Genet. Sel. Evol. 38 (2006) 535–550

DOI: 10.1051/gse:2006020

Original article

Sequence heterogeneity and phylogenetic relationships between the *copia* retrotransposon in *Drosophila* species of the *repleta* and *melanogaster* groups

Luciane M. De Almeida[a], Claudia M.A. Carareto[a]*

Universidade Estadual Paulista (UNESP), Departamento de Biologia, 15054-000 São José do Rio Preto, SP, Brazil

(Received 2 February 2006; accepted 19 April 2006)

Abstract – Although the retrotransposon *copia* has been studied in the *melanogaster* group of *Drosophila* species, very little is known about *copia* dynamism and evolution in other groups. We analyzed the occurrence and heterogeneity of the *copia* 5'LTR-ULR partial sequence and their phylogenetic relationships in 24 species of the *repleta* group of *Drosophila*. PCR showed that *copia* occurs in 18 out of the 24 species evaluated. Sequencing was possible in only eight species. The sequences showed a low nucleotide diversity, which suggests selective constraints maintaining this regulatory region over evolutionary time. On the contrary, the low nucleotide divergence and the phylogenetic relationships between the *D. willistoni* / *Zaprionus tuberculatus* / *melanogaster* species subgroup suggest horizontal transfer. Sixteen transcription factor binding sites were identified in the LTR-ULR *repleta* and *melanogaster* consensus sequences. However, these motifs are not homologous, neither according to their position in the LTR-ULR sequences, nor according to their sequences. Taken together, the low motif homologies, the phylogenetic relationship and the great nucleotide divergence between the *melanogaster* and *repleta copia* sequences reinforce the hypothesis that there are two *copia* families.

***copia* retrotransposon / nucleotide diversity / *copia* families**

1. INTRODUCTION

Ty1-copia is present as a highly heterogeneous group of retrotransposons within all higher eukaryotes [10, 21]. The *Drosophila* retrotransposon *copia*, which is structurally similar to retroviral proviruses, is 5.4 kb in length and flanked by 276-bp direct long terminal repeats (LTR). According to the review of Biémont and Cizeron [2], *copia* sequences were identified by PCR and

* Corresponding author: carareto@ibilce.unesp.br

Southern blot analysis in 52 different species of the genus *Drosophila*. Twenty-two species out of this total belong to the *melanogaster* group, seven to the *willistoni* group, seven to the *obscura* group, six to the *saltans* group, two to the *immigrans* group, one to the *mesophragmatica* group, and one to the *pinicola* group. Although the retrotransposon *copia* is harbored by the genome of these 52 species, nucleotide sequences have been described only for eight species of the *melanogaster* group, two species of the *repleta* group, *D. willistoni* and *Zaprionus tuberculatus*. *Drosophila copia* phylogeny studies are difficult to carry out, not only because of the low number of *copia* sequences in the group, but also because these sequences are partial, most of them concerning 5' long terminal repeats (LTR) and untranslated leader regions (ULR).

The 5' LTR-ULR contains sequences responsible for controlling *copia* transcription, which is a rate-limiting step in the retrotransposition process [3]. The 5' LTR contains promoter sequences and the transcription start site [19, 20]. The ULR contains several repeated sequence motifs which function as enhancers [9, 26, 34, 35, 45, 49]. These repeat motifs are binding sites for host regulatory proteins, and the strength of an enhancer is often positively correlated to the number of repeat motifs it contains [42]. Because of the functional importance of these regulatory sequences, the noncoding LTR-ULR sequences have been commonly used in phylogenetic studies [14, 25] and in retrotransposon regulation studies [9, 10, 26, 34, 35, 49].

The retrotransposon *copia* has been intensively studied in the *melanogaster* group of *Drosophila* species and proven to be a good model system for studying regulatory interactions between retrotransposons and their host genomes [8, 10, 20, 26, 34]. However, very little is known about *copia* dynamism and evolution in other species of the genus *Drosophila.* In order to broaden our knowledge about the evolutionary history and dynamism of this element, we analyzed the occurrence and heterogeneity of the *copia* 5'LTR-ULR partial sequence and their phylogenetic relationships in 24 species of three subgroups of the *Drosophila repleta* group and their relationships with all the corresponding *copia* sequences of Drosophilidae found in GenBank.

2. MATERIALS AND METHODS

2.1. Fly stocks

All species and strains (isofemales) used in this study are listed in Table I. Each list includes the taxonomic nomenclature [18], location, and either stock number or collection date.

Table I. Taxonomic list of *Drosophila* species of the *repleta* group used in this study (according to Durando *et al.,* 2000).

Subgroup	Cluster	Species	Location	Stock number	Date
mercatorum		*D. paranaensis*	Novo Horizonte, SP, Brazil		1998
		D. mercatorum	Novo Horizonte, SP, Brazil		1998
		D. aldrichi	Tamaulipas, Mexico	15081125.0	
	mulleri	*D. mulleri*	Tucson, Arizona, USA	150811371.7	
		D. wheeleri	Catalina Island, Los Angeles, USA	150811501.3	
		D. navojoa	Tehuantepec, Mexico	150811374.1	
	mojavensis	*D. arizonae*	Tucson, Arizona, USA	A 1015	
mulleri		*D. mojavensis*	Grand Canyon, Arizona, USA	A870	
		D. koepferae	Tapia-Tucumam, Argentina		1990
		D. serido	Milagres, BA, Brasil	150811431.3	
	buzzatii	*D. buzzatii*	Novo Horizonte, SP, Brazil		1998
		D. gouveai	Morro do Chapéu, MG, Brazil		
		D. antonietae	Bela Vista, MS, Brazil		1990
		D. seriema	Serra do Cipó, MG, Brazil		1990
		D. longicornis	Tucson, Arizona, USA	150811311.8	
		D. pachuca	Chapingo, Mexico	150811391.0	
	longicornis	*D. propachuca*	Hidalgo, Mexico	150811411.1	
		D. hexastigma	Zapotitlan, Puebla, Mexico	150811302.2	
		D. spenceri	Gyaymas, Sonora, Mexico	150811441.2	
	eremophila	*D. eremophila*	Zapotitlan, Puebla, Mexico	150811292.1	
		D. mettleri	Tucson, Arizona, USA	CAT397	
		D. anceps	Infernillo, Michoacán, Mexico	150815031.1	
		D. ritae	Zapotitlan dos Salinas, Puebla, Mexico		2002
hydei		*D. hydei*	Encarnación, Paraguay		2000

2.2. DNA amplification and sequencing

The regions of *copia* focused on in this study were the 5' LTR and the untranslated leader region (ULR). PCR reactions were performed in final volumes of 25 μL, using approximately 200 ng of template DNA, 100 μM of each primer, 200 μM of dNTP, 1.5 mM of $MgCl_2$, 1.25 μL of DMSO and 1 unit of TaqBead Hot Start Polymerase (Promega) in 1× polymerase buffer. After an initial denaturation step of 5 min at 94 °C, 35 cycles consisting of 1 min at 94 °C, 1 min at 55 °C, and 1 min at 72 °C were carried out, followed by a final extension step of 15 min at 72 °C. The primers used were the following: CoBuz1 (5'-CCCNTATTCCTCCTTCAAAAA-3') and CoBuz2 (5'-CCGCGAAATTAAGAAACGAG-3'), which anneal into the LTR-URL *copia* region and amplify a 615 bp long fragment (nucleotides 10

to 625). These primers were designed based on the *D. buzzatii* (X96972) and *D. koepferae* (X96971) *copia* sequences obtained from GenBank, which contain a polymorphism between both species in the fourth position of CoBuz1. It is important to point out that the region amplified by the *primers* CoBuz1 and CoBuz2 corresponds to the 5' LTR-ULR region studied by Jordan and McDonald [26]. The amplified fragments were separated by electrophoresis in 1% agarose gel. The PCR products were cloned into a TA cloning vector (Invitrogen). Both strands of three clones chosen randomly were sequenced for each species, and the consensus sequence was used for the phylogenetic analysis. The *copia* sequences obtained in the present study were deposited in NCBI GenBank (accession numbers from AY655745 to AY655750 and DQ494345 and DQ494346).

2.3. Evolutionary analysis

The multiple alignments of *copia* consensus sequences were performed with CLUSTAL W [47]. The evolutionary relationships among *copia* sequences were assessed using the maximum parsimony method (branch and bound algorithm), as implemented in PAUP v.4.0b10 [46]. The distance matrix used was built according to the HKY model [23], which was determined as the best fit for the data by a likelihood ratio test using MODELTEST 2.0 [38]. The *copia* sequences of *D. melanogaster* (X02599), *D. yakuba* (AF063885), *D. teissieri* (AF063883), *D. sechellia* (AF063879), *D. simulans* (AF063882), *D. mauritiana* (AF063872), *D. erecta* (AF063870), *D. willistoni* (AF175766), *D. buzzatii* (AF063868) and *D. koepferae* (X96971) were obtained from GenBank. The *Z. tuberculatus* sequence was obtained from McDonald *et al.* [35].

2.4. Neutrality tests

The levels of nucleotide diversity and number of segregating sites were determined for the copia LTR and ULR regions using the DnaSP program [40], in order to evaluate whether the sequences evolve randomly or have been subjected to functional constraints.

2.5. Identification of transcription factor binding sites

The *copia* 5' LTR-ULR sequence in the *melanogaster* group has been shown to contain some motifs, which are binding sites to trans regulatory proteins and

regulate their transcriptional activity [9,10,34,35,49]. Variants of 5' LTR-ULR with different numbers of motifs differ in their abilities to drive expression [34]. In order to identify LTR-ULR binding sites of transcriptional factors, the *repleta* and *melanogaster* consensus sequences were submitted to the Alibaba2 software [22] (http://www.gene-regulation.com/pub/programs.html) that is currently considered as the most effective tool for predicting transcription factor binding sites in an unknown DNA sequence.

3. RESULTS

3.1. Distribution of *copia* among species

PCR analysis showed *copia* sequences of the expected size (615 bp) in 18 of the 24 *repleta* species (Fig. 1). Strong amplification was observed in species belonging to the *buzzatii* (*D. koepferae, D. buzzatii, D. serido, D. gouveai, D. antonietae* and *D. seriema*), *mulleri* (*D. aldrichi, D. mulleri* and *D. wheeleri*) and *longicornis* (*D. pachuca, D. propachuca, D. hexastigma* and *D. spenceri*) clusters. Only faint bands were observed in *D. hydei, D. paranaensis, D. mercatorum, D. navojoa,* and *D. mojavensis*. Even after several repetitions, no amplification was observed in *D. arizonae, D. longicornis, D. eremophila, D. mettleri, D. anceps* and *D. ritae*.

3.2. Evolutionary analysis

The aim of the sequencing analysis was to investigate the natural *copia* LTR-URL nucleotide variation and to propose a phylogenetic relationship between these sequences and those described in the literature. The most unrooted parsimonious tree is shown in Figure 2. From a total of 742 characters, 484 were phylogenetically informative. The consistency index was 0.7800 and the retention index was 0.9156. Branch support was calculated by bootstrap analysis consisting of 1000 replicates. Two well-defined groups of *copia* sequences can be seen in the tree, one containing the species of the *repleta* group (*D. serido*, *D. buzzatii*, *D. mojavensis*, *D. gouveai*, *D. seriema*, *D. koepferae*, *D. antonietae* and *D. pachuca*) and the other containing *Z. tuberculatus*, *D. willistoni* and the species of the *melanogaster* group (*D. sechellia*, *D. erecta*, *D. mauritiana*, *D. simulans*, *D. yakuba*, *D. melanogaster* and *D. teissieri*). It is worth pointing out that the *copia* tree does not fit the host phylogeny, since *Z. tuberculatus*, *D. willistoni*, *D. melanogaster, D. teissieri, D. yakuba* and *D. simulans* constitute a monophyletic group of sequences clearly separated

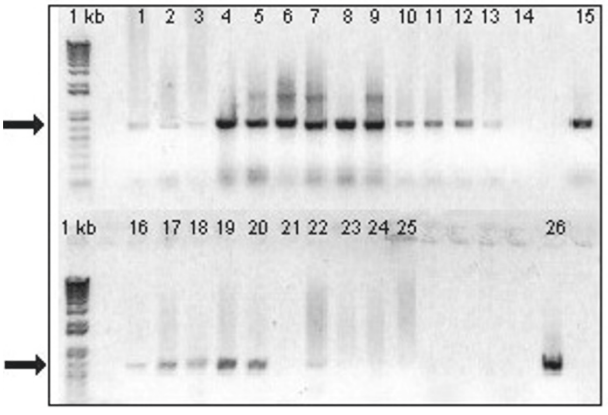

Figure 1. Ethidium bromide-stained agarose gel of PCR products obtained with *copia*-specific primers (CoBuz1 and CoBuz2) used as template genomic DNA from the following species (**A):** 1. *D. hydei*; 2. *D. paranaensis*; 3. *D. mercatorum*; 4. *D. koepferae*; 5. *D. buzzatii*; 6. *D. serido*; 7. *D. gouveai*; 8. *D. antonietae*; 9. *D. seriema*; 10. *D. aldrichi*; 11. *D. mulleri*; 12. *D. wheeleri*; 13. *D. navojoa*; 14. *D. arizonae*; 16. *D. mojavensis*; 17. *D. pachuca*; 18. *D. propachuca*; 19. *D. hexastigma;* 20. *D. spenceri*; 21. *D. longicornis*; 22. *D. eremophila*; 23. *D. mettleri*; 24. *D. anceps*, 25. *D. ritae* and 15 and 26 PTZ18 plasmid containing *copia* element of *D. koepeferae* (positive control). The arrows point to the expected 615 bp fragments.

from the other sequences of the *melanogaster* species group. Also, the *copia* sequences of *D. mojavensis* (cluster *mojavensis*) and *D. pachuca* (cluster *longicornis*) constitute a monophyletic group with species of the *buzzatii* cluster, the first with *D. buzzatii* and the second with *D. koepferae* and *D. antonietae*.

The distance matrix is shown in Table II. The smallest divergence within the *repleta* group was 0.01 (*D. koepferae* and *D. pachuca*) and the greatest was 0.18 (between *D. koepferae* and *D. mojavensis*). In the species of the *melanogaster* group, the values varied from 0.00 (between *D. teissieri* and *D. melanogaster*) to 0.06 (between *D. sechellia* and *D. melanogaster*; *D. teissieri* and *D. sechellia*). In spite of the low divergence rates within each species group, the intergroup rates between the *melanogaster* and *repleta* groups were very high. The smallest divergence was 0.61 (between *D. yakuba* and *D. seriema*; *D. yakuba* and *D. gouveai*) and the greatest was 0.82 (between *D. koepferae* and *D. sechellia*). On the contrary to these high rates, the nucleotide divergence between the *melanogaster* subgroup and *D. willistoni* and *Z. tuberculatus* was very small: 0.00 between *D. willistoni* and *D. teissieri*, and 0.02 between *Z. tuberculatus* and *D. yakuba*. A low rate was also observed between *Z. tuberculatus* and *D. willistoni* (0.03). These results, in agreement

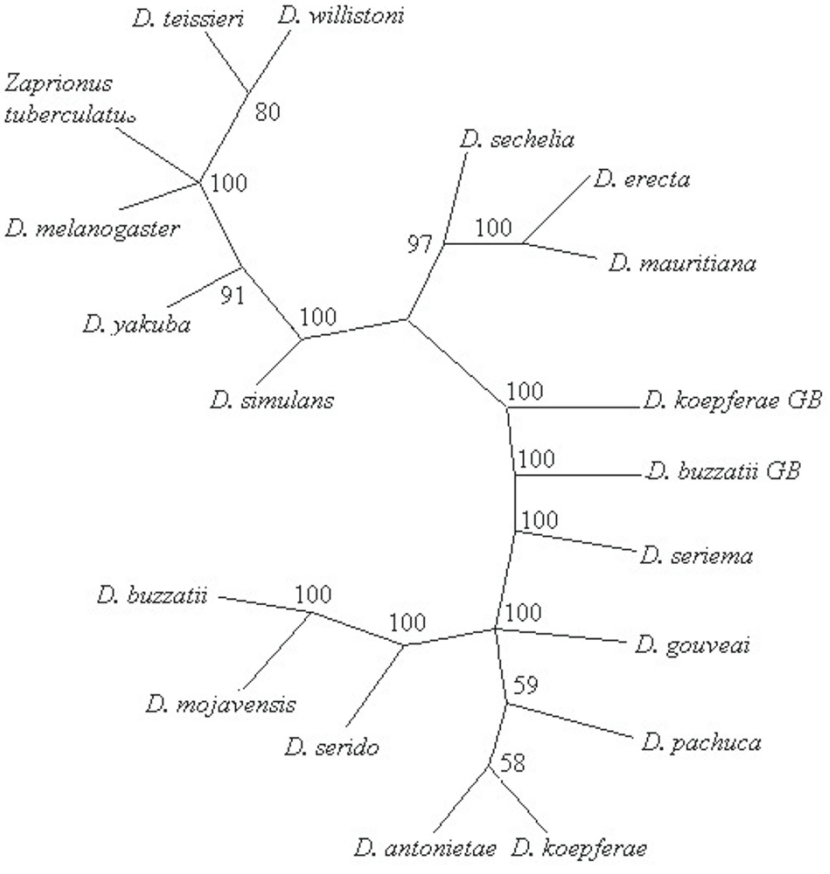

Figure 2. Phylogenetic analysis of LTR-ULR *copia* nucleotide sequences. The cladogram was generated by parsimony analysis as implemented by PAUP 4.0 b10 (Swofford, 2000). The consistency index is 0.7800 and the retention index is 0.9156. Branch support was calculated by bootstrap analysis consisting of 1000 replicates.

with the phylogenetic analysis, indicate the occurrence of two significantly divergent *copia* groups of sequences, one carried by genomes of the species of the *melanogaster* group / *D. willistoni* / *Zaprionus tuberculatus* and the other by the *repleta* species group.

3.3. *Neutrality tests*

The nucleotide diversity and the number of segregating sites were calculated for the six species of the *buzzatii* cluster, in order to determine whether

Table II. Genetic distances between *copia* nucleotide sequences calculated using the HKY method (Hasegawa *et al.*, 1985). GB: GenBank.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1 *Z. tuberculatus*	-																	
2 *D. willistoni*	0.03	-																
3 *D. melanogaster*	0.03	0.00	-															
4 *D. simulans*	0.05	0.04	0.04	-														
5 *D. mauritiana*	0.05	0.04	0.04	0.04	-													
6 *D. erecta*	0.05	0.04	0.04	0.04	0.01	-												
7 *D. sechellia*	0.07	0.06	0.06	0.05	0.02	0.01	-											
8 *D. yakuba*	0.02	0.01	0.01	0.04	0.04	0.04	0.05	-										
9 *D. teissieri*	0.03	0.00	0.00	0.04	0.04	0.04	0.06	0.01	-									
10 *D. buzzatii* (GB)	0.67	0.66	0.66	0.67	0.66	0.67	0.70	0.63	0.66	-								
11 *D. buzzatii*	0.69	0.52	0.67	0.67	0.67	0.67	0.67	0.64	0.68	0.05	-							
12 *D. koepferae* (GB)	0.78	0.73	0.74	0.79	0.78	0.81	0.82	0.74	0.75	0.12	0.15	-						
13 *D. koepferae*	0.67	0.52	0.65	0.65	0.65	0.67	0.68	0.62	0.66	0.09	0.07	0.01	-					
14 *D. serido*	0.69	0.54	0.67	0.68	0.68	0.68	0.71	0.65	0.68	0.05	0.06	0.12	0.07	-				
15 *D. gouveai*	0.66	0.52	0.64	0.63	0.64	0.65	0.66	0.61	0.64	0.08	0.07	0.04	0.02	0.07	-			
16 *D. antonietae*	0.67	0.52	0.65	0.65	0.65	0.67	0.68	0.62	0.67	0.08	0.06	0.10	0.11	0.06	0.02	-		
17 *D. seriema*	0.65	0.56	0.63	0.63	0.64	0.65	0.65	0.61	0.64	0.08	0.08	0.04	0.03	0.09	0.01	0.02	-	
18 *D. pachuca*	0.69	0.61	0.67	0.67	0.67	0.69	0.69	0.64	0.68	0.09	0.08	0.02	0.01	0.08	0.02	0.02	0.03	-
19 *D. mojavensis*	0.71	0.57	0.69	0.70	0.69	0.69	0.70	0.66	0.69	0.05	0.02	0.18	0.08	0.06	0.08	0.07	0.09	0.11

selective constraints have been imposed upon the *copia* LTR and ULR sequences. Our data showed that the polymorphism is 0.06718 within the LTR, with 45 segregating sites, and 0.01979 within the ULR, with 18 segregating sites. Hence, the ULR conservation is 3.4 times bigger than that of the LTR, which suggests that some degree of selective constraint has been imposed upon the ULR compared to the LTR.

3.4. Identification of transcription factor binding sites

The repeated motifs within enhancers are usually binding sites for host factors, which regulate element expression. We identified the motifs of the *copia* ULR sequences present in the *melanogaster* and the *repleta* group using the Alibaba 2 program [22]. Table III shows the identified motifs, the sequence of each motif and the number of repetitions of each one. The most frequently found motif was the CCAAT/enhancer binding protein (C/EBP is involved in the control of head segmentation in *Drosophila*). This motif presents 13 repetitions in the *repleta* species group and 11 in the *melanogaster* species group. It has been shown that the number of C/EBP repetitions is responsible for different levels of *copia* expression in *D. melanogaster* [27]. Motifs such as HB (hunchback factor), TBP (TATA-binding protein), Oct-1 (octamer-binding factor), FTZ (fushi tarazu) and Zen (Zerknuellt 1) were found in both species groups. On the contrary, some motifs had their occurrence limited to just one group of species. The motifs E47 (hairy); CFF (complex forming factor) and Embry (embryo DNA binding protein) were present only in the *melanogaster* species group, while Kr (Krueppel), Odd (Odd-skipped), NF-kappa; Ant (antennapedia) and Oct-2.1 (octamer-binding factor) were present only in the *repleta* species group.

4. DISCUSSION

Transposable elements can be classified, according to their structure, into classes, subclasses, families, and subfamilies. Jordan and McDonald [25, 26] suggested that there were two *copia* families within the genus *Drosophila*: the *melanogaster* family with three subfamilies, and the *repleta* family (based on the analysis of only two species of the *repleta* group). Looking for other families and subfamilies within the *repleta* group, we increased the number of analyzed species to 24 (one species belonging to the *hydei* subgroup, two to the *mercatorum* subgroup, and 21 to the *mulleri* subgroup), but the *copia* sequences could only be obtained from eight species. The greatest HKY distance

Table III. Transcription-factor binding sites in the *repleta* and *melanogaster copia* consensus-sequence according to Alibaba2 (Grabe, 2002). **Motifs**: the factor name and the consensus sequence identified in TRANSFAC database; ***repleta* group**: the sequences identified as motifs in the *copia repleta* sequence; **nt position**: the position of each motif in the *copia* sequence; ***melanogaster* group**: the sequences identified as motifs in the *copia melanogaster* sequence.

Motifs (Consensus)	*repleta* group sequences	nt position	Motifs (Consensus)	*melanogaster* group sequences	nt position
TBP (TATA-binding protein)			TBP		
(TCCTTmwwnA)	**TCCTTCAAAA**	10-19	(yTTTATAnny)	**GTTTATATTT**	37-46
(wATTTArGnr)	**AATTTAAGCG**	584-593			
Hb (hunchback factor)			Hb		
(nrTTTTTTnk)	**AGTTTTTTGT**	329-338	(symAwAAAAm)	**GCCATAAAAC**	70-80
			(wTwnATTAwA)	**ATAAATTAAA**	244-253
			(mAAwmAAyTr)	**AAATAAATTG**	322-331
E47 (Hairy)	ABSENT		E47		
			(ssnGCwGGTG)	**GGTGCAGGTG**	152-161
CFF (complex forming factor)	ABSENT		CFF		
			(wTArmTTTwAA)	**ATAACTTAAA**	174-183
Oct-1 (octamer-binding factor)			Oct-1A		
(mwATrymAAT)	**AAATACCAAT**	170-179	(wwwnTAAAAC)	**AATATAAAAC**	183-192
(ykwCATwTwA)	**TTACATTTAA**	389-399			
Oct-2.1 (octamer-binding factor)			Oct-2.1	ABSENT	
(TmATkrrCAT)	**TAATGAGCAT**	54-63			
Embry (embryo DNA binding protein)	ABSENT		Embry		
			(CAATTArmyw)	**CAATTAAATT**	294-303
FTZ (fushi tarazu)			FTZ		
(AAwTAnyAw)	**TCTTAATTTC**	601-610	(AAwTAnyAwT)	**AAATAGCATT**	334-343
Zen-1 (Zerknuellt 1)			Zen-1		
(nykwCATTTA)	**ATTACATTTA**	388-397	(nykwCATTTA)	**TAAATTATAA**	247-256
Kr (Krueppel)		399-408	Kr	ABSENT	
(rAmnGGmAAA)	**AAATGGAAAA**				
Odd (Odd-skipped)		493-502	Odd	ABSENT	
(TAkmTnAwAn)	**TATATTAAAA**				
NF-Kappa		306-315	NF-Kappa	ABSENT	
(kGnGAAAyyC)	**TGTGAAATTC**				
Ant (antennapedia)		264-273	Ant	ABSENT	
(AATAmTwww)	**AATAATAAAA**				
C/EBP (CCAAT/enhancer binding protein)		260-269	C/EBP		
(wrnAAAyAAw)	**AGTAAATAAT**	452-461	(TkGnmAATwA)	**TTGAAAATAA**	168-177
(TTGGwAAnww)	**TTGGAAATTT**				
C/EBPalp (CCAAT/enhancer binding protein)			C/EBPalp		
(TmwnTATTTy)	**TAAGTATTTT**	44-53	(sywAmACmAs)	**GTTAAACAAC**	22-31
(wykATTknCA)	**TTTATTTGCA**	84-93	(sAkwTyGTrA)	**CATATTGTAA**	79-88
(sATTkTGnrm)	**GATTTTGTGC**	108-117	(TnwkTATTTy)	**TTTTTATTTC**	102-111
(AwTAnwAAwT)	**AATACAAATT**	170-179	(knnTTTGCwk)	**TTTTTTGCTG**	130-139
(nwTTGTGnmA)	**CATTGTGAAA**	303-312	(wmwTyATTTn)	**TATTTATTTA**	217-226
(rwTGsArAAn)	**AATGGAAAAG**	400-409	(wATTrmnmAA)	**TATTAAGAAA**	229-238
(TrTTTsCwrw)	**TATTTGCTGA**	477-486	(ArATyGTsnm)	**AAATTGTGAA**	299-308
			(nwTTGkGnmA)	**GATTGTGAAA**	314-323
			(mwmATAAAkw)	**AAAATAAATT**	321-330
C/EBPbeta (CCAAT/enhancer binding			C/EBPbeta		
protein)	**GTGGGTGTTG**	133-142	(TTrTGCmAnA)	**TTATGCCATA**	66-75
(kwGGGyGTkr)	**AAGTGCTTAA**	149-158			
(wkTssTTAAw)	**TTTTTCCAAC**	226-335			
(wTTTyyCmAy)	**GACATTTCTT**	533-542			
(GwmATTTCyn)					

value among the *copia* sequences in the *repleta* group was 0.18 (*D. mojavensis* and *D. koepferae*), which reinforces the idea that this group of species harbors a single *copia* subfamily. But the great divergence of these sequences compared to those of the *melanogaster* group (0.82) between *D. koepferae* and *D. sechellia*) confirms the proposition of Jordan and McDonald [25, 26] that there are at least two *copia* families in the genus *Drosophila*. This hypothesis is reinforced by the evolutionary relationships presented here. The unrooted tree obtained by the parsimony method shows two main groups: one with *copia* sequences of the *repleta* species and another with *copia* sequences of the *melanogaster* species, *Z. tuberculatus* and *D. willistoni*.

The wide distribution, the heterogeneous occurrence of *copia* in the *Drosophila* and *Sophophora* subgenera suggest that *copia* might have been present in the common ancestor of the genus *Drosophila* and have been vertically transmitted over evolutionary time. Hence, the divergence rates between the species groups should be, as observed, so great that they cannot be recognized anymore, or the sequences may have been lost by stochastic events. This might be the case of the *repleta* species, in which only faint bands (subgroup *mercatorum*: *D. hydei*, *D. mercatorum*; subgroup *mulleri*, *mojavensis* cluster: *D. navojoa*, *D. mojavensis*) or no amplification at all (subgroup *mulleri*, *mojavensis* cluster: *D. arizonae*, *longicornis* cluster: *D. longicornis*, and all four studied species of the *eremophila* cluster) were observed.

Moreover, very low distance values were also found between species belonging to a different genus (species of the *melanogaster* and *willistoni* groups and the *Zaprionus* genus) and between different groups of species within the *Sophophora* subgenus (*melanogaster* and *willistoni* groups). Taking together the nucleotide divergence and the parsimony analysis, the results may be indicative of horizontal transfers between the *D. willistoni* / *Zaprionus tuberculatus* / *melanogaster* species subgroup. However, it is not possible to infer the direction of the postulated events. Horizontal *copia* transfers have been previously proposed between species of the *melanogaster* subgroup [25], between *D. melanogaster* and *D. willistoni* [27], and between *D. melanogaster* and *D. simulans* [41]. When we included in our analysis the *copia* reported by Jordan and McDonald [25] plus sequences of eight species from the *repleta* group, we observed the inconsistencies reported by Jordan and McDonald [25] and Jordan *et al.* [27]. Additional phylogenetic incongruences can be observed between *Z. tuberculatus* and species of the *melanogaster* subgroup (*D. melanogaster* / *D. yakuba*), and between *D. mojavensis*, *D. pachuca* and species of the *buzzatii* cluster. Since geographic and temporal overlap between donor and recipient species are the minimum requirement to infer horizontal

transfer, only the event between *Z. indianus* and *D. yakuba* or *D. melanogaster* might suggest such a transfer because the three species share their range distribution in Africa [28,48]. Horizontal transfer between transposable elements of *Drosophila* has been shown to be a very frequent event. In addition to the classical reports [4,5,11–13,16,17,44], other examples have been published more recently [1,7,24,30,32,39]. It has been postulated that cross-species transfers may be an effective strategy by which TEs avoid inactivation over evolutionary time [33,37,43]. Since *copia* is known to be subject to effective host-mediated repression, selective pressure might favor its horizontal transfer over evolutionary time [34,41]. Nevertheless, it is important to point out that a potential area of weakness for this kind of research is the presence of several copies of the transposable element in the same species (ancestral polymorphism). Comparisons of paralogous copies of elements and varying rates of the sequence evolution of TE copies within and between species are factors which can yield incongruent phylogenies even under conditions of strict vertical transmission, as stressed by Zupunski *et al.* [50]. Another possibility, sequence similarity between distantly related species due to conservation of small motifs [6], could explain the similarity between *copia* sequences of species that do not share the same environments, since the LTR-ULR *copia* regions analyzed are regulatory protein-binding domains which control *copia* expression and spreading in natural populations [26].

Positive diversifying selection acting between *copia* families (*melanogaster* and *repleta*) and negative purifying selection acting within these families were reported by Jordan and McDonald [26] when studying the *copia* LTR-URL natural variation among seven species of the *melanogaster* group and two species of the *repleta* group. According to these authors, the ULR nucleotide diversity (π) in the *repleta* group of species was 2.2 times greater than that of the LTR. By increasing the number of *repleta* species analyzed, we showed that this ratio is even higher (3.4). This result reinforces the hypothesis of functional constraint of *copia* ULR regulatory regions within the *repleta* family. The π value in the *repleta* LTR and ULR was 3.0 and 2.8 times higher, respectively, than in the *copia melanogaster* family; these ratios are lower than those reported by Jordan and McDonald [26]. Despite the differences between the two studies, which are probably due to the smaller number of *repleta* species studied by those authors, the negative purifying selection explains the very low distance values within each species group. The fact that the nucleotide diversity in the *repleta* LTR and ULR is greater than in those of the *melanogaster* group and that their distance values are higher reinforces the idea that the *copia* of the *repleta* group is a more ancestral family than that of the *melanogaster* group.

Another approach used by us to compare the *copia* sequences harbored by the *melanogaster* species on the one hand and by the *repleta* species on the other was to analyze the LTR-ULR transcription factor binding sites. DNA-binding transcription factors play a central role in transcription regulation [29]. In this work, we found a similar repetition number of the C/EBP, TBP, FTZ and Zen motifs in the *repleta* and *melanogaster copia* families. However, these motifs are not homologous, neither according to their position in the LTR-ULR sequences nor according to their sequences. Moreover, it is known that regulatory regions can maintain their functions in spite of structural reorganization, as a result of species-specific losses and gains of transcription factor binding sites [15, 31, 36]. Of the 16 motifs found in this study, E47, CFF and Embry were absent in the *repleta* species, and Oct-2.1, Kr, Odd, Nf-kappa and Ant were absent in the *melanogaster* species (Tab. III). On the contrary, the occurrence of several motifs in common in the *repleta* and *melanogaster* LTR-ULR *copia* sequences could be an indication that both *copia* families regulate the TE activity in the same way. Because LTR retroelements may be continually generating variation within their noncoding regions, continuous opportunities might exist for natural selection to favor the evolution of adaptive enhancer configurations. Hence, diversifying selection could explain so highly divergent sequences between the *repleta* and *melanogaster* species groups and so greatly conserved sequences within these groups. Taking together the low homology of the motif sequences, the phylogenetic relationships and the high level of nucleotide divergence between the *melanogaster* and *repleta copia* sequences, the occurrence of at least two retrotransposon *copia* families in *Drosophila* seems to be a robust hypothesis. However, additional analysis including species from other groups of the *Sophophora* and the *Drosophila* subgenera might fill the gap and clarify whether the discontinuity of *copia* sequences between the *repleta* and *melanogaster* groups is real or not.

ACKNOWLEDGEMENTS

The authors wish to thank Oriol Cabré for providing us with plasmid PTZ18 and W. Etges, F.M. Sene, L. Madi-Ravazzi, F.R. Torres, C.R. Ceron, H.E.M.C. Bicudo, and M. Manfrin for providing us with strains of *Drosophila.* This work was supported by FAPESP (Grant 00/11313-0 to C.M.A.C and fellowship 97/14646-5 to L.M.A) and CNPq.

REFERENCES

[1] Almeida L.M., Carareto C.M., Multiple events of horizontal transfer of the *Minos* transposable element between *Drosophila* species, Mol. Phylogenet. Evol. 35 (2005) 583–594.
[2] Biémont C., Cizeron G., Distribution of transposable elements in *Drosophila* species, Genetica 105 (1999) 43–62.
[3] Boeke J.D., Garfinkel D.J., Styles C.A., Fink G.R., Ty elements transpose through an RNA intermediate, Cell 40 (1985) 491–500.
[4] Brunet F., Godin F., David J.R., Capy P., The mariner transposable element in the Drosophilidae family, Heredity 73 (1994) 377–385.
[5] Brunet F., Godin F., Bazin C., Capy P., Phylogenetic analysis of Mos1-like transposable elements in the Drosophilidae, J. Mol. Evol. 49 (1999) 760–768.
[6] Capy P., Bazin C., Higuet D., Langin T., Dynamics and Evolution of transposable elements, 1st edn., Landes Bioscience, Springer, Heidelberg, 1998.
[7] Castro J.P., Carareto C.A., *P* elements in *saltans* group of *Drosophila:* a new evaluation of their distribution and number of genomic insertion sites, Mol. Phylogenet. Evol. (2004) 383–387.
[8] Cavarec L., Heidmam T., The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins, Nucleic Acids Res. 21 (1993) 5041–5049.
[9] Cavarec L., Jensen S., Heidmam T., Identification of a strong transcriptional activator for the *copia* retrotransposon responsible for its differential expression in *Drosophila hydei* and *melanogaster* cell line, Biochem. Bioph. Res. Co. 203 (1994) 392–399.
[10] Cavarec L., Jensen S., Casella J.F., Cristescu S.A., Heidmam T., Molecular cloning and characterization of a transcription factor for the *copia* retrotransposon with homology to BTB-containing lola neurogenic factor, Mol. Cell. Biol. 17 (1997) 482–494.
[11] Clark J.B., Kidwell M.G., A phylogenetic perspective on *P* transposable element evolution in *Drosophila*, Proc. Natl. Acad. Sci. USA 94 (1997) 11428–11433.
[12] Clark J.B., Maddison W.P., Kidwell M.G., Phylogenetic analysis supports horizontal transfer of *P* transposable elements, Mol. Biol. Evol. 11 (1994) 40–50.
[13] Clark J.B., Altheide T.K., Sclosser M.J., Kidwell M.G., Molecular evolution of *P* transposable elements in genus *Drosophila I.* The *saltans* and *willistoni* species groups, Mol. Biol. Evol. 12 (1995) 902–913.
[14] Csink A.K., McDonald J.F., Analysis of *copia* sequence variation within and between *Drosophila* species, Mol. Biol. Evol. 12 (1995) 83–93.
[15] Cuadrado M., Sacristan M., Antequera F., Species-specific organization of CpG island promotors at mammalian homologous genes, EMBO Rep. 2 (2001) 586–592.
[16] Daniels S.B., Peterson K.R., Straubaugh L.D., Kidwell M.G., Chovnick A., Evidence for horizontal transmission of the *P* transposable elements between *Drosophila* species, Genetics 124 (1990) 339–355.
[17] Daniels S.B., Chovnick A., Boussy I., Distribution of *hobo* transposable elements in the genus *Drosophila*, Mol. Biol. Evol. 8 (1990) 589–606.

[18] Durando C.M., Baker R.H., Etges W.J., Heed W., Wasserman M., DeSalle R., Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters, Mol. Phylogenet. Evol. 16 (2000) 296–307.
[19] Emori Y., Shiba T., Kanaya S., Inouye S., Yuki S., Saigo K., Determination of the nucleotide sequences of *copia* and *copia*-related RNA in *Drosophila* VLP, Nature 315 (1985) 773–776.
[20] Flavell A.J., Levis R., Simon M.A., Rubin G.M., The 5'termini of RNAs encoded by the transposable element *copia*, Nucleic Acids Res. 9 (1981) 6279–6291.
[21] Flavell A.J., Dunbar E., Anderson R., Pearce S.R., Hartley R., Kumar A., *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants, Nucleic Acids Res. 20 (1992) 3639–3644.
[22] Grabe N., AliBaba2: context specific identification of transcription factor binding sites, In Silico Biol. 2 (2002) S1-S15.
[23] Hasegawa M., Kishino H., Yano T., Dating of the human-ape splitting by molecular clock of mitochondrial DNA, J. Mol. Evol. 21 (1985) 160–174.
[24] Heredia F., Loreto E.L.S., Valente V.L.S., Complex evolution of *gypsy* in Drosophilid species, Mol. Biol. Evol. 21 (2004) 1831–1842.
[25] Jordan I.K., McDonald J.F., Evolution of *copia* retrotransposon in *Drosophila melanogaster* species subgroup, Mol. Biol. Evol. 15 (1998) 1160–1171.
[26] Jordan I.K., McDonald J.F., Interelement selection in the regulatory region of the *copia* retrotransposon, J. Mol. Evol. 47 (1998) 670–676.
[27] Jordan I.K., Matyunina V.L., McDonald J.F., Evidence for recent horizontal transfer of long terminal repeat retrotransposon, Proc. Natl. Acad. Sci. USA 96 (1999) 12621–12625.
[28] Lachaise D., Cariou M.-L., David J.R., Lemeunier F., Tsacas L., Ashburner M., Historical biogeography of the *Drosophila melanogaster* species subgroup, Evol. Biol. 22 (1988) 159–225.
[29] Long F., Hong L., Chang H., Sumazin P., Zhang M., Zilberstein A., Genome-wide prediction and analysis of function-specific transcription factor binding sites, In Silico Biol. 4 (2004) 395–410.
[30] Loreto E.L.S., Valente V.L.S., Zaha A., Silva J.C., Kidwell M.G., *Drosophila mediopunctata P* elements: A new example of horizontal transfer, J. Hered. (2001) 375–381.
[31] Ludwig M., Bergman C., Patel N., Kreitman M., Evidence for stabilizing selection in a eukaryotic enhancer element, Nature (2000) 564–567.
[32] Martinez-Sebastian M., Hernández M., Mejias B., Gas M., Pérez A., Pascual L., De Frutos R., Evolutionary patters of *gypsy* and *bilbo* retrotransposon families in the *Drosophila* species of the *obscura* group, Mol. Phylogenet. Evol. 22 (2002) 254–266.
[33] Maruyama K., Hartl D.L., Evidence for interspecific transfer of transposable element *mariner* between *Drosophila* and *Zaprionus*, J. Mol. Evol. 33 (1991) 514–524.
[34] Matyunina L.V., Jordan I.K., McDonald J.F., Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation, Proc. Natl. Acad. Sci. USA 93 (1996) 855–864.

[35] McDonald J.F., Matyunina L.V., Wilson S.W., Jordan I.K., Bowen N.J., Miller W.J., LTR retrotransposon and the evolution of eukaryotic enhancers, Genetica 100 (1997) 3–13.
[36] Piano F., Parisi M.J., Karess R., Kambysellis M.P., Evidence for redundancy but not trans factor-cis element coevolution in the regulation of *Drosophila* Yp genes, Genetics 152 (1999) 605–616.
[37] Pinsker W., Haring E., Hangemann S., Miller W.J., The evolutionary life history of *P* transposons: from horizontal invaders to domesticated neogenes, Chromosoma 110 (2001) 148–158.
[38] Posada D., Cradall K.A., MODELTEST: testing the model of DNA substitution, Bioinformatics 14 (1998) 817–818.
[39] Robertson H.M., Soto-Adames F.N., Walden K.K., Avancini R.M., Lampe D.J., The *mariner* transposon of animals: horizontally jumping genes, in: Syvanen M., Kado C.I. (Eds.), Horizontal gene transfer, Academic Press, San Diego, CA, 2002.
[40] Rozas J., Rozas R., DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis, Comput. Appl. Biosci. 13 (1997) 307–311.
[41] Sánchez-Gracia A., Maside X., Charlesworth B., High rate of horizontal transfer of transposable elements in *Drosophila*, Trends Genet. 21 (2005) 200–203.
[42] Serfling E., Lubbe A., Dorsch-Hasler K., Schaffner W., Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes, EMBO J. 4 (1985) 3851–3859.
[43] Silva J.C., Loreto E.L., Clark J.B., Factors that affect the horizontal transfer of transposable elements, Curr. Issues Mol. Biol. 6 (2004) 57–71.
[44] Simmons G.M. Horizontal transfer of *hobo* transposable element within the *Drosophila melanogaster* species complex: evidence of DNA sequencing, Mol. Biol. Evol. 9 (1992) 1050–1060.
[45] Sneddon A., Flavell A.J., The transcriptional control of the *copia* retrotransposon, Nucleic Acids Res. 17 (1989) 4025–4035.
[46] Swofford D., PAUP: Phylogenetic analysis using parsimony (and other methods), Version 4.0.b10. Sinauer, Sunderland, Massachusetts, 2000.
[47] Thompson J.D., Higgins D.G., Gibson T.J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[48] Tsacas L., L'identité de *Zaprionus* vittiger Coquillet et révision des espèces tropicales affines, Bull. Soc. Ent. Fr. 85 (1980) 141–153.
[49] Wilson S.W., Matyunina L.V., McDonald J.F., An enhancer region within the *copia* untranslated leader contains binding sites for *Drosophila* regulatory proteins, Gene 209 (1998) 239–246.
[50] Zupunski V., Gubenack F., Kordis D., Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons, Mol. Biol. Evol. (2001) 1849–1863.