Mol Genet Genomics
DOI 10.1007/s00438-008-0324-7

ORIGINAL PAPER

Conserved motifs and dynamic aspects of the terminal inverted repeat organization within *Bari*-like transposons

Roberta Moschetti · Sarantis Chlamydas · Renè Massimiliano Marsano · Ruggiero Caizzi

Received: 14 November 2007 / Accepted: 15 January 2008

Abstract In this work the structural variations of Terminal Inverted Repeats (TIR) of *Bari* like transposons in Drosophila species has been studied. The aim is to try and assess the relevance of different variants in the evolutionary distribution of *Bari* elements. *Bari* is a member of the widespread *Tc1* superfamily of transposable elements that has colonized most species of the Drosophila genus. We previously reported the structure of two related elements that differ in their TIR organization: *Bari1* harbouring 26-bp TIR (short TIRs) and *Bari2* with about 250-bp TIR (long TIR). While elements with short TIRs are complete and potentially autonomous, long ones are invariably composed of defective copies. The results show that in *D. pseudobscura*, *D. persimilis* and *D. mojavensis*, there is a third class of *Bari* elements, *Bari3*, that exhibit a long TIR structure and are not defective. Phylogenetic relationships among reconstructed transposases are consistent with the three subfamilies sharing a common origin. However, the final TIR organization into long or short structure is not related by descent but appears to be lineage-specific. Furthermore, we show that, independently of origin and organization, within the 250-bp terminal sequences there are three regions that are conserved in both sequence and position suggesting they are under functional constraint.

Keywords Transposable elements · Tc1-mariner like · Terminal inverted repeats · Comparative genomics

Communicated by S. Hohmann.

Nucleotide sequence data from this article have been deposited in the EMBL/GenBank databases with the accession numbers: AM493769, AM493770, AM493771, AM493772.

R. Moschetti · S. Chlamydas · R. Caizzi (✉)
Department of Genetics and Microbiology,
University of Bari, Via Amendola 165/A, 70126 Bari, Italy
e-mail: r.caizzi@biologia.uniba.it

R. Massimiliano Marsano
Department of Genetics,
Biology of Microorganisms, Antropology,
Evolution, University of Parma,
Viale Usberti 11/A, 43100 Parma, Italy

Introduction

The terminal ends of class II transposons (Finnegan 1992) have an important role in the transposition mechanism because they harbour the binding sites for the transposase protein. Hence, modifications of these sequences can affect the success of an element both in evolution and within a particular host genome. In this study, we report the organization of terminal repeats of *Bari* elements in distantly related Drosophila species with the aim to track the evolutionary modifications that affect their distribution in different species of Drosophila genus.

Bari is a member of the *Tc1-mariner* superfamily of DNA transposons that are widespread in vertebrates, invertebrates and fungi. The vast majority of *Tc1*-like (TLE) and *mariner*-like (MLE) elements analysed so far are defective. This is mainly due to a variety of point mutations and/or deletions that distrupt the overall organization of a complete and functional transposon. These defective elements represent ancient relics of autonomous copies and are still informative in tracking the evolutionary biology of an individual family during phylogenesis. All autonomous elements share a common TA target site that is duplicated after transposition events and a conserved D,D(34)E signature in the catalytic C-terminal domain of the transposase gene, the only protein they encode. The catalytic signature is also present in the bacterial transposon Tn7, in many

insertion sequences (IS) of a wide spectrum of bacteria, in the phage Mu, in the integrase gene of HIV virus and in the retroviridae in general and in the RAG1 recombinase of the vertebrate V(D)J recombination system of the immunoglobulin genes (reviewed in Plasterk et al. 1999).

Autonomous *Tc1-mariner* elements move around the genome by a conservative "cut and paste" mechanism that appears in vitro, to be independent of species-specific host factors (Lampe et al. 1996; Vos et al. 1996), although a host protein has been identified as cofactor in the *Sleeping Beauty* transposition in vivo (Zayed et al. 2003). Such an apparent plainness in the transposition process allows these transposons to spread more easily, horizontally between different species. Actually, it is on this peculiar property that these transposons have been proposed as generalized transformation vectors (Plasterk et al. 1999; Izsvák et al. 2000).

One of the earliest steps of the transposition mechanism is the binding of the transposase, through a bipartite N-terminal domain, to a short stretch of DNA present at the terminal inverted repeats (TIR) of the transposon. The structure of TIRs of different families varies both in size and number of transposase-binding sites. For example, *Tc1* of *C. elegans* and *mos1* of *D. mauritiana*, the founder elements of the *Tc1-mariner* superfamily, have TIRs shorter than 100 bp and a single binding site per repeat; *Tc3* of *C. elegans* has a long TIR, >400 bp, and two binding sites per repeat, although the second one is not necessary for the transposition (Fisher et al. 1999). Another group of *Tc1*-like elements possesses an IR-DR structure, which consists of about 250-bp-long inverted repeats (IR) each possessing two 15–18 bp direct repeats (DR). The DRs are the DNA-binding sites for the transposase and are generally located at the beginning and at the end of the 250-bp TIR. Examples of this group are elements discovered in insects (Franz and Savakis 1991; Merriman et al. 1995; Ke et al. 1996), amphibians (Lam et al. 1996) and fishes (Radice et al. 1994; Izsvák et al. 1995) including *Sleeping Beauty* (SB), a very efficient reconstructed artificial transposon (Ivics et al. 1997). The differences in the TIR structure parallels the different DNA-binding domain compositions present in the transposase proteins encoded by *Tc1-mariner*-like elements. This, in turn, guarantees the specificity in the recognition process between one transposase to its own ends during the paired end complex (PEC) formation (reviewed in Brillet et al. 2007).

Bari1 represents an exception to the terminal repeats organization in that it was found to possess two DRs notwithstanding the fact that the TIR is short, i.e.26 bp. Thus, the second DR is located outside the inverted repeats. The *Bari1* transposon was orginally discovered in *D. melanogaster* (Caizzi et al. 1993) as a very homogeneous element, since almost all of the copies in the genome were shown to be complete and virtually identical to each other and also to the copies present in the species of the melanogaster complex, *D. simulans, D. mauritiana* and *D. sechellia*, that split from *D. melanogaster* around 2.3–2.7 millions years ago (MYA) (Li et al. 1999). A variant and defective element possessing a ~250-bp-long IR-DR structure was then detected in two more distantly related species, *D. erecta* and *D. diplacantha* (Moschetti et al. 1998); this presented the possibility of studying the structural modifications occurring during the evolutionary history of the *Bari* transposon since *Bari1* of *D. melanogaster* and the "erecta type" elements share 70% nucleotide identity and 86% amino acid similarity in their transposase genes. In the survey of the euchromatic complement of transposable elements in the sequenced *D. melanogaster* genome, the extreme similarity between *Bari1* copies was definitely ascertained and it was also found that the "erecta type", called in this work *Bari2*, coexists with *Bari1* in the same genome (Kaminker et al. 2002). The cohabitation of related copies in a genome and the absence of one or the other element in phylogenetically closer species is a quite common phenomenon among the *Tc1-mariner* members; yet, the timing and the events responsible for the diversification and the evolutionary relationship between autonomous and defective elements are poorly understood.

To get more insight into the biology of *Bari*-like elements and the TIR diversification, we have analysed the genomes of representative drosophiline species whose sequencing is in an advanced state (http://flybase.net) and also by cloning elements from the not-yet-sequenced species. New *Bari*-like elements, both complete and defective, have been discovered that allowed us to track their evolutionary history and to detect a conserved motif in the TIR sequence in the phylogenetically distinct but related elements.

Materials and methods

Molecular analysis

The source of DNA from different Drosophila species is the same of that isolated and described in a previous paper (Moschetti et al. 1998). PCR experiments were carried out with the degenerate oligo sequence 5′TCAGCTGCAGT CAAAA(TG)TATTT(TA)CAC. The underlined nucleotides containing *PstI* and *PvuII* restriction sites were added for cloning purposes. PCR was carried out in a 50-µl final volume containing 50 ng genomic DNA, appropriate buffer, 1 U of Platinum Taq polymerase (Invitrogen, Carlsbad, CA, USA) for 30 of the following cycles: 94°C for 30″, 60°C for 30″, 72° for 90″. Initial melting was at 94°C for 3′ and final extension at 72°C for 5′. PCR products were cloned in pGEM-T vector or in *PstI* digested pUC19

plasmid. Southern blot hybridization of PCR products was carried out at 64°C in standard buffer (Moschetti et al. 1998). Filters were washed for 15′ at two stringency conditions: 1X SSC, 0.1% SDS, 60°C (reduced stringency) or 0.1XSSC, 0.1% SDS, 64°C (high stringency). Probes *Bari1* and *Er1* were labelled with ^{32}P-dATP by oligo priming.

In silico search

Genomic Databases of Flybase browser (http://flybase.net; version FB2007_02) were searched by BLASTn using alternatively the DNA sequence of *Bari1* (X67681) and *Er1* (Y13853). Sequences extracted were singularly analysed to find the best alignment with the two queries with the Blast2 tool (http://www.ncbi.nlm.nih.gov). TBLASTn with deducted transposases of *Bari1* and *Er1* was used for those genomes that showed a low score with the BLASTn procedure. Only elements showing the best hits were used in successive alignments analysis. Global DNA alignment was done by Pairwise Sequence Alignment tool (http://genome.cs.mtu.edu/align.html).

Sequence analysis, alignments and phylogenetic analysis

Amino acid alignments of conceptual nucleic acid translations were accomplished with MacVector and ClustalW (Thompson et al. 1994). Two conserved regions of *Tc1* like transposases were considered in the multiple alignments: the bipartite helix-turn-helix motifs in the N-terminal and the DDE catalytic domain in the C-terminal. The alignments were enriched with the annotated sequences of *Minos* (Z29098), *Paris* (U26938), *S* (U33463), *Tc1* (K01135) and with the following elements retrieved from the TBLASTn analysis: *grim*, from *D. grimshawe* (CH916366, position 24403520-24405168), *vir-S* from *D. virilis* (CH940657, position 855111-856809). Alignments were accomplished using ClustalX, after which bootstrapped trees (1,000 replicates) were created with the neighbour joining method (Saitou and Nei 1987). Trees were drawn with TreeView.

Results

The TIRs of Bari1 and Er1 (Bari2) elements possess three imperfect DRs

We previously reported that at both ends of *Bari1* and *Er1* (erecta type) elements there are two 18-bp-long DRs, although the overall TIR organization is different in these two strictly related elements (Moschetti et al. 1998). However, a careful re-examination of the ~250-bp terminal sequences revealed the presence of an additional DR in both *Bari1* and *Er1* elements, located between the above DRs. The position and sequence of these three DRs in the left and right ends are highly conserved (Fig. 1a, b). Following the nomenclature of Cui et al. (2002), the outer DRs are called Lo and Ro, the inner ones are called Li and Ri, and the new discovered internal ones are called Lm and Rm (m, in the middle). The left and right TIR sequences of *Bari1* and *Er1* elements are not identical to each other, as in other *Tc1*-like elements (Merriman et al. 1995; Radice et al. 1994). Still, 12 of 18-nt are invariably found in all DRs, suggesting a functional constraint of this architecture.

In order to investigate if a structure with three DRs exists in other TLEs, we have analysed the TIR structure of Drosophila *Paris*, *S* and *minos* and of mosquito *Tucur* and *Tiang* elements. We found that *Paris*, *S* and *minos* do have

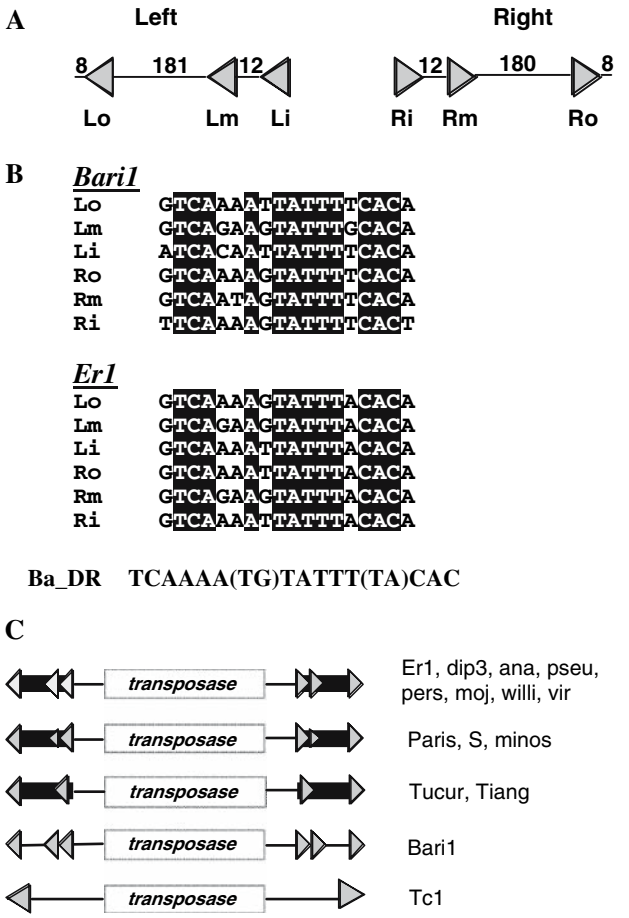

Fig. 1 The terminal repeats organization. **a** Schematic representation of left and right inverted repeats in *Bari1* and *Er1* elements. *Gray triangles* represent the position of the 18-nucleotide DR and numbers indicate the nucleotides separating them. The names of individual DRs are indicated at the bottom. **b** Alignment of all DRs in *Bari1* and *Er1* elements. Identical nucleotides are shadowed. Below is the degenerate sequence used in the PCR experiments. **c** An overall picture of the TIR structure in different *Tc1*-like elements whose name are reported on the right. A ~250-bp-long TIR structure is represented by black boxes. The DRs are represented as in **a**. A larger triangle in *Tc1* represents a 54 bp long TIR

Lm and Rm but they compose only one half of their 18-bp DR. *Tucur* and *Tiang* possess only the outer and inner DRs, the last ones being in a subterminal position (Fig. 1c). Taken together, these observations revealed a plasticity in the TIR organization and the specificity of three complete DRs in the *Bari*-like elements only.

In a pairwise comparison of DRs in all combinations, we note that the *Er1* DRs are more similar to each other than *Bari1* DRs, thus suggesting that the presence of long TIRs may influence an higher sequence homogeneity within the DRs.

Finding Bari-like elements in distantly related species

To address the general issue of TIR structure evolution, we have identified new *Bari*-like sequences in the genomes of different Drosophila species by experimental cloning and in silico search.

On the basis of similarity of the outer and inner DRs, we designed a 16-nucleotide degenerate oligo, called Ba_DR (bottom line in Fig. 1b), that should amplify both kinds of elements present in a genome, either with a short (*Bari1*) or with a long (*Bari2*) TIR structure. Then the identity of each amplified band can be characterized by blot hybridization and by sequencing the cloned PCR bands. We realized that the presence of multiple DRs in a single element could produce more PCR fragments depending upon which DR is functioning as the template in the PCR reaction. This is clearly shown to occur when we used plasmid DNA harbouring *Bari1* or *Er1* inserts (Fig. 2a, lanes 1, 2). Nevertheless, we will rely on the cloning and sequencing of individual PCR bands to get sufficient information regarding both the TIR structure and transposase gene if at least one outer DR is present in the cloned fragment. An example of this strategy is shown in Fig. 2. Among 17 different genomic DNA samples tested, we obtained PCR bands in 13 samples. Seven of those can be tentatively assigned as *Er1* homologs by the hybridization results (Fig. 2b, c). The remaining six PCR positive strains are not related to *Bari1* because of not showing cross hybridization in the same conditions (data not shown). Most of the cloned and sequenced fragments analysed thus far present a high level of rearrangments that affect their TIR and transposase composition, rendering them uninformative for the present study. Here, we will report the data of the informative elements cloned from *D. sechellia, D. mimetica, D. montana* and *D. flavomontana* (*sec-Er, mim, mon, flav* in Table 1), together with the elements discovered by in silico search.

Finding Bari-like elements in the sequenced genomes

A BLASTn analysis has been performed over the whole genomes of Drosophila species present in the FlyBase browser using *Bari1* and *Er1* separately as queries. The

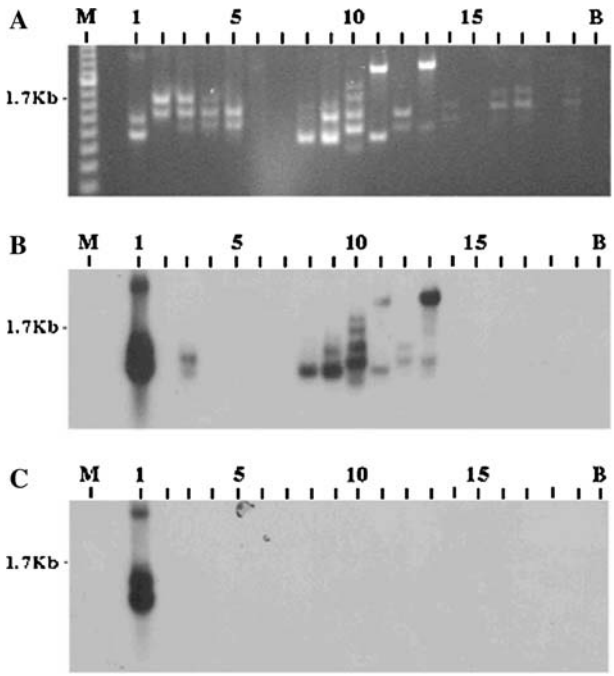

Fig. 2 PCR analysis from different genomic DNAs. **a** Ethidium bromide agarose gel showing the PCR fragments obtained with the degenerate Ba_DR primer and with the following template DNAs: (1) plasmid pU-Er1; (2) plasmid pU-Bari1; (3) *D. melanogaster;* (4) *D. pseudoobscura;* (5) *D. miranda;* (6) *D. affinis;* (7) *D. willistoni;* (8) *D. americana americana;* (9) *D. novamexicana;* (10) *D. lumnei;* (11) *D. littoralis;* (12) *D. flavomontana;* (13) *D. montana*; (14) *D. mojavensis mojavensis*; (15) *D. buzzatii*; (16) *D. mulleri*; (17) *D. repleta*; (18) *D. hydei*; (19) *D. polychaeta;* **b** is blank control; M is the 200-bp Marker DNA (Promega). **b** Autoradiography of the Southern blot hybridisation of the panel shown in **a** with Er1 probe and washed at medium stringengy. **c** The same filter after washing at high stringency. The 1.7-kb marks the position of a full length *Bari1* element

sequences showing the highest BLAST score (that is, a low *E*-value) were extracted and individually analysed in pairwise comparisons. Some genomes gave numerous hits with one or the other query. For example, *D. ananassae* contains at least 22 hits with a high score (*E*-value at least e-125) using *Bari1* as query. All of these 22 sequences are also found in the hits with the *Er1* query, but with a lower score (*E*-value e-42). Thus, *D. ananassae* appears to possess *Bari1*-like sequences only. Conversely, in *D. erecta* we found at least 30 sequences showing an higher score with the *Er1* query than the score obtained with the *Bari1* query, suggesting the presence of *Er1*-like elements and the absence of *Bari1*-like sequences in this species. By this comparative procedure, we were unanbiguously able to identify homologous elements in *D. sechellia, D. simulans, D. yakuba, D. ananassae* and *D. virilis*.

BLASTn search failed to find elements with significant scores in the genomes of *D. pseudobscura, D. persimilis, D. willistoni* and *D. mojavensis*. However, a TBLASTn analysis using the transposase amino acid sequences of

Table 1 List of *Bari*-like elements analysed

Element	Species	Size (bp)	TIR structure[a]	Amino acids of deducted transposase[b]	Reference elements
Bari1	*D. melanogaster*	1,726	SIR-3DR	339	X67681
mel-Er (consensus)	*D. melanogaster*	1,730	LIR-3DR	334*	chrX, 21971934-21973307 2R 691120-692691
sec-Ba	*D. sechellia*	1,727	SIR-3DR	339	CH481259 6449-8175
sec-Er	*D. sechellia*	1,271	ND	329*	This work
sim-Ba	*D. simulans*	1,726	SIR-3DR	339	CM000362.1 3123885-3125610
sim-Er	*D. simulans*	1,221	ND	310*	CM000361.1 19767684-19770083
yak	*D. yakuba*	467	LIR-3DR (deducted)	–	CH892673 259500-259966
Er1	*D. erecta*	1,638	LIR-3DR	325*	Y13853
dip3	*D. diplacantha*	1,676	LIR-3DR	337*	Y13852
ana	*D. ananassae*	1,727	LIR-3DR	337*	CH902621 3446828- 3448554
mim	*D. mimetica*	1,224	ND	334*	this work
pseudo	*D. pseudoobscura*	1,706	LIR-3DR	339	chr_U 18169953- 18168248
pers	*D. persimilis*	1,704	LIR-3DR	339	CH479182 160711-159008
willi	*D. willistoni*	1,715	LIR-3DR	339	CH964217 234341-236055; 162081-163569
moja	*D. mojavensis*	1,717	LIR-3DR	339	CH933806 6273693-6275409
vir	*D. virilis*	1,645	LIR-3DR	322	CH940657 79741-81385
mon	*D. montana*	1,161	ND	288*	This work
flav	*D. flavomontana*	1,433	ND	327*	This work

[a] TIR are classified as Long Inverted Repeat (*LIR*) and Short Inverted Repeat (*SIR*) followed by the number of Direct Repeats. *ND* not determined

[b] Asterisks denote defective ORFs

Bari1 and *Er1* as queries allowed us to identify homologous elements also in these species.

A list of all elements that we have selected and further analysed is reported in Table 1. The sequences that appeared to harbour complete terminal repeats were then analysed for the presence of a long or a short TIR and the number of DRs.

Bari-like elements fall in three distinct clades

An all-by-all comparison of the DNA sequences extracted is shown in Fig. 3 as dot plot graphics that allow rapid visualization of the extent of similarity within the sequences. There is a large variation in their relative similarity. The most divergent sequence we found was that of the *willi* element. In fact, no alignments above the threshold parameters of the program were detected in the comparison with *mel-er*, *sim-er*, *sec-er*, *Er1*, *dip3*, *vir*, *mon* and *flav* sequences.

In order to better understand the phylogenetic relationship between the different elements, we analysed the deduced transposase proteins. Elements showing defective ORFs were manually corrected in their nucleotide indels so that the start and stop codons were placed at the same position with respect to the start and stop codons of *Bari1* ORF. In a pairwise comparison, the complete and recostructed transposases showed an amino acid identity above 50% (Table 2), including the more divergent *willi* element, indicating the presence of a clear phylogenetic relationship among all of the *Bari*-like elements. The amino acid regions extending from the DNA-binding domain to the DDE domain of transposases (about 275 amino acids) were

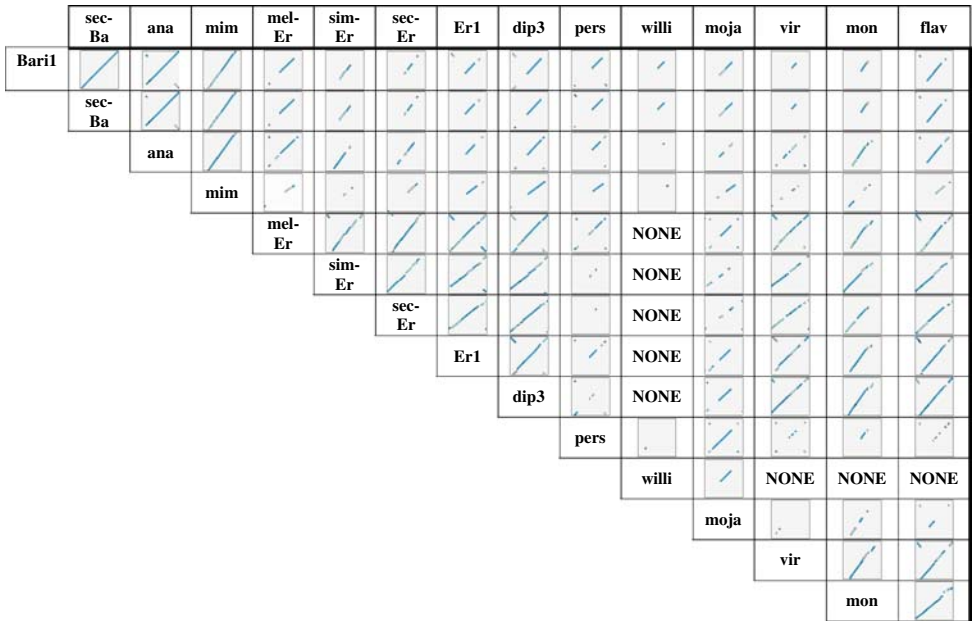

Fig. 3 The nucleotide similarity of *Bari* like elements. *Dot* plot comparison of *Bari*-like sequence listed in Table 1. Elements *sim-Ba* and *pseu* are not reported because they are nearly 100% identical to *Bari1* and *pers*, respectively. Sequences were treated as a "whole", irrespective of their length differences

Table 2 Pairwise percent amino acids identity among *Bari*-like deducted transposases

	ana	Mim	mel-Er	sim-Er	sec-Er	Er1	dip3	pers	willi	moja	vir	mon	flav
Bari1	81	76	59	61	56	64	67	67	66	64	56	54	60
	ana	81	56	57	54	62	64	64	62	61	55	56	59
		Mim	54	53	51	58	60	60	57	57	52	50	55
			mel-Er	64	92	73	73	59	54	57	57	55	60
				sim-Er	63	75	76	59	57	58	60	63	63
					sec-Er	71	72	57	52	55	56	54	58
						Er1	83	64	59	64	65	65	71
							dip3	66	60	65	63	64	69
								pers	66	74	55	54	60
									willi	63	50	53	56
										moja	55	55	59
											vir	63	63
												mon	62

aligned by ClustalX and the generated phylogram is shown in Fig. 4. Three clusters are clearly distinguishable: the Bari1 clade, which is present in the species of melanogaster group only; the Bari2 clade, which is distributed in both Drosophila and Sophophora species. There is an incongruence in the transposase phylogeny of *sim-er* that does not cluster with the elements from the drosophila complex, *sec-er* and *mel-er*, but appears to be closer to *Er1* and *dip* elements. A case that may represent an ancient horizontal transmission event, the Bari3 clade, is represented by the *willi*, *moja* and *pseu* (plus *pers*) elements (*pers* transposase is 99.7% identical to the *pseu* transposase).

Taken together, these data indicate that *Bari* elements have colonized most species of the Drosophila genus diversifying in at least three related subfamilies.

Alignment of different TIR shows conserved DRs

As reported in Table 1 and Fig. 1c, all *Bari*-like elements possessing at least one complete terminal sequence carry an inverted repeat structure with three DRs. This probably holds also for the shorter sequence of the *yak* element, where only two DRs were found, one outer and one middle. Taking into account that the distinction between long or short TIRs could be applied only to elements where both left and right ends are available, we reported in Table 3 the nucleotide identity between the two TIRs of elements belonging to the three Bari subfamilies. In every case, the elements possessing almost complete long TIRs showed a nucleotide identity above 90%. The lower values for the *vir, pseu* and *pers* elements are only apparent because their

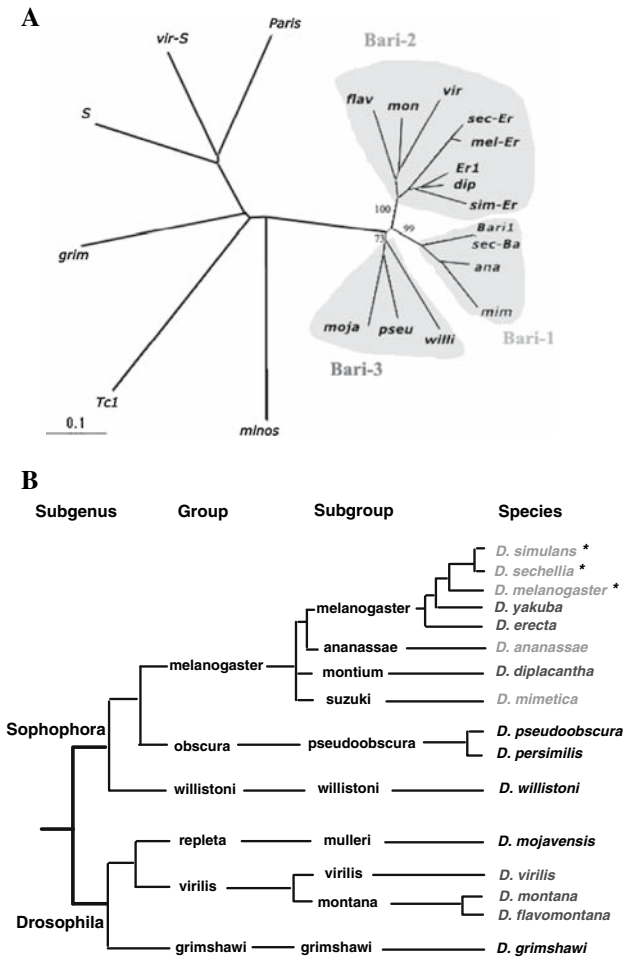

Fig. 4 Phylogenetic relationship between *Bari*-like and other *Tc1*-like transposases. **a** An unrooted neighbour-joining (*NJ*) phylogram is shown. Maximum parsimony (*MP*) produced a tree with similar overall topology. *Small numbers* are the percent of the NJ bootstrap value out of 1,000 replicates supporting the branch groups. The region analysed covers 275 amino acids encompassing the helix-turn-helix and the DDE domains of transposases. The following *Tc1*-like transposases were included in the tree: *Paris* (U26938), *S* (U33463), *minos* (Z29098), *Tc1* (X01005) and two new elements discovered in this study, *vir-S* from *D. virilis* and *grim* from *D. grimshawi*. **b** Schematic representation of the phylogenetic relationship between the species analysed in this study. Drosophila species are *grey coloured* according to the three Bari subfamilies shown in **a**. Species with asterisks contain *Bari1* and *Bari2* elements

TIRs are different in length. Excluding the gaps, the identity raises above 82%. As a comparison, *Bari1*, which has short TIRs, shows only 48% global identity in the ~250 bp corresponding regions. Interestingly, the *moja* element is the only one that has two TIRs identical to each other. Table 3 also reports the occurrence in the genomes of copies harbouring two complete or almost complete TIR structures together with the occurrence of TIR traces as revealed by the number of HSP of the Blast alghoritm (Atschul et al. 1997).

Table 3 Global identity of the ~250 bp TIRs and their occurrence in the genomes

	TIR's match percentage	N. elements with two almost complete TIRs	Number of HSP's better than 10.0 without gapping[a]
Bari1	48	5	28
ana	93	1	29
mel-Er	92	0	31
Er1	93	1	29
vir[b]	65	0	45
willi	96	1	164
pseu	78	1	44
pers	80	5	210
moja	100	11	357

[a] Effective High scoring Segment Pair (*HSP*) length: 18

[b] Right TIR is shorter (221 nt) elements are shadowed in a grey scale according to the clusters of Fig. 4

Noteworthy, the *ana* element, which is the closer one to *Bari1*, still possess 93% identity in its long TIRs and, when compared to *Bari1*, shows an high level of identity restricted to the first 27 and to the last 50 nucleotides of the terminal ends. In addition, these two TIR regions are the only ones that show similarity with the *Bari2* and *Bari3* elements. Taken together, these observations suggest that the maintenance or the loss of long terminal repeat organization is an ongoing process during species invasion.

The alignment of ~250 nt of all selected terminal sequences is shown in Fig. 5. The significant similarity is quite apparent within the DRs, which is independent of origin and overall organization.

Discussion

Full length and defective Bari-like elements: role of terminal repeat architecture

Numerous studies have established that the predominant mode of evolution governing the dispersion of *Tc1-mariner* elements through different species relies on three main forces: (1) the generation of random mutations that impair mobility giving rise to defective elements which are vertically transmitted (vertical inactivation); (2) the stochastic loss in some evolutionary lineages by genetic drift and/or natural selection; (3) the occasional invasion of a different species (horizontal transmission) that allows the survival of a transposon from the decay that occurred in the first two events (Lohe et al. 1995). However, other processes are also involved in maintaining the dispersion of transposons: the response of the host genome to control the deleterious effect of transposition; the role that defective elements may have in the transposition mechanism and finally the recruitment of some parasitic function into the adapted host

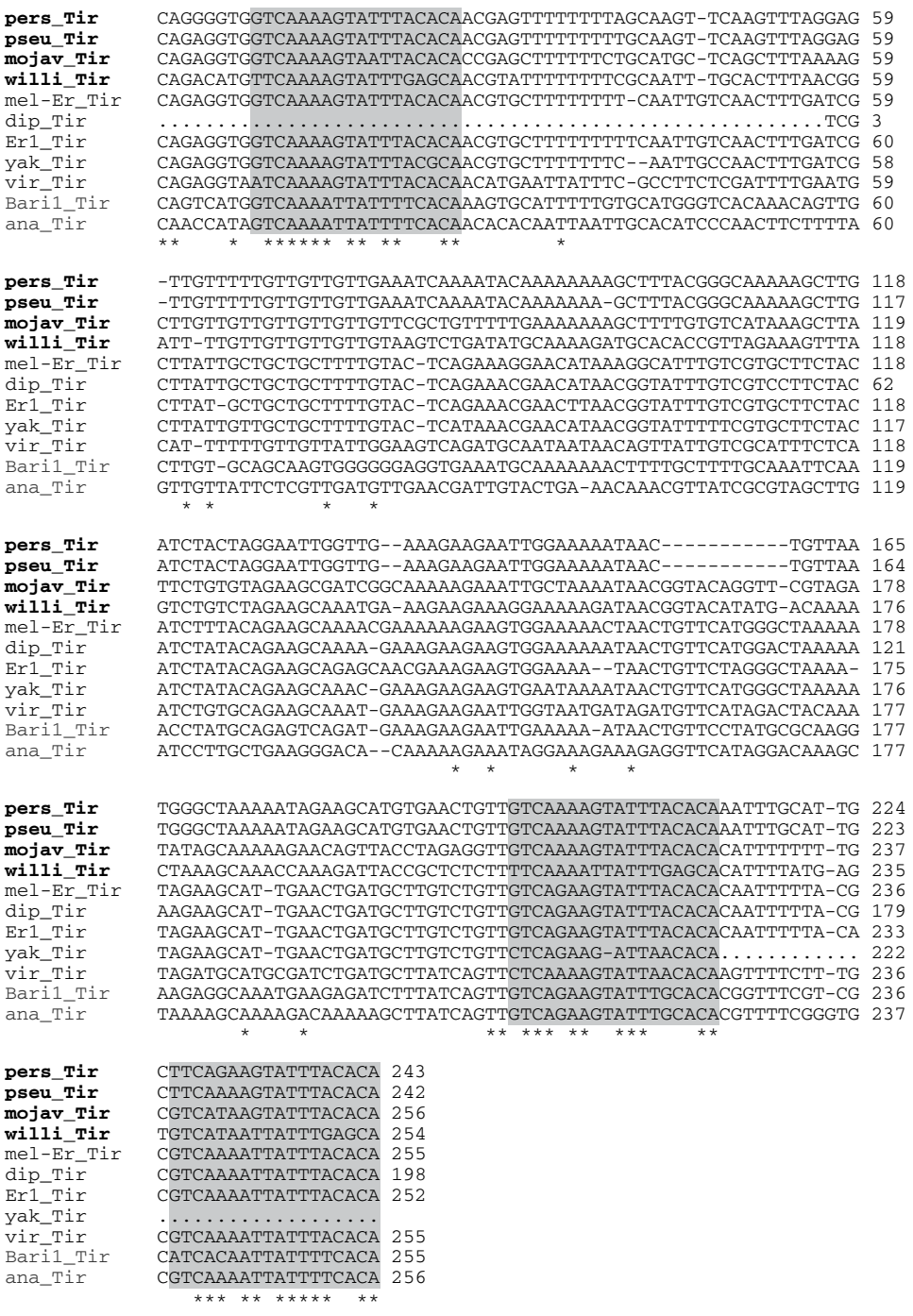

Fig. 5 Alignment of TIRs. A ClustalW multiple alignment of left inverted repeats from different *Bari*-like elements. *Dots* indicate missing nucleotides. The *shadowed* areas represent the DR regions. Sequences are *grey coloured* in accord with the clades of Fig. 4. *Asterisks* below the alignment mark the conserved sites in all sequences

function (Kidwell and Lisch 2001). Presumably, the picture of the transposon *Bari* in different species represents the balance of these events leading to the current dispersion of one of the most represented TLE in the Drosophila genus (Moschetti et al. 1998; Biémont and Cizeron 1999), if compared to *Minos* (Arcà and Savakis 2000), *S* (Merriman et al. 1995) and *Paris* (Petrov et al. 1995).

Three subfamilies of *Bari* elements exist in the species of the Drosophila genus (Fig. 4) and all of them encompass full length and defective copies. The full-length ones are probably autonomous and active (Junakovic et al. 1997; Lerat et al. 2003). All of the *Bari1* copies identified in the four species of the melanogaster complex, that split from a common ancestor around 2.3–2.7 MYA (Li et al. 1999), are full length and almost identical to each other, suggesting that they have a recent origin. However, within the *Bari1* clade, this element is also present in the form of a defective copy in *D. ananassae* with a long TIR structure and in *D. mimetica* with an

unknown TIR structure. These last two species split from *D. melanogaster* around 12-14 MYA (Lemeunier et al. 1986). A parsimonious consequence of these observations is that the long TIR in ancestor *Bari1* has been lost only in the lineage leading to the species of the melanogaster complex.

The *Bari2* subfamily is composed of defective copies only that have a long TIR structure, which arose before the divergence of Sophophora and Drosophila subgenera around 40 MYA (Russo et al. 1995). Thus, they represent relics of a diversified form of *Bari* element with a distinct evolutionary history. The ~250-bp TIR sequence of *Bari2* and *Bari1* (*ana* element) is poorly conserved except for the regions surrounding the DR motifs, the first 27 and the last 50 nucleotides.

Within the *Bari3* clade, full length and defective copies exist and all possess long TIRs. The TIRs of *Bari3* elements show very little nucleotide global identity with the sequence of *Bari2* and *Bari1* (*ana* element). Thus, the evolutionary history of *Bari3* elements has also been accompanied by a severe modification in the nucleotide sequence in the TIR, but still maintained a long TIR architecture. Interestingly, in *D. mojavensis*, there are 11 copies of *moja* elements that are 100% identical to each other and all have 100% nucleotide identity in their left and right TIRs. This situation does not occur with the related *pseu*, *pers* and *willi* elements where the left and right TIRs are 80, 80 and 96% identical to each other, respectively. We hypothesize that the *moja* element is a new and recent form of the *Bari* transposon that is in the initial state of its evolution. In this context, the comparison of *pers* and *pseu* elements appears relevant. These two elements were found in two species, *D. persimilis* and *D. pseudoobscura* that split from a common ancestor 1 MYA (Wang and Hey 1996). These two elements are nearly identical to each other, a situation similar to the *Bari1* elements in the species of the melanogaster complex. Both elements possess long TIRs with 80% global nucleotide identity and a complete transposase ORF, yet the number of full-length copies is 5 in *D. persimilis* and only 1 in *D. pseudoobscura* (Table 3). There are also numerous defective copies in both genomes that harbour only one of the two TIRs, and in addition, the traces of TIR sequences (HSP) are higher in *D. persimilis* than in *D. pseudoobscura*. We interpret these data as an example of different evolutionary events occurring in two strictly related organisms. In other words, *pers* and *pseu* are in an evolutionary phase in which they are losing the complete identity within their TIRs, apparently with a different rate in the two species.

Role of the conserved DR motif in full-length Bari-like elements

We have shown that within the *Bari* members that belong to the three subfamilies and possess a complete TIR, a significant conservation in both position and sequence occurs in three locations, the DR motifs (Fig. 5). Since the DRs are reported to be the binding site of the transposase (Fisher et al. 1999; Cui et al. 2002; Izsvák et al. 2002), their presence could be necessary for the transposition mechanism of autonomous, full-length copies of the transposon. At present, the only functional TLE that posseses a similar three DR structure is *Sleeping Beauty* (*SB*), an in vitro reconstructed transposon derived from naturally occurring and defective elements present in the genome of teleost fishes (Ivics et al. 1997). Molecular studies have revealed the role of an individual DR sequence in the mechanism of *SB* transposition (Izsvák et al. 2002). The proposed model predicts a DNA-protein complex, where a tetrameric form of the transposase interacts with the external and internal DRs of both ends (Lo, Li, Ro and Ri). The stability and the efficiency of transposition are enhanced by the presence of the Lm (and may be Rm), which appears not involved directly in the DNA-protein interaction. Since *SB* has long TIRs whereas *Bari1* has short TIRs, the long TIR structure does not appear necessary per se to the transposition as it was previously suggested (Plasterk et al. 1999); what matters for transposition is the presence of the three DRs located in the right position at each end. So, why has full-length *Bari1* lost the long TIR structure whereas full-length copies of *Bari3* (*moja* element) have long TIRs? A simple explanation is that the presence or absence of long TIR structure in a subfamily is a stochastic event that may or may not occur depending on the time and mode of invasion in a species. This appears quite improbable because there is a large excess of long TIR structures over the short one. An alternative hypothesis derives from the observation that the vast majority of defective *Bari*-like elements have a long TIR structure, indicating an intrinsic instability associated with this organization. Therefore, we hypothesize that *Bari1* would represent an "old" element that experienced the instability associated with the presence of long TIRs and hence has eliminated the source of its instability with time. On the other end, as previously discussed, the *Bari3* (*moja*) element is a "young" element still in its evolutionary infancy.

Role of the conserved DR motif in defective Bari-like elements

The presence of conserved DR motifs in defective *Bari*-like elements suggests a functional constraint unrelated to transposition because these copies are non-autonomous. Population genetics models of competition between autonomous and non-autonomous elements predict the persistence of defective copies at the expense of autonomous ones (Le Rouzic and Capy 2006). In addition, defective copies may play a role in regulating transposition of autonomous elements. For example, transposase monomers encoded by the

defective copies can compete with wild-type transposase during the assembly of the multimeric transposase complex (dominant-negative complementation) (Lohe and Hartl 1996). In addition, the presence of transposase-binding sites in the defective copies can titrate the wild-type transposase by acting as a sink. In both cases, the regulatory contribution of defective copies will operate in genomes where both full length and defective elements coexist as in most species of the melanogaster complex, which harbour both full length *Bari1* and defective *Bari2* copies. If this reflects intermediate steps in the *Bari* evolution, other species which are devoid of full-length *Bari* elements would be the final outcome. The question then arises as to what forces give rise to conservation of DRs. One possibility is that DRs represent the target site for host-encoded proteins. For example, the DNA-binding site of CENP-B mammalian centromeric protein has similarity with the TIR of *pogo*, another subclass of the *Tc1-mariner* superfamily (Kipling and Warburton 1997). Another possibility stems from recent finding of master loci that are made to a great extent of TE remnants and that control germline transposition by Piwi-interacting RNAs (piRNAs) (Brennecke et al. 2007, Gunawardane et al. 2007). A key feature of this control is sequence conservation because the role of the RNA in the complex is specific recognition of intermediates of transposition. Consistent with this interpretation, defective copies of *Tc1* elements in teleost fishes are abundantly transcribed in sense and antisense orientation, a prerequizite for the formation of double-stranded RNA in the RNA interference system (Krasnov et al. 2005; Sijen and Plasterk 2003). Also consistent with this interpretation is our finding that six out of the 39 *Bari*-related sequences in the piRNA database (GEO accession number GSE6734) correspond to the DRs.

Acknowledgments We thank C. Caggese for discussion and S. Marconi, F. Manodoro for technical support. P. D'Addabbo helped execute the figures. This investigation received financial support from the University of Bari (Progetti di Ateneo) and from Ministero dell'Istruzione, Università e Ricerca, Italy (PRIN).

References

Arcà B, Savakis C (2000) Distribution of transposable element *Minos* in the genus Drosophila. Genetica 108:263–267

Atschul SF, Madden TL, Scaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gappped BLAST and PSI-BLAST: a new generation of protein database search program. Nucleic Acids Res 25:3389–3402

Biémont C, Cizeron G (1999) Distribution of transposable elements in Drosophila species. Genetica 105:43–62

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell 128:1–15

Brillet B, Bigot Y, Augé-Gouillou C (2007) Assembly of the Tc1 and mariner transposition initiation complexes depends on the origin of their transposase DNA binding domains. Genetica 130:105–120

Caizzi R, Caggese C, Pimpinelli S (1993) Bari-1, a new transposon like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics 133:335–345

Cui Z, Geurts AM, Liu G, Kaufman CD, Hackett PB (2002) Stucture-Function analysis of the inverted terminal repeats of the sleeping beauty transposon. J Mol Biol 318:1221–1235

Finnegan DJ (1992) Transposable elements. Curr Opin Genet Dev 2:861–867

Fisher SE, van Luenen HGAM, Plasterk RHA (1999) Cis requirements for transposition of Tc1-like transposons in C. elegans. Mol Gen Genet 262:268–274

Franz G, Savakis C (1991) Minos, a new transposable element from *Drosophila hydei*, is a member of the Tc1 family of transposons. Nucleic Acids Res 19:6646

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC (2007) A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5´ End Formation in Drosophila. Science 315:1587–1590

Ivics Z, Hackett PB, Plasterk RH, Izsvák Z (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91:501–510

Izsvák Z, Ivics Z, Hackett PB (1995) Characterization of a Tc1-like transposable element in zebrafish (*Danio rerio*). Mol Gen Genet 247:312–322

Izsvák Z, Ivics Z, Plasterk RH (2000) Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. J Mol Biol 302:93–102

Izsvák Z, Khare D, Behlke J, Heinemann U, Plasterk RH, Ivics Z (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. J Biol Chem 277:34581–34588

Junakovic N, Di Franco C, Terrinoni A (1997) Evidence for a host role in regulating the activity of transposable elements in *Drosophila melanogaster*: the case of the persistent instability of Bari 1 elements in Charolles stock. Genetica 100:149–154

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M, Celniker SE (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol 3:research0084.1–0084.20

Ke Z, Grossman GL, Cornel AJ, Collins FH (1996) Quetzal: A transposon of the Tc1 family in the mosquito *Anopheles almimanus*. Genetica 98:141–147

Kidwell MG, Lisch D (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55:1–24

Kipling D, Warburton PE (1997) Centromeres, CENP-B and Tigger too. Trends Genet 13:141–145

Krasnov A, Koskinen H, Afanasyev S, Molsa H (2005) Transcribed Tc1-like transposon in salmonid fish. BMC Genomics 6: article n.107

Lam WL, Seo P, Robison K, Virk S, Gilbert W (1996) Discovery of amphibian Tc1-like transposon family. J Mol Biol 257:359–366

Lampe DJ, Churchill MEA, Robertson HM (1996) Purified mariner transposase is sufficient to mediate transposition in vitro. EMBO J 15:5470–5479

Lemeunier F, David JR, Tsacas L, Ashburner M (1986) The melanogaster species group. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of drosophila. New York Academic Press, New York, pp 147–256

Lerat E, Rizzon C, Biémont C (2003) Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. Genome Res 13:1889–1896

Le Rouzic A, Capy P (2006) Population genetics models of competition between transposable element subfamilies. Genetics 174:785–793

Li YJ, Satta Y, Takahata N (1999) Paleo-demography of the *Drosophila melanogaster* subgroup: Application of the maximum likelihood method. Genes Genet Syst 74:117–127

Lohe AR, Hartl DL (1996) Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. Mol Biol Evol 13:549–555

Lohe AR, Morijyama EN, Lidholm DA, Hartl DL (1995) Horizontal transmission, Vertical inactivation, and Stochastic loss of *mariner*-like transposable elements. Mol Biol Evol 12:62–72

Merriman PJ, Grimes CD, Ambroziak J, Hackett DA, Skinner P, Simmons MJ (1995) S elements: a family of Tc1-like transposons in the genome of *Drosophila melanogaster*. Genetics 141:1425–1438

Moschetti R, Caggese C, Barsanti P, Caizzi R (1998) Intra- and interspecies variation among Bari-1 elements of the melanogaster species group. Genetics 150:239–250

Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER (1995) Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. Proc Natl Acad Sci USA 92:8050–8054

Plasterk RHA, Izsvak Z, Ivics Z (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet 15:326–332

Radice AD, Bugaj B, Fitch DH, Emmons SW (1994) Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. Mol Gen Genet 244:606–612

Russo CAM, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol 12:391–404

Saitou N, Nei M (1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sijen T, Plasterk RHA (2003) Transposon silencing in the *Caenorhabtisis elegans* germ line by natural RNAi. Nature 426:310–314

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vos JC, De Baere I, Plasterk RHA (1996) Transposase is the only nematode protein required for in vitro transposition of Tc1. Genes Dev 10:755–761

Wang R, Hey J (1996) The Speciation history of *Drosophila pseudoobscura* and Close Relatives: Inferences from DNA Sequence Variation at the Period Locus. Genetics 144:1113–1126

Zayed H, Izsvak Z, Khare D, Heinemann U, Ivics Z (2003) The DNA-binding protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nucleic Acids Res 31:2313–2322