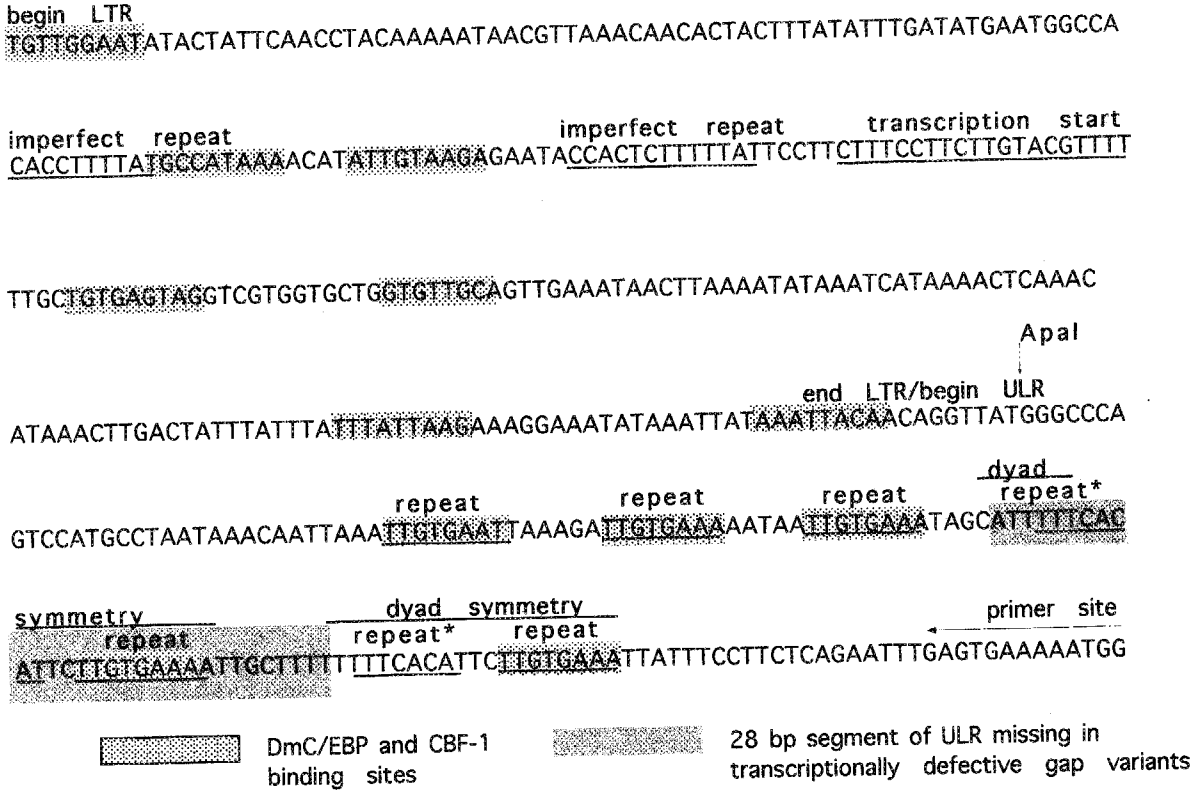

Figure 4. Binding sites for the *Drosophila* regulatory proteins DmC/EBP and CBF-1 are present in multiple copies within the *copia* long terminal repeat and untranslated leader. A computer search established that the 9 bp repeating motifs found within the *copia* ULR are similar to binding sites for the *Drosophila* regulatory protein C/EBP. Binding assays established that these sequences are indeed capable of binding Dm C/EBP. In addition, a second binding factor isolated from the nuclei of *Drosophila* cell lines (*copia* binding factor-1) was also shown to bind to these same C/EBP binding sites.

Size variation maps to the LTR-ULR of naturally occurring copia *elements*

We recently 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 (Csink & McDonald, 1995). Relative to *D. melanogaster copia* elements isolated from an Iquitos, Peru population, all of the *D. simulans* and *D. mauritiana* elements sequenced in this study were found to have a 39 bp gap within the 5′ LTR located just 5′ to the transcriptional start site (Figure 3). 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 ULR. This 28 bp ULR gap was also found in two of nine *copia* elements isolated from the Loua *D. melanogaster* population. The *copia* ULR contains seven dispersed copies of a conserved 9-bp motif [(T)TGTGAA(A/T)(A/T)], two (inverted repeat) copies of which are contained within the 28 bp region missing within the ULR gap variants (Figure 3).

We recently found that these same *copia* size variants are present in varying frequencies within the genomes of other members of the *melanogaster* subgroup and *D. willistoni* (Matyunina, Jordan & McDonald, 1996). More recently, we have isolated a new class of *copia* size variant from *Zaprionus tuberculatus* (Zap-gap) (Figure 3) (Jordan & McDonald, 1997). This variant contains a 67 bp gap at the same ULR position where the 28 bp gap is located in the ULR-gap variants described above. In addition, the Zap-gap variant has the same 39 bp LTR gap as the double-gap variants described above.

The copia *ULR has evolved by a series of regional duplications*

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 28 bp region within the ULR that contains two inverted repeats of the 9 bp motif described above is repeated twice in the full length variants, but only once in the ULR-gap variant (Figure 3). These repeated motifs are bordered by runs of T's that are known to facilitate recombination during reverse transcription (Burns & Temin, 1994). Sequences immediately adjacent to the region missing in the LTR of the double-gap variant are repeated within the LTR of the full-length variant, indicating that the evolution of full-length variants also involved at least one regional duplication within the LTR (Matyunina, Jordan & McDonald, 1996).

The 9 bp repeating motif within the copia *ULR is a binding site for at least two host-encoded proteins*

A computer search revealed a significant similarity between the 9 bp motif that is repeated seven times within the full-length *copia* ULR and the consensus C/EBP (CCAAT/enhancer binding protein) binding site [T(T/G)NNG(C/T)AA(T/G)] (Figure 4). C/EBP is a mammalian transcriptional activator (Graves, Johnson & McKnight, 1986). It has recently been reported that C/EBP is a trans-regulator of the human LTR retroelement, HIV-1 (human immunodeficiency virus-1) (Ruocco et al., 1996).

A homologue of mammalian C/EBP has recently been identified in *Drosophila* and implicated in the regulation of tissue-specific gene expression (Montell et al., 1992). To determine if the 9 bp repeating motifs within the *copia* ULR are capable of binding *Drosophila* C/EBP (DmC/EBP), we carried out a series of mobility shift assays with purified protein (Wilson, Matyunina & McDonald, 1997). We utilized three different probes in our assays: a sequence encompassing the entire LTR/ULR, a sequence encompassing the ULR only, and a short 28-bp oligomer identical in sequence to the region mssing in the ULR gap variants. Incubation of the 28-bp oligomer probe with DmC/EBP gives a single protein-DNA complex, whereas 2–4 complexes (depending on probe and incubation conditions) resulted from incubations with the ULR gap and full-length ULR probes.

In addition, we have recently identified another nuclear protein factor, *copia* binding factor-1 (CBF-1), which also binds to the 9 bp motifs within the *copia* ULR. CBF-1 was first isolated from the *Drosophila* Schneider-2 cell (S2) line. S2 cells have been previously found not to contain DmC/EBP (P. Rorth, personal communication). As was the case for DmC/EBP, CBF-1 bound specifically to each ULR probe. To ensure that CBF-1 was not DmC/EBP and to verify that our nuclear extracts did not contain DmC/EBP, which may be interacting with CBF-1, we competed incubations of our probes and S2 extract with DmC/EBP antibodies. The antibodies that interfere with the binding of C/EBP with our probes had no effect on the binding profiles of our probes with nuclear extracts, indicating that CBF-1 is indeed distinct from DmC/EBP (Wilson, Matyunina & McDonald, 1997).

Naturally occurring variation in the number of repeating motifs within the copia *LTR-ULR is of functional significance*

To determine if naturally occurring variation in the number of motifs within the *copia* LTR-ULR may be of functional (potentially adaptive) significance, we conducted a series of transient expression assays to test the relative ability of the full-length, double-gap, and ULR-gap variants to drive expression of a bacterial CAT reporter gene in Iquitos *D. melanogaster* (Matyunina, Jordan & McDonald, 1996). Transient expression assays are useful in identifying both cis- and trans-regulatory effects that influence transcriptional initiation extrachromosomally. The results summarized in Figure 5 indicate that: i) all three of the naturally occurring variants are expressed at levels significantly above negative controls ($P < 0.0025$); ii) the full-length variant is expressed at levels significantly higher than either the ULR-gap (t-test, $P < 0.001$) or the double-gap ($P < 0.001$) variants; iii) the ULR-gap variant is significantly higher than that associated with the presumed ancestral double-gap variant. From these results we conclude that the naturally occurring variation in number of 9 bp motifs contained within the *copia* LTR-ULR is of potential adaptive significance.

Host-regulatory control of copia *expression is variable among* Drosophila *populations and species*

To determine if host encoded trans-regulatory variation may be contributing to interspecific differences in levels of *copia* expression, we tested the ability of the three LTR-ULR variants described above to drive expression of a bacterial CAT reporter gene in mem-

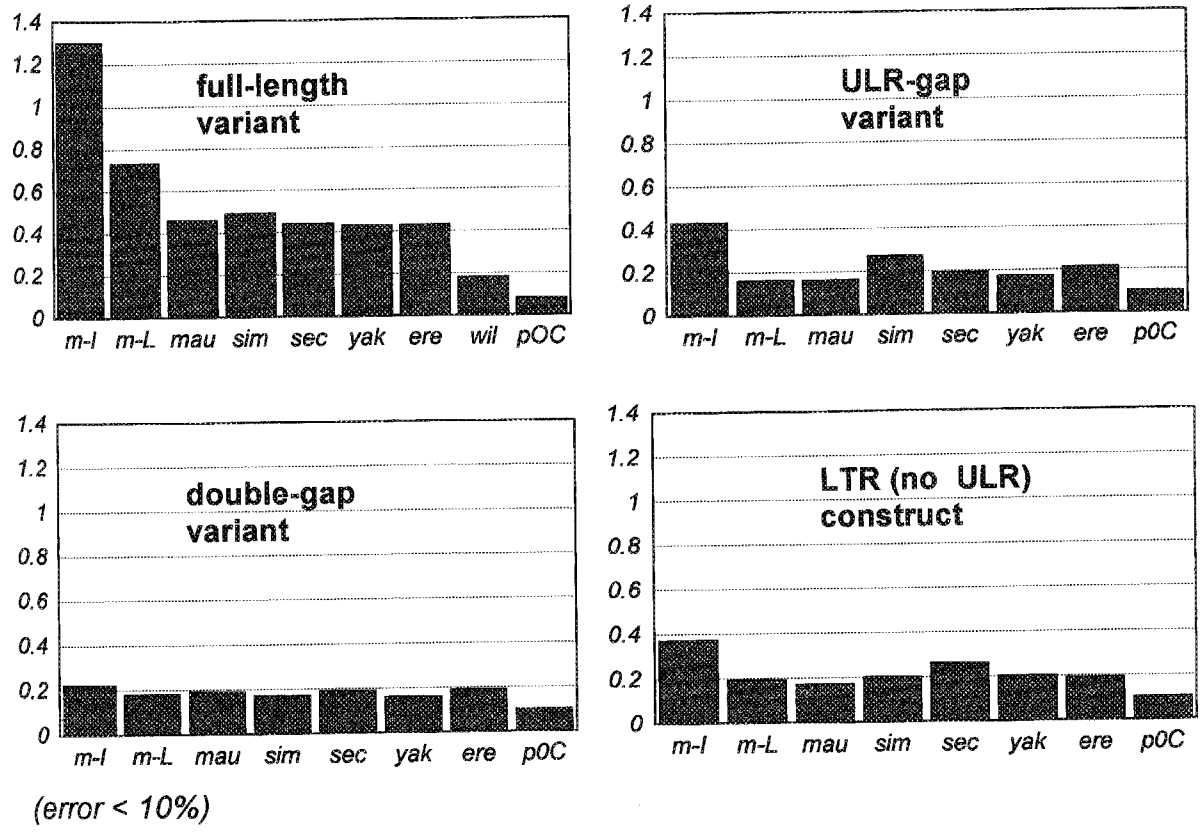

Figure 5. The relative ability of naturally occurring *copia* LTR/ULRs to drive transient expression of the CAT reporter gene in *Drosophila* larvae. *Copia* LTR/ULR's representing the three major LTR/ULR size variants in *Drosophila* species were cloned in frame with the bacterial chloramphenicol acetyl transferase gene (CAT) in pCAT (Promega). The basic pCAT vector, which lacks any promoter, was used as our negative control plasmid. 4.6 nl of reporter plasmid DNA was injected per embryo. Ten active 1st and 2nd instar larvae representing each species strain examined were harvested 48 h after injection and individually homogenized and the supernatant assayed for CAT inactivity (Matyunina, Jordan & McDonald, 1996). Shown are mean units of CAT activity (1 unit = pmole of acetyl groups from acetyl coenzyme A to chloramphenicol in 1 min at 37 °C per larvae) per larva. Error variation (SE/X) was $\leq$ 10%.

bers of the *melanogaster* species group and the outlier *D. willistoni* (Matyunina, Jordan & McDonald, 1996). Our results, which are summarized in Figure 5, indicate 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 species 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 $\approx$ 70% in strains representing other members of the *melanogaster* species group. 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. No significant difference was associated with the nearly inactive double-gap variant. The ULR-gap variant is expressed at highest levels in the Iquitos strain and significantly lower in all other species strains tested ($P < 0.001$).

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 demonstrate that removal of the entire ULR results in nearly the same levels of expression associated with the naturally occurring ULR-gap variant. As a positive control for all of the above studies, we tested the ability of the *D. melanogaster hsp70* promoter to drive CAT expression in all of the species strains in which the LTR-ULR-CAT constructs were tested. No significant difference in *hsp-70*-CAT expression was detected between any of the strains (Matyunina, Jordan & McDonald, 1996).

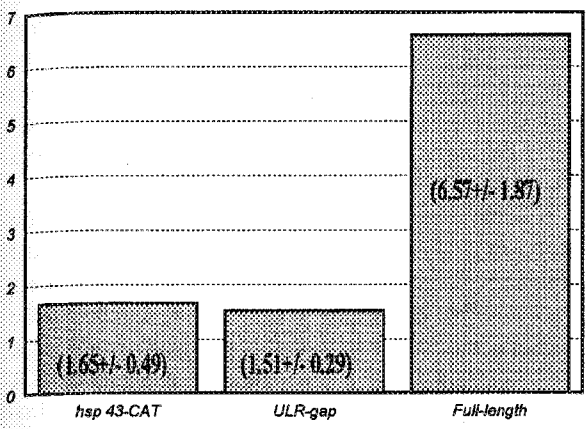

Figure 6. Effect of the insertion of a *copia* ULR containing either five (ULR-gap) or seven (full-length) copies of the 9 bp motif on transient expression of the *hsp 70* minimal promoter. The full-length ULR exerts a significant enhancer effect of *hsp 70* expression. The ULR-gap variant had no significant effect on *hsp 70* expression. The control was the *hsp-70* minimal promoter fused to CAT. Microinjections at CAT activity readings were carried out as described in Figure 5.

The full-length copia *ULR is capable of enhancing expression of a heterologous chromosomal gene*

The results summarized up to now indicate that naturally occurring *copia* variants lacking a 28 bp region within their ULR have significantly reduced expression relative to variants containing the 28 bp region. To determine whether the *copia* ULR with or without the 28 bp gap has properties characteristic of a eukaryotic enhancer, we conducted transient expression assays to test if a *copia* full-length ULR and a ULR lacking the 28 bp region could influence the ability of a heterologous *hsp 70* minimal promoter to drive expression of a bacterial CAT reporter gene in *D. melanogaster*. The results summarized in Figure 6 indicate that when the full-length ULR is placed upstream of the *hsp 70* minimal promoter, it stimulates expression four-fold relative to the control. The ULR lacking the 28 bp region, on the other hand, had no significant enhancer effect on the *hsp 70* minimal reporter relative to the control.

Summary and perspectives

We propose that LTR retrotransposons may constitute a mechanism by which eukaryotic enhancers evolve and are subsequently distributed throughout eukaryotic genomes over evolutionary time. This model is based on three hypotheses: 1) that reverse transcriptase mediated LTR retroelement replication is prone to the generation of short regional duplications that are characteristic of eukaryotic enhancers; 2) that at least some of this size variation within LTR/ULRs can lead to the emergence of new LTR retroelement enhancers and/or to an increase in the strength of existing enhancers thereby providing an opportunity for selection; 3) that enhancers that evolve by this mechanism within LTR retroelements also have the potential to exert *cis*-regulatory effects on the expression of chromosomal genes when distributed throughout the genome by retrotransposition.

Frequent slippage and recombination events during reverse transcriptase mediated LTR retroelement replication is a well established fact (Parthasarathi et al., 1995; Preston & Dougherty, 1996). Although short duplications, deletions, and rearrangements due to such errors can be generated throughout the LTR retroelement genome, active LTR retroelements would only be expected to accumulate size variation within their non-encoding LTR and ULRs where reading frames would not be disrupted. Consistent with this prediction, we have found *copia* LTR/ULR size variation segregating within and between *Drosophila* species. Moreover, this size variation appears to be the result of short regional duplications of the type known to be generated during LTR retroelement replication.

The second hypothesis underlying our model is that at least some of the size variation within the LTR/ULRs may have a significant effect on LTR retroelement expression and thus potentially be subject to natural selection. To determine if naturally occurring *copia* LTR/ULR size variation may be of functional significance, we tested the ability of the variant LTR/ULRs to transiently drive expression of a reporter gene in *Drosophila* larvae. Our results demonstrate that full-length elements are associated with the highest level of transcriptional activity followed by ULR-gap and double gap variants, respectively. With regard to the ULR region, our results indicate that the duplication event that increased the number of 9 bp motifs from five (ULR-gap variant) to seven (full-length variants) resulted in a significant elevation in the ULR's ability to enhance *copia* driven expression. Since *copia* transcription correlates with *copia* retrotransposition (Pasyukova et al., 1997), our results predict that full-length elements may be selectively favored in genetic contexts which support *copia* expression. Further studies are required to test this hypothesis rigorously. However, our preliminary findings, that the Iquitos,

Peru genome strongly supports *copia* expression and is fixed for full-length *copia* elements, are consistent with the predictions of the model.

The third hypothesis upon which our model rests is that those changes within a LTR/ULR that enhance LTR retroelement expression will also have the potential to enhance expression of chromosomal genes. As a preliminary test of this hypothesis, we again employed transient expression assays to monitor the ability of *copia* ULRs from the ULR-gap and full-length variants to enhance expression of a heterologous gene promoter. We found that only the full-length ULR was able to significantly enhance expression of the minimal *hsp 70* promoter. Thus, the same ULR regional duplication that increased *copia*-driven transient expression also resulted in enhanced expression of an adjacent chromosomal promoter.

Recent findings indicate that eukaryotic enhancers act by facilitating the formation of stable domains within which promoter activity is permitted (e.g., Moon & Ley, 1990; Walters et al., 1995, 1996). Although the effect of enhancers on the expression of extrachromosomal gene constructs, as monitored in transient assays, is considered to be the same as for constructs stably integrated within chromosomes, the later situation is clearly more complex. Stably integrated genetic constructs have been shown to be relatively more efficient at creating a transcriptionally active state within regions of inactive chromatin when they contain an enhancer (Walters et al., 1996). By implication, we envision that LTR retrotransposons with the strongest enhancers may have the highest probability of being expressed in any chromosomal context.

The view that LTR retrotransposons and other classes of transposable elements located within heterochromatin are *ipso facto* biologically inactive is no longer tenable. Recent studies have shown that heterochromatin is far more complex than once envisioned (Lohe & Hilliker, 1995). In particular, heterochromatin in *Drosophila*, which was once thought to be genetically inert, has recently been shown to consist of 'islands' of transcriptional activity that are associated with middle repetitive sequences (e.g., Devlin, Bingham & Wakimoto, 1990; Mitchelson et al., 1993). At least some of these middle repetitive sequences have recently been identified as transposable elements (Zang & Sprading, 1995). Because *copia* and many other LTR retrotransposons are known to reside within both euchromatic and heterochromatic regions of the genome, it seems likely that selective pressure may be placed on these elements to evolve enhancers that can function in both chromosomal contexts. In this regard, studies are currently in progress to determine if the transcriptional superiority associated with full-length *copia* elements in our transient assays is maintained when the *copia* constructs are stably integrated into a variety of chromosomal locations.

Clearly, additional studies on retrotransposon enhancer evolution in *Drosophila* and other biological systems will be required before general conclusions can be drawn. However, our preliminary studies on the evolution of enhancer sequences within the *copia* ULR in *Drosophila* are consistent with the model that LTR retroelements and perhaps other retrotransposons (Brosius & Tiedge, 1996) constitute drive mechanisms for the evolution of eukaryotic enhancers. The subsequent distribution of these enhancers throughout genomes by retrotransposition may play a significant role in eukaryotic regulatory evolution.

Acknowledgement

Our research is supported by National Science Foundation Grant DEB-9307220 to J.F.M.

References

Atchison, M.L., 1988. Enhancers: mechanism of action and cell specificity. Annu. Rev. Cell Biol. 4: 127–153.

Berg, D.E. & M.M. Howe (editors), 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.

Brosius, J. & H. Tiedge, 1996. Reverse transcriptase: mediator of genomic plasticity. Virus Genes 11: 163–179.

Burns, D.P. & H.M. Temin, 1994. High rates of frameshift mutations within homo-oligomeric runs during a single cycle of retroviral replication. J. Virol. 68: 4196–4203.

Casacuberta, J., S. Vernhettes & M-A. Grandbastien, 1995. Sequence variability within the tobacco retrotransposon Tnt 1 population. EMBO J. 14: 2670–2678.

Csink, A.K. & J.F. McDonald, 1990. *Copia* expression is variable among natural populations of *Drosophila*. Genetics 126: 375–385.

Csink, A.K. & J.F. McDonald, 1995. Analysis of *copia* sequence variation within and between *Drosophila* species. Mol. Biol. Evol. 12: 83–93.

Devlin, R.H., B. Bingham & B.T. Wakimoto, 1990. The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila melanogaster*. Genetics 127: 553–565.

Graves, B.J., P.F. Johnson & S.L. McKnight, 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV *tk* gene. Cell 44: 565–576.

Gerasimova, T.I., D.A. Gdula, D.V. Gerasimova, O. Simonova & V.G. Corces, 1995. A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. Cell 82: 587–597.

Jordan, I.K. & J.F. McDonald, 1997. Evolution of the *copia* retrotransposon in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. (submitted).

Koken, S.E., J.L. van Wamel, J. Goudsmit, B. Berkhout & J.L. Geelen, 1992. Natural variants of the HIV-1 long terminal repeat: analysis of promoters with duplicated DNA regulatory motifs. Virology 191: 968–972.

Lohe, A.R. & A.J. Hilliker, 1995. Return of the H-word (heterochromatin). Curr. Biol. 5: 746–755.

Makalowski, W., 1995. SINES as agenomic scrap yard: an essay on genomic evolution, pp. 81–104 in The Impact of Short Interspersed Elements (SINES) on the Host Genome, edited by R.J. Maraia. R.G. Landes Co., Austin, TX.

Maniatis, T., S. Goodbourn & J.A. Fisher, 1987. Regulation of inducible and tissue-specific gene expression. Science 236: 1237–1244.

Matyunina, L.V., I.K. Jordan & J.F. McDonald, 1996. Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation. Proc. Natl. Acad. Sci. (USA) 93: 7097–7102.

McClintock, B., 1951. Chromosome organization and genic expression. Cold Spr. Harb. Symp. Quant. Biol. 16: 13–47.

McClintock, B., 1956. Controlling elements and the gene. Cold Spr. Harb. Symp. Quant. Biol. 21: 197–216.

McDonald, J.F., 1995. Transposable elements: possible catalysts of organismic evolution. Trends Ecol. Evol. 10: 123–126.

McDonald, J.F., 1993. Evolution and consequences of transposable elements. Curr. Opin. Gen. Dev. 3: 855–864.

Mitchelson, A.M., M. Simonelig, C. Williams & K. O'Hare, 1993. Homology with *Saccharomyces cerevisiae* RNA14 suggests that phenotypic suppression in *Drosophila melanogaster* by suppressor of forked occurs at the level of RNA stability. Genes Dev. 7: 241–249.

Miller, W.J., N. Paricio, S. Hagemann, M.J. Martinez-Sebastian, W. Pinsker & R. de Frutos, 1995. Structure and expression of clustered P element homologues in *Drosophila subobscura* and *Drosophila guanche*. Gene 156: 167–174.

Moon, A.M. & T.J. Ley, 1990. Conservation of the primary structure, organization, and function of the human and mouse β-globin locus-activating regions. Proc. Natl. Acad. Sci. (USA) 87: 7693–7697.

Montell, D.J., P. Rorth & A.C. Spradling, 1992. Slow border cells, a locus required for a developmentally regulated cell migration during oogenesis, encodes *Drosophila* C/EBP. Cell 71: 51–62.

Nilsson, M. & S. Bohm, 1994. Inducible and cell type-specific expression of VL30 U3 subgroups correlate with their enhancer design. J. Virol. 68: 276–288.

Olson, P. & H.M. Temin, 1992. Unusually high frequency of reconstitution of long terminal repeats in U3-minus retrovirus vectors by DNA recombination or gene conversion. J. Virol. 66: 1336–1343.

Parthasarathi, S., A. Varela-Echavarria, Y. Ron, B.D. Preston & J.P. Dougherty, 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. J. Virol. 69: 7991–8000.

Pasyukova, E., S. Nuzhdin, L. Wei & A. Flavell, 1997. Germ line transposition of the *copia* retrotransposon in *Drosophila melanogaster* is restricted to males by tissue-specific control of *copia* RNA levels. Mol. Gen. Genet. (in press).

Preston, B.D. & J.P. Dougherty, 1996. Mechanisms of retroviral mutation. Trends Microb. 4: 16–21.

Robins, D.M. & L.C. Samuelson, 1992. Retrotransposons and the evolution of mammalian gene expression. Genetica 86: 191–202.

Rorth, P. & D.J. Montell, 1993. *Drosophila* C/EBP: a tissue-specific DNA-binding protein required for embryonic development. Genes Dev. 6: 2299–2311.

Ruocco, M.R., X. Chen, C. Ambrosino, E. Dragonetti, W. Liu, M. Mallardo, G. De Falco, C. Palmieri, G. Franzoso, I. Quinto, S. Ventura & G. Scala, 1996. Regulation of HIV-1 long terminal repeats by interaction of C/EBP (NF-IL6) and NfkappaB/Rel transcription factors. J. Biol. Chem. 271: 22479–22486.

Serfling, E., M. Jasin & W. Schaffner, 1985. Enhancers and eukaryotic gene transcription. Trends Genet. 1: 224–230.

Skalka, A.M. & S.P. Groff (editors), 1993. Reverse Transcriptase. Cold Spr. Harb. Press, Plainview, N.Y.

Smith, P.A. & V.G. Corces, 1995. The *suppressor-of-hairy-wing* protein regulates the tissue-specific expression of the *Drosophila gypsy* retrotransposon. Genetics 139: 215–228.

Spana, C., D.A. Harrison & V.C. Corces, 1988. The *Drosophila melanogaster suppressor-of-hairy-wing* protein binds to specific sequences of the *gypsy* retrotransposon. Genes Dev. 2: 1414–1423.

Tijan, R. & T. Maniatis, 1994. Transcriptional activation: a complex puzzle with few easy pieces. Cell 77: 5–8.

Walters, M.C., W. Magis, S. Fiering, J. Eidemiller, D. Scalzo, M. Groudine & D.I.K. Martin, 1996. Transcriptional enhancers act in *cis* to suppress position-effect variegation. Genes Dev. 10: 185–195.

White, S.E., L. Habera & S.R. Wessler, 1994. Retrotransposons in the flanking regions of normal plant genes: a role of *copia*-like elements in the evolution of gene structure and expression. Proc. Natl. Acad. Sci. (USA) 91: 11792–11796.

Wessler, S.R., T.E. Bureau & S.E. White, 1995. LTR-retrotransposons and LITES: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5: 814–821.

Wilson, S., L. Matyunina & J.F. McDonald, 1997. An enhancer-like region with the *copia* untranslated leader which is variable in natural populations contains binding sites for *Drosophila* regulatory proteins. Nucl. Acids Res. (submitted).

Zang, P. & A.C. Spradling, 1995. The *Drosophila* salivary gland chromocenter contains highly polytenized subdomains of mitotic heterochromatin. Genetics 139: 659–670.

Zang, J. & H.M. Temin, 1994. Retrovirus recombination depends on the length of sequence identity and is not error prone. J. Virol. 68: 2409–2414.