J Mol Evol (1998) 47:670–676

JOURNAL OF MOLECULAR EVOLUTION

Interelement Selection in the Regulatory Region of the *copia* Retrotransposon

I. King Jordan, John F. McDonald

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

Received: 5 February 1998 / Accepted: 14 May 1998

Abstract. We report the results of an analysis of naturally occurring cis-regulatory variation within and between two families of the *copia Drosophila* long terminal repeat (LTR) retrotransposon. The *copia* 5′ LTR and adjacent untranslated leader region (ULR) consists of a number of well-characterized sequence motifs which play a role in regulating expression of the element. In order to understand the evolutionary forces which may be responsible for generating and maintaining *copia* regulatory sequence variation, we have quantified levels of naturally occurring *copia* LTR-ULR nucleotide variation and subjected the data to a series of tests of neutrality. Our analysis indicates that the *copia* LTR-ULR has been subject to negative purifying selection within families and positive adaptive selection between families. We discuss these findings with respect to the regulatory evolution of retrotransposons and the phenomenon of interelement selection.

Key words: *Drosophila* — *copia* — Retrotransposons — Regulatory evolution — Selection

Introduction

Long terminal repeat (LTR) retrotransposons are a class of repetitive, mobile DNA sequences which transpose via the reverse transcription of an RNA intermediate. Retrotransposons are ubiquitous components of eukaryotic genomes, and insertions of these elements are a primary source of spontaneous mutation (Berg and Howe 1989). Retrotransposon insertions are known to cause species specific differences in patterns of gene expression and chromosome structure (McDonald 1993, 1998; White et al. 1994). Retrotransposons have also been shown to be important in genome organization and can even take on basic cellular functions (Biessmann et al. 1992; Levis et al. 1993; SanMiguel et al. 1996). These facts taken together indicate a tremendous potential for retrotransposons to affect the evolution of their host species (Wessler et al. 1995; Miller et al. 1996; McDonald 1995, 1998). The study of retrotransposons from an evolutionary perspective is a burgeoning area of research (e.g. Capy 1998). The interaction between host genomes and retrotransposons provides fertile ground for studying coevolutionary processes. One particular area of coevolutionary investigation focuses on how host trans factors interact with retrotransposon cis sequences to regulate the expression of the elements and how this relates to the maintenance and spread of retrotransposons in natural populations.

A number of studies, conducted by our lab and others, have utilized the *Drosophila* LTR retrotransposon *copia* as a model system for studying the interaction between element regulatory sequences and host factors which regulate element expression. *Copia* is ~5 kb in length and consists of two LTRs flanking a single open reading frame (Fig. 1). The *copia* transcript which initiates in the 5′ LTR and terminates in the 3′ LTR contains homology to the gag and pol loci of retroviruses (Mount and Rubin 1985). Deletion analysis initially indicated that the 5′ region of *copia,* which consists of the 5′ LTR and adjacent untranslated leader region (ULR), contains regula-

Correspondence to: I. King Jordan; *e-mail:* king@parvati.lv-whi.nevada.edu

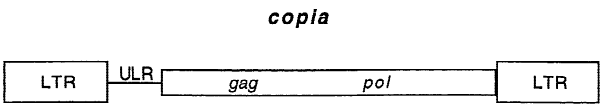

Fig. 1. Genomic organization of the *copia* retrotransposon (not to scale). *Copia* is ~5 kb in length and consists of a single open reading frame with homology to the *gag* and *pol* loci of retroviruses (Mount and Rubin 1985) flanked by two long terminal repeats (LTR). The untranslated leader region (ULR) is adjacent to the 5′ LTR.

tory sequences responsible for controlling *copia* transcription (Sneddon and Flavell 1989), a rate-limiting step in the retrotransposition process (Boeke et al. 1985). Subsequent studies have identified a series of cis-acting regulatory sequences in this region of *copia* which interact with host factors to regulate expression of the element. The 5′ LTR contains promoter sequences and the start site of transcription (Arkhipova et al. 1995). The ULR contains a number of sequence motifs which bind host regulatory factors and function as enhancers for the element (Sneddon and Flavell 1989; Cavarec and Heidmann 1993; Cavarec et al. 1994, 1997; Matyunina et al. 1996; Wilson et al. 1998). While these types of studies have been very informative, they have relied largely on experimental manipulation of canonical *copia* regulatory sequences. In order to add another dimension to the study of *copia* regulation, we have investigated the naturally occurring molecular variation in the 5′ LTR-ULR regulatory region of the *copia* element. Analysis of naturally occurring molecular variation of *copia* regulatory sequences can help achieve an understanding of *copia* regulatory features in an evolutionary context.

The neutral theory of molecular evolution (Kimura 1983) provides both a theoretical framework for evaluating genetic variability in natural populations and a falsifiable null hypothesis. A number of rigorous statistical tests of neutrality based on the analysis of the nucleotide sequence variation of a given gene or genes within and between species have been devised and implemented within the last decade (e.g., Wayne and Simonsen 1998). One of the most widely employed tests of neutrality is the McDonald–Kreitman (1991) test. This test is based on the null hypothesis that the ratio of replacement to synonymous nucleotide changes should be the same both within species (polymorphism) and between species (fixed differences) if all nucleotide variations are selectively neutral. Replacement changes are nucleotide substitutions which change the amino acid sequence of a protein, while synonymous or silent changes are nucleotide substitutions which do not change the amino acid sequence. The McDonald–Kreitman test consists of a 2 × 2 contingency table with four classes of observations derived from a nucleotide sequence alignment: (1) the number of polymorphic replacement changes, (2) the number of polymorphic synonymous changes, (3) the number of fixed replacement changes, and (4) the number of fixed synonymous changes. Selection will tend to affect levels of polymorphism and divergence in opposite directions and yield a significant deviation from neutrality (McDonald and Kreitman 1991).

While the McDonald–Kreitman test was originally designed with protein coding sequences in mind, it can be adapted to analyze noncoding regulatory sequences as well. In this case, the test compares different functional classes of mutations rather than synonymous versus amino acid replacement changes (Hudson 1993; Jenkins et al. 1995; Ludwig and Kreitman 1995). For example, a number of cis regulatory sequences which bind trans regulatory proteins have been identified within the *copia* LTR-ULR. Thus the *copia* LTR-ULR can be divided into sequences which bind transcription factors or other important trans regulatory proteins (replacement sites) and those which do not (silent sites) (Jenkins et al. 1995). Application of these data to a modified version of the McDonald–Kreitman test can be used to test the neutrality of the LTR-ULR regulatory sequences (Jenkins et al. 1995; Ludwig and Kreitman 1995). In our survey of naturally occurring *copia* LTR-ULR variation, we employed this test of neutrality and others in an attempt to gain an understanding of the evolutionary forces responsible for establishing and maintaining *copia* regulatory sequence variation.

Materials and Methods

Copia *Sequence Analysis*

The *melanogaster* family *copia* sequences isolated from populations of *D. melanogaster, D. simulans, D. mauritiana, D. sechellia, D. teissieri, D. yakuba,* and *D. erecta* species of the *melanogaster* subgroup as well as one *repleta* family *D. buzzatii copia* sequence have been reported elsewhere (Csink and McDonald 1995; Matyunina et al. 1996; Jordan and McDonald 1998). Additional *melanogaster* family *copia* sequences analyzed in this study were retrieved from Genbank and correspond to the following accession numbers: *D. melanogaster,* X02599, AC001281, AC002138, and J01075; and *D. simulans,* D10880. The *repleta* family *copia* sequences were isolated from *D. buzzatii* and *D. koepferae* and correspond to Genbank accession numbers X96972 and X96971, respectively.

A total of 26 *melanogaster* family (13 *D. melanogaster,* 4 *D. simualns,* 1 *D. mauritiana,* 2 *D. sechellia,* 1 *D. teissieri,* 2 *D. yakuba,* and 3 *D. erecta*) and 3 *repleta* family (2 *D. buzzatii* and 1 *D. koepferae*) *copia* LTR-ULR sequences was aligned using the PILEUP program of the Wisconsin GCG computer package followed by visual inspection and adjustment. Levels of nucleotide diversity (π) and numbers of segregating sites (S) were then calculated using the DnaSP program (Rozas and Rozas 1997). Nucleotide diversity (π) was calculated using the method of Lynch and Crease (1990) with the Jukes and Cantor (1969) correction.

Statistical Tests

The Tajima (1989) and Fu and Li (1993) tests of neutrality, each of which rely on different estimates of the neutral mutation parameter θ, were performed using the DnaSP program (Rozas and Rozas 1997).

Table 1. *Copia* LTR-ULR regulatory sequences

Regulatory site	Reference(s)
LTR	
Initiator (Inr)	Arkhipova et al. (1995), Flavell et al. (1981)
Downstream element (De)	Arkhipova et al. (1995), Flavell et al. (1981)
ULR	
$tRNA^{Met}$ primer binding site (Pbs)	Kikuchi et al. (1986)
Antennapedia homeodomain (Antp HD) binding site	Cavarec et al. (1994)
9.2.1.AB protein binding site	Cavarec et al. (1997)
Drosophila CCAAT/enhancer binding protein (CEBP)	Wilson et al. (1998)
Copia binding factor-1 (CBP-1) binding sites	
Box-B-binding factor-2 site (BBF-2)	Wilson et al. (1998)

Tajima (1989) initially proposed this type of test with the rationale that uniquely derived estimates of θ will be different when selection is present. A test statistic is calculated comparing the different estimates of θ. A negative value of the test statistic indicates that purifying selection is acting on a sequence. Tests of this type can be performed using any pair of estimates of θ given that they are sufficiently different when selection is present (Li 1997). Fu and Li (1993) proposed a similar test using different estimates of θ.

We also employed a modified version of the McDonald–Kreitman test (1991) using the DnaSP program (Rozas and Rozas 1997). In the modified McDonald–Kreitman test, nucleotide substitutions in regulatory protein binding sites were taken as replacement changes while nucleotide substitutions in non-protein-binding sites were taken as silent changes (Jenkins et al. 1995; Ludwig and Kreitman 1995). The McDonald–Kreitman test was further modified such that comparisons between nucleotide variation within element families (polymorphisms) and between element families (divergence) were substituted for the standard comparisons of variation within and between species. This modification was necessary because the unique evolutionary dynamics of multicopy transposable elements (including possible horizontal transfers) do not allow element nucleotide variation to be definitively partitioned along species lines.

Results

LTR-ULR Binding Sites

For the last decade the *Drosophila* LTR retrotransposon *copia* has been used as a model system to study the regulatory interactions between retroelements and their host genomes. A number of studies have documented both host trans factors and element cis regulatory sequences which mediate the regulation of *copia* expression (Sneddon and Flavell 1989; Cavarec and Heidmann 1993; Cavarec et al. 1994, 1997; Matyunina et al. 1996; Wilson et al. 1998). *Copia* cis regulatory binding sites have been mapped to the element's 5′ LTR and adjacent ULR (Table 1). The LTR contains the initiator and downstream element which regulate transcriptional initiation (Flavell et al. 1981; Arkhipova et al. 1995). The *copia* ULR sequence begins with the $tRNA^{Met}$ primer binding site where reverse transcription is initiated (Kikuchi et al. 1986). The ULR has also been shown to contain a number of sequence motifs which bind host trans regulatory proteins (Cavarec et al. 1994, 1997; Matyunina et al. 1996; Wilson et al. 1997).

In this study, we conducted a survey of naturally occurring *copia* LTR-ULR sequence variation. We determined the level of variation in this regulatory region of the *copia* element and consider the results with respect to the location of regulatory protein binding sites. Figure 2 shows the location of these binding sites in the *copia* LTR-ULR region relative to the nucleotide diversity present among naturally occurring *copia* elements. Nucleotide variation tends to be reduced in *copia* regulatory binding sites, which suggests that these regulatory sequences may have been subject to selective constraint over evolutionary time.

Nucleotide Variation

A number of *copia* LTR-ULR sequences isolated from host species of the *melanogaster* subgroup have been characterized previously in our lab (Csink and McDonald 1995; Matyunina et al. 1996; Jordan and McDonald 1998). In addition to these sequences, a number of other *copia* LTR-ULR sequences from Genbank were used in our study. The *Drosophila copia* sequences thus far isolated can be grouped into two families: the *melanogaster* family and the *repleta* family (Jordan and McDonald 1998). In our survey of *copia* LTR-ULR sequence variation, we quantified levels of nucleotide polymorphism within each family and fixed differences between the two families (Table 2). Numbers of segregating sites (S) and levels of nucleotide diversity (π) were determined for the LTR and ULR using the DnaSP program (Rozas and Rozas 1997).

Although only three *repleta* family sequences were available for this study, these elements are more diverged than members of the *melanogaster* family in both the LTR and the ULR. The *repleta* family sequences show nucleotide diversities 4.2 and 6.1× higher than the *melanogaster* family sequences in the LTR and ULR, respectively. The higher level of nucleotide diversity suggests that *repleta* family elements may represent a more

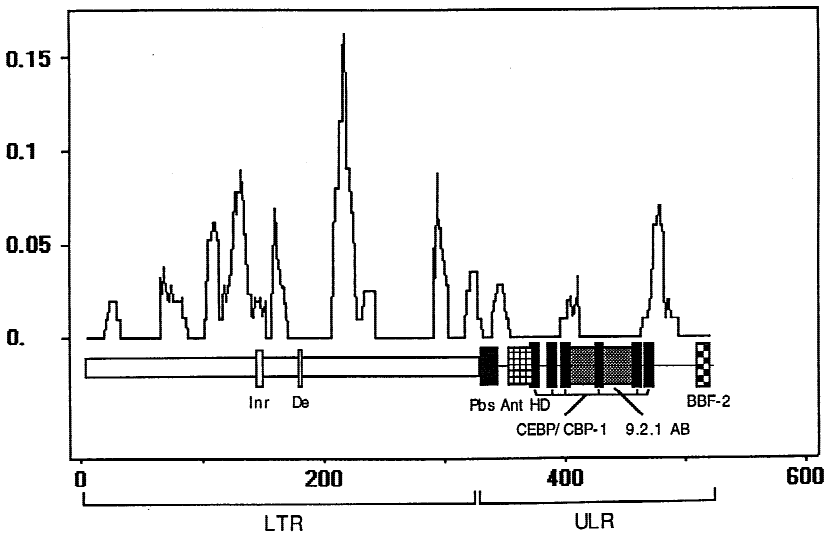

Fig. 2. A sliding window was used to visualize levels of nucleotide diversity (π) along the *copia* LTR-ULR regulatory region of *melanogaster* family elements. The location and identity of *copia* cis-regulatory sequences are shown relative to the level of nucleotide diversity (π).

Table 2. Variation in the *copia* LTR-ULR regulatory region

	LTR	ULR
Polymorphism		
melanogaster family (n = 26)		
Segregating sites (S)	22	7
Nucleotide diversity (π)[a]	0.02206	0.00697
Length polymorphism		
<10	14	12
>10	4	1
repleta family (n = 3)		
Segregating sites (S)	30	9
Nucleotide diversity (π)[a]	0.09351	0.04255
Length polymorphism		
<10	3	2
>10	0	0
Fixed differences		
Segregating sites (S)	90	45
Length polymorphism		
<10	2	1
>10	1	0
Total nucleotide diversity (π)[a]	0.14980	

[a] Nucleotide diversity (π) was calculated as described under Materials and Methods.

ancestral group. The level of LTR-ULR variation of *melanogaster* family *copia* elements is quite low. This is despite the fact that previous surveys have demonstrated that this is the most rapidly evolving region of LTR retrotransposons (Kulguskin et al. 1981; Lankenau et al. 1990; Lyubomirskaya et al. 1990; Mizrokhi and Mazo 1991). In both *copia* families the LTR has higher levels of nucleotide diversity than the ULR. The LTRs are 3.2× more diverse in the *melanogaster* and 2.2× more diverse in the *repleta* families than the ULRs. This suggests that a higher degree of selective constraint has been imposed on the ULR than the LTR due perhaps to the higher density of regulatory sites in the ULR.

In order to determine definitively if *copia* LTR-ULR binding site sequences are more conserved than nonbinding site sequences, levels of nucleotide diversity for each binding site were compared with the overall LTR-ULR nucleotide diversity among *melanogaster* family elements. In 10 of the 12 binding sites examined, levels of binding site nucleotide diversity are significantly lower ($P < 0.01$, t test) than the overall LTR-ULR nucleotide diversity. Only the initiator in the LTR and one of the CEBP/CBP-1 binding sites in the ULR did not show significantly lower levels of nucleotide diversity. These data indicate that these particular sites are less constrained and may not be as critical for *copia* regulation.

Size Variation

The *copia* LTR-ULR shows a great deal of naturally occurring size variation (Table 2). While the *melanogaster* family elements are more conserved at the nucleotide level, they show more size variation than *repleta* family elements thus far sequenced. The majority of the *melanogaster* family size variation consists of small gaps (<10 bp) due to changes in the number of homonucleotide repeats; however, there are also large gaps (>10 bp) present in both the LTR and the ULR. Indeed all *melanogaster* family elements can be grouped into discrete classes (full-length, ULR gap, and double-gap variants) based on the presence or absence of identically located gaps in their LTRs and/or ULRs (Csink and McDonald 1995; Matyunina et al. 1996). Comparison of the sequences of the gap variants indicates that full-length variants arose from the smaller gap variants by a series of regional duplications (Matyunina et al. 1996; McDonald et al. 1997). Independent phylogenetic evidence, which indicates that the discrete classes of gap variants represent *copia* subfamilies, is consistent with this hypothesis (Jordan and McDonald 1998). Interestingly much of the

Table 3. Tajima's (1989) and Fu and Li's (1993) tests of neutrality

Neutrality test[a]	Regulatory region	Test statistic value[a]	Significance level
Tajima	LTR-ULR	$D = -1.72570$	$0.10 > P > 0.05$
	LTR	$D = -1.52898$	$P > 0.10$
	ULR	$D = -1.89882$	$P < 0.05$
Fu and Li	LTR-ULR	$D = -2.50960$	$0.10 > P > 0.05$
		$F = -2.65722$	$P < 0.05$
	LTR	$D = -2.13450$	$0.10 > P > 0.05$
		$F = -2.28208$	$0.10 > P > 0.05$
	ULR	$D = -2.71761$	$P < 0.05$
		$F = -2.88402$	$P < 0.05$

[a] Test statistics were calculated as described under Materials and Methods.

ULR size variation maps to regions which contain regulatory binding sites, and this variation is known to be of functional significance (Matyunina et al. 1996).

Tests of Neutrality

In order to evaluate effects of selection versus neutrality on *copia* LTR-ULR sequence variation, we performed the Tajima (1989) and Fu and Li (1993) tests of neutrality, which compare different estimates of θ, the neutral mutation parameter. Using each of these tests, we compared the patterns of nucleotide variation among *melanogaster* family *copia* sequences over the entire *copia* LTR-ULR as well as within the LTR and ULR regions individually (Table 3). For each test, a negative test statistic value was obtained, which is indicative of purifying selection. However, in most cases only the ULR showed a significant deviation ($P < 0.05$) from neutrality. These results are consistent with the lower levels of nucleotide diversity in the ULR. The presence of a high density of *copia* regulatory binding sites in the ULR likely imparts a more stringent degree of functional constraint in this region than in the LTR. The results of these statistical tests taken together with the lower levels of nucleotide variation in the regulatory sites indicate that the *copia* LTR-ULR regulatory region has been subject to selective constraint.

We also looked for evidence that positive selection might have played a role in establishing the patterns of variation which exist between families of *copia* elements. We employed a modified version of the McDonald–Kreitman (1991) test to compare the ratio of binding site (replacement) to non-binding site (silent) nucleotide differences within (polymorphic) and between (fixed) *copia* families. For the two *copia* families, the ratio of polymorphic binding site to non-binding site substitutions is significantly different ($P < 0.05$, 2×2 *G* test of independence) than the ratio of fixed binding site to non-binding site substitutions. The significance of this result is due to a relative excess of fixed binding site changes

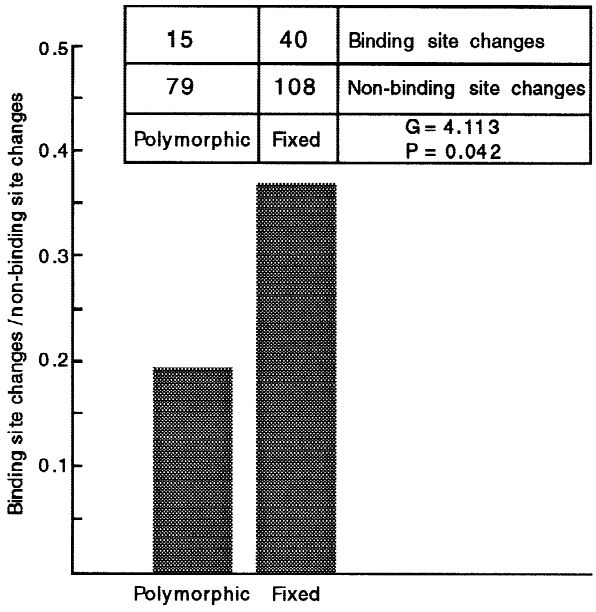

Fig. 3. A *G* test with Williams' correction was performed on the 2×2 contingency table comparing the ratio of binding site to non-binding site LTR-ULR nucleotide changes within (polymorphic) and between (fixed) *copia* families. The deviation from neutrality is due to a relative excess of fixed binding site changes.

(Fig. 3). This pattern of variation indicates that the departure from neutrality is due to adaptive nucleotide substitutions (McDonald and Kreitman 1991; Brookfield and Sharp 1994) between the LTR-ULR regulatory regions of the two *copia* families. Therefore, positive diversifying selection appears to have contributed to the establishment of the major families of *copia* elements, while negative purifying selection appears to have been a factor in maintaining the nucleotide integrity of regulatory sequences within *copia* families.

Discussion

The results of our analysis of naturally occurring *copia* LTR-ULR nucleotide variation indicates that both negative purifying selection and positive diversifying selection have played a role in the evolution of *copia* LTR-ULR regulatory regions. The negative values of the test statistics comparing different estimates of θ, taken together with the low levels of nucleotide variation associated with these regulatory regions, indicate that *copia* LTR-ULR sequence evolution has been selectively constrained within the *melanogaster* family of *copia* elements. Additionally, the results of the modified McDonald–Kreitman test indicate that positive diversifying selection has played a significant role in establishing the LTR-ULR nucleotide differences which characterize the *melanogaster* and *repleta* families of *copia* elements.

Selection on *copia* sequences is most likely occurring at the genomic level rather than at the organismic level,

since it is unlikely that *copia* elements provide a selective advantage to their hosts. In fact, the replicative transposition life cycle of LTR retrotransposons clearly exacts a cost on host genomes as is evidenced by the fact that >50% of all spontaneous mutations in *Drosophila* having a significant phenotypic effect are associated with LTR retrotransposons (Berg and Howe 1989). *Copia* elements, like other transposable elements, are believed to maintain themselves despite this cost to their hosts, due to their ability to replicate faster than their hosts through transposition (Hickey 1982). Elements which transpose most efficiently or at the highest rate are most likely to survive and propagate (Deininger 1992). This situation leads to competition between elements within genomes and the phenomenon of interelement selection (McDonald et al. 1997).

Since, in most circumstances, the adaptive interests of the host and element may be in conflict, the dynamics of host–element coevolution may be heuristically viewed as an ''arms race,'' where host genomes are continually evolving mechanisms to repress the replicative transposition of the elements, while the elements are under continual selective pressure to escape repressive controls. An evolved cis regulatory sequence(s) which allows an LTR retrotransposon successfully to transcribe and replicate within a host genome would be expected to be selectively maintained until such time as a successful counter response has been evolved by the host. Our results suggest that this is the current state of affairs for the *melanogaster* family of *copia* elements. It is important to keep in mind, however, that the relationship between LTR retrotransposons and their host genomes is a dynamic one and that what is observed for a particular family of elements at a particular point in evolutionary time cannot necessarily be extrapolated to other host–element systems or even to the same host–element system at other stages in its evolutionary history.

Previous studies have demonstrated that functionally significant *copia* LTR-ULR regulatory sequence variation exists both within and between *Drosophila* species (Matyunina et al. 1996). Thus the opportunity clearly exists for positive interelement selection on *copia* LTR-ULR regulatory sequences. Our results indicate that such positive selection contributed to the LTR-ULR nucleotide divergence which exists between the *melanogaster* and the *repleta* families of *copia* elements. Whether this selective divergence occurred gradually as the two host genome lineages diverged or was more sudden, as might have been necessitated, for example, by a horizontal transfer event, is presently unknown.

Transposable elements and LTR retrotransposons are intimately associated with the host genomes in which they reside. As we learn more about the molecular variation of transposable elements and the nature of the molecular interactions which exist between these elements and their host genomes, we will be better able to reconstruct the processes which underlie host–element coevolution.

References

Arkhipova IR, Lyubomirskaya NV, Ilyin YV (1995) Drosophila retrotransposons. R.G. Landes, Austin, TX

Berg DE, Howe MM (eds) (1989) Mobile DNA. American Society for Microbiology, Washington, DC

Biessmann H, Valgeirsdottir K, Lofsky A, Chin C, Ginther B, Levis RW, Pardue ML (1992) HeT-A, a transposable element specifically involved in ''healing'' broken chromsome ends in Drosophila melanogaster. Mol Cell Biol 12:3910–3918

Boeke JD, Garfinkel DJ, Styles CA, Fink GR (1985) Ty elements transpose through an RNA intermediate. Cell 40:491–500

Brookfield JF, Sharp PM (1994) Neutralism and selectionism face up to DNA data. Trends Genet 10:109–111

Capy P (ed) (1998) Evolution and impact of transposable elements. Kluwer, Dordrecht, The Netherlands

Cavarec L, Heidmann T (1993) The Drosophila *copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. Nucleic Acids Res 21:5041–5049

Cavarec L, Jensen S, Heidmann T (1994) Identification of a strong transcriptional activator for the *copia* retrotransposon respponsible for its differential expression in *Drosophila hydei* and *melanogaster* cell lines. Biochem Biophys Res Commun 203:392–399

Cavarec L, Jensen S, Casella JF, Cristescu SA, Heidmann T (1997) Molecular cloning and characterization of a transcription factor for the *copia* retrotransposon with homology to the BTB-containing lola neurogenic factor. Mol Cell Biol 17:482–494

Csink AK, McDonald JF (1995) Analysis of *copia* sequence variation within and between Drosophila species. Mol Biol Evol 12:83–93

Deininger PL, Batzer MA, Hutchison CA 3rd, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311

Flavell AJ, Levis R, Simon MA, Rubin GM (1981) The 5′ termini of RNAs encoded by the transposable element *copia:* Nucleic Acids Res 9:6279–6291

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Hickey DA (1982) Selfish DNA: A sexually-transmitted nuclear parasite. Genetics 101:519–531

Hu WS, Temin HM (1990) Retroviral recombination and reverse transcription. Science 250:1227–1233

Hudson RR (1993) Levels of DNA polymorphism and divergence yield important insights into evolutionary processes. Proc Natl Acad Sci USA 90:7425–7426

Jenkins DL, Ortori CA, Brookfield JF (1995) A test for adaptive change in DNA sequences controlling transcription. Proc R Soc Lond B Biol Sci 261:203–207

Jordan IK, McDonald JF (1998) Evolution of the *copia* retrotransposon in the *Drosophila melanogaster* subgroup. Mol Biol Evol 15:1160–1171

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Monro HN (ed) Mammalian protein evolution. Academic Press, New York, pp 21–132

Kikuchi Y, Ando Y, Shiba T (1986) Unusual priming mechanism of RNA-directed DNA synthesis in copia retrovirus-like particles of Drosophila. Nature 323:824–826

Kimura, M. (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kulguskin VV, Ilyin YV, Georgiev GP (1981) Mobile dispersed genetic element MDG1 of Drosophila melanogaster: Nucleotide sequence of long terminal repeats. Nucleic Acids Res 9:3451–3464

Lankenau DH, Huijser P, Jansen E, Miedema K, Hennig W (1990) DNA sequence comparison of micropia transposable elements from

Drosophila hydei and Drosophila melanogaster. Chromosoma 99: 111–117

Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen FM (1993) Transposons in place of telomeric repeats at a Drosophila telomere. Cell 75:1083–1093

Li W (1997) Molecular Evolution. Sinauer, Sunderland, MA

Ludwig MZ, Kreitman M (1995) Evolutionary dynamics of the enhancer region of even-skipped in Drosophila. Mol Biol Evol 12: 1002–1011

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7:377–394

Lyubomirskaya NV, Arkhipova IR, Ilyin YV. Kim AI (1990) Molecular analysis of the gypsy (mdg4) retrotransposon in two *Drosophila melanogaster* strains differing by genetic instability. Mol Gen Genet 223:305–309

Matyunina LV, Jordan IK, McDonald JF (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

McDonald JF (1993) Evolution and consequences of transposable elements. Curr Opin Genet Dev 3:855–864

McDonald JF (1995) transposable elements: Possible catalysts of organsimic evolution. Trends Ecol Evol 10:123–126

McDonald JF (1998) Transposable elements, gene silencing and macroevolution. Trends Ecol Evol 13:94–95

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654

McDonald JF, Matyunina LV, Wilson SW, Jordan IK, Bowen NJ, Miller WJ (1997) LTR retrotransposons and the evolution of eukaryotic enhancers. Genetica 100:3–13

Miller WJ, Kruckenhauser I, Pinsker W (1996) The impact of transposable elements on genome evolution in animals and plants. In: Tomiuk J, Woerhm K, Sentker A (eds) Transgenic organisms—Biological and social implications. Birkhauser, Verlag Basel, pp 21–34

Mizrokhi LJ, Mazo AM (1991) Cloning and analysis of the mobile element gypsy from *D. virilis.* Nucleic Acids Res 19:913–916

Mount SM, Rubin GM (1985) Complete nucleotide sequence of the Drosophila transposable element copia: Homology between *copia* and retroviral proteins. Mol Cell Biol 5:1630–1638

Rozas J, Rozas R (1997) DnaSP version 2.0: A novel software package for extensive molecular population genetics analysis. Comput Appl Biosci 13:307–311

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Sneddon A, Flavell AJ (1989) The transcriptional control regions of the *copia* retrotransposon. Nucleic Acids Res 17:4025–4035

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Wayne ML, Simonsen KL (1998) Statistical tests of neutrality in the age of weak selection. Trends Ecol Evol 13:236–240

Wessler SR, Bureau Te, White SE (1995) LTR-retrotransposons and MITEs: Important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal plant genes: A role for *copia*-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91:11792–11796

Wilson SW, Matyunina LV, McDonald JF (1998) An enhancer region within the *copia* untranslated leader contains binding sites for *Drosophila* regulatory proteins. Gene 209:239–246