Current Zoology, 2016, 62(6), 667–674
doi: 10.1093/cz/zow095
Advance Access Publication Date: 11 September 2016
Article

Article

Rare horizontal transmission does not hide long-term inheritance of SINE highly conserved domains in the metazoan evolution

Andrea Luchetti* **and Barbara Mantovani**

Dipartimento di Scienze Biologiche, Geologiche e Ambientali—Università di Bologna, Via Selmi 3, Bologna 40126, Italy

*Address correspondence to Andrea Luchetti. E-mail: andrea.luchetti@unibo.it.

Received on 15 February 2016; accepted on 5 August 2016

Abstract

Transposable elements (TEs) are self-replicating, mobile DNA sequences which constitute a significant fraction of eukaryotic genomes. They are generally considered selfish DNA, as their replication and random insertion may have deleterious effects on genome functionalities, although some beneficial effects and evolutionary potential have been recognized. Short interspersed elements (SINEs) are non-autonomous TEs with a modular structure: a small RNA-related head, a body, and a long interspersed element-related tail. Despite their high turnover rate and *de novo* emergence, the body may retain highly conserved domains (HCDs) shared among divergent SINE families: in metazoans, at least nine HCD-SINEs have been recognized. Data mining on public molecular databases allowed the retrieval of 16 new HCD-SINE families from cnidarian, molluscs, arthropods, and vertebrates. Tracking the ancestry of HCDs on the metazoan phylogeny revealed that some of them date back to the Radiata–Bilateria split. Moreover, phylogenetic and age *versus* divergence analyses of the most ancient HCDs suggested that long-term vertical inheritance is the rule, with few horizontal transfer events. We suggest that the evolutionary conservation of HCDs may be linked to their potential to serve as recombination hotspots. This indirectly affects host genomes by maintaining active and diverse SINE lineages, whose insertions may impact (either positively or negatively) on the evolution of the genome.

Key words: horizontal transfer, Metazoan genome, retrotransposons, selfish DNA, SINEs highly conserved domain, vertical inheritance.

The bulk of eukaryotic genomes is composed of repeated DNA sequences, either arranged in tandem or interspersed in the genome (Richard et al. 2008). Interspersed repeats comprise transposable elements (TEs): DNA sequences able to move in different chromosomal locations and to increase their copy number. They can be classified into two major categories: Class I elements replicate via an RNA intermediate (retrotransposons) and Class II elements move via DNA intermediates (transposons; Wicker et al. 2007). In both classes most elements are autonomous, that is, they code for the enzymatic machinery necessary for their replication/mobilization. On the other hand, non-autonomous elements occur: they can replicate by using the enzymes coded by autonomous TEs (Ohshima and Okada 2005; Yang et al. 2009).

TEs are generally considered selfish DNA, as they replicate independently, regardless of the consequences for host species survival or, in more general terms, for costs they may impose on host fitness (Doolittle and Sapienza 1980; Orgel and Crick 1980; Hua-Van et al. 2011; Werren 2011; Ågren and Wright 2015). It is well known, in fact, that their random genomic insertions may cause severe alterations of genomic functions by disrupting gene structures, inducing gene down-regulation or even causing non-homologous recombination (Werren 2011). On the other hand, a number of studies highlighted positive interactions: for example, some TE insertions have been reported to have a major role in post-transcriptional gene expression regulation (Zhang et al. 2015), adaptation (Gonzalez et al. 2010), or even in the emergence of evolutionary novelties (Okada et al. 2010). Although the selfish nature of TEs has been recently debated (Fedoroff 2012), Orgel and Crick (1980) already suggested that just because "some selfish DNA may acquire a useful function and confer a selective advantage on the organism" this is not sufficient to question their selfishness, in the same way that endosymbionts such as *Wolbachia* are considered selfish genetic elements despite occasionally providing benefits to the host. The high representation of TEs in eukaryotic genomes may unavoidably lead to the domestication of some of these sequences for cellular functions, a process known as exaptation (Gould and Vrba 1982; Brosius and Gould 1992; Werren 2011). As recently argued, TEs may have positive interactions in some instances but their insertions can be still considered a threat for the host genome (Ågren 2014).

Here, we analyze a specific instance of selfish DNA: short interspersed elements (SINEs), non-autonomous class I retrotransposons, carrying evolutionary conserved domains. Their sequence structure is made up by 3 main modules: the head, the body, and the tail. The head derives from a small RNA (usually a tRNA or a 5S rRNA) and contains RNA pol III promoters for transcription. The tail is homologous to the tail of the long interspersed element (LINE) autonomous partner and allows the capture of LINE-encoded reverse transcriptase used for SINE mobilization. The body is the sequence between the head and the tail; it has no obvious functional meaning and its origin is often undetermined. SINEs are usually considered as elements with a high genomic turnover rate and *de novo* emergence (Kramerov and Vassetzky 2011): in fact, it is rare to find the same SINE family in different species. This is due also to their ability to exchange functional modules, for example replacing the head or the tail which would allow the maintenance of the retrotransposition competence (Takahashi and Okada 2002; Kramerov and Vassetzky 2011; Luchetti and Mantovani 2013a). On the other hand, in some remarkable instances, body domains have been found to be conserved among different SINE families, even distributed across distantly related organisms: these are termed highly conserved domains (HCDs). To date, 9 HCDs have been characterized in metazoan genomes: CORE, V, Nin, Deu, Inv, Ceph, Pln, Meta, and Mesc (Gilbert and Labuda 1999, 2000; Ogiwara et al. 2002; Nishihara et al. 2006, 2016; Akasaki et al. 2010; Piskurek and Jackson 2011; Luchetti and Mantovani 2013a; Matetovici et al. 2016). The taxonomic distribution of HCD-SINEs can be either restricted to few related species, such as Ceph-SINEs in cephalopods (Akasaki et al. 2010), or it can be widespread across the animal kingdom, such as Nin-SINEs (Piskurek and Jackson 2011) (Table 1). Moreover, they may share a common and complex origin: for example, Nin is part of the larger Deu domain (Piskurek and Jackson 2011) and Inv is the 5′-end part of Nin (Luchetti and Mantovani 2013a). Two HCDs can be also combined in the same SINE, such as CORE + Deu in the lancelet Bf1SINE1 element (Nishihara et al. 2006) or Inv + Pln in cricket, termite, and viviparous cockroach SINEs (Luchetti and Mantovani 2013a).

Taking into account the rapid evolutionary rate of most SINEs, why HCDs should be conserved for such vast timescales is a major enigma. Three main hypotheses have been proposed to explain such conservation: i) HCDs are important for SINE–LINE tail exchange (Gilbert and Labuda 2000), ii) HCDs are recombination hotspots facilitating SINE module exchanges (Luchetti and Mantovani 2013a), and iii) HCDs have some functional meaning conferring advantages to the host genome (Nishihara et al. 2006; Deragon 2012). None of the three hypotheses has been yet demonstrated. The first explanation appears less likely because no such domains have been found in LINE sequences; in addition, LINEs are fast evolving sequences and, therefore, domains should diverge rapidly (Deragon 2012). The other 2 models, though, appear to be more reliable and are not mutually exclusive.

Table 1. SINEs' HCDs distribution (Gilbert and Labuda 1999, 2000; Ogiwara et al. 2002; Nishihara et al. 2006, 2016; Akasaki et al. 2010; Piskurek and Jackson 2011; Luchetti and Mantovani 2013a; Matetovici et al. 2016)

	CORE	V	Nin	Deu	Inv	Ceph	Pln	Meta	Mesc
Chordata	+	+	+	+				+	
Hemichordata					+				
Echinodermata	+		+	+	+			+	
Arthropoda			+		+		+		
Mollusca	+	+	+			+		+	+
Annelida			+					+	
Cnidaria		+	+					+	

To have a better understanding of HCD-SINEs and their possible impact on the host genome, it is first necessary to have a clear picture of their evolution and inheritance. Presently known HCDs (Table 1) show a patchy taxonomic distribution across distantly related animal phyla, suggestive of horizontal (lateral) transfer (i.e., the passage of genetic material beyond reproductive boundaries; Schaak et al. 2010). The distinction between vertical inheritance and horizontal transfer is, therefore, crucial to understand whether the phylogenetic distribution of HCDs is the result of a recent origin by lateral spreading or due to a long-term vertical inheritance. We, therefore, investigated the HCD-SINEs occurrence in public database releases of genomic resources, and studied their phylogenetic relationships together with the possibility of lateral transfers.

Materials and Methods

New HCD-SINEs were searched for by BLAST (Altschul et al. 1990) analysis of GenBank databases nt, est, tsa, wgs (accessed on September 2015), using known HCD consensus sequences obtained from Vassetzky and Kramerov (2013). BLAST search was performed with the *blastn* algorithm (default parameters), and all scored BLAST alignments with length > 20 bp have been considered. Database sequences showing similarity with the same HCD consensus sequence, and obtained from the same taxon, were then aligned. Regions of homology, which likely correspond to interspersed repeats, were identified on the basis of nucleotide similarity. Sequence alignments were carried out with MAFFT v.7 (Katoh and Standley 2013), using L-INS-i parameters, and 50%-majority rule consensus sequences were obtained using Seaview (Gouy et al. 2010). GenBank accession numbers of scored database sequences are given in Supplementary Table S1 and new SINE consensus sequences are in Supplementary Figure S1. These interspersed repeats were then classified as SINEs on the basis of homology with a small RNA gene at the 5′-end and of the AT-rich tail at the 3′-end (Kramerov and Vassetzky 2011). In order to check for homology with previously known elements, consensus sequences were used to search in Repbase Update (Bao et al. 2015; accessed on September 2015).

Table 2. Taxonomic distribution and structural features of the newly identified HCD-SINEs

Phylum/Class	Species	HCD	SINE	*N*	Consensus length (bp)
Chordata/Amphibia	*Rana clamitans*	V	*Racla*	10	329
Echinodermata/Asteroidea	*Patiria miniata*	CORE	*Pami*	9	288
Arthropoda/Insecta	*Thermobia domestica*	V	*Thedo*	10	246
Arthropoda/Insecta	*Empusa pennata*	V	*Empe*	6	241
Arthropoda/Crustacea	*Portunus trituberculatus*	CORE	*Potri*	14	283
Arthropoda/Crustacea	*Petrolisthes cinctipes*	CORE	*Peci*	5	288
Arthropoda/Crustacea	*Gandalfus yunohana*	CORE	*Gayu*	4	355
Arthropoda/Crustacea	*Procambarus clarckii*	CORE	*Procla*	11	238
Arthropoda/Crustacea	*Homarus americanus*	CORE	*Homa*	4	351
Arthropoda/Chelicerata	*Limulus polyphemus*	CORE	*Lipo*	6	265
Mollusca/Scaphopoda	*Antalis entalis*	CORE + V	*Ante*	43*	388
Mollusca/Gastropoda	*Aplysia californica*	CORE	*Aply*	7	353
Mollusca/Gastropoda	*Lymnaea stagnalis*	Ceph	*Lyst*	5	197
Mollusca/Bivalvia	*Crassostrea gigas*	V	*Crag*	10	259
Mollusca/Bivalvia	*Elliptio complanata*	CORE	*Elco*	4	179
Cnidaria/Anthozoa	*Stylophora pistillata*	CORE	*Styp*	7	296

Notes: HCD, highly conserved domain; *N*: number of sequences retrieved carrying SINEs; *, Roche 454 sequencer reads, only one read span the entire *Ante* length.

Maximum Likelihood phylogenetic trees were calculated on Nin, V, CORE, and Meta domains sequence datasets, using K80+Γ (Nin, V, CORE) or TN92+Γ+I (Meta) as best estimated substitution models and 500 bootstrap replicates for nodal support, through MEGA v.6 (Tamura et al. 2013). All Nin domain sequences taken from Piskurek and Jackson (2011) were used: AmnSINE (*Homo sapiens* and *Gallus gallus*), SINE3-1a (*Danio rerio*), SINE3_IP (*Ictalurus punctatus*), OSSINE1 (*Oncorhynchus mykiss*), LmeSINE1a (*Latimeria menadoensis*), SacSINE1 (*Squalus acanthias*), EbuSINE1 (*Eptatretus burgeri*), Bf1SINE1 (*Branchiostoma floridae*), SINE2-3_SP (*Strongylocentrotus purpuratus*), Isc-Nin-DC (*Ixodes scapularis*), Lgi-Nin-DC (*Lottia gigantea*), Aca-Nin-DC (*Aplysia californica*), Cte-Nin-DC (*Capitella teleta*), and Nve-Nin-DC (*Nematostella vectensis*). For comparisons with presently found CORE- and V-SINEs, we included in the analyses the following elements drawn from Repbase Update. CORE domain dataset: MIR3 (*H. sapiens*), MON1 (*Ornithorhynchus anatinus*), CoeSINE3 (*Latimeria chalumnae*), MIR-Xt (*Xenopus tropicalis*), Bf1SINE1 (*B. floridae*), SINE2-4c_SP (*S. purpuratus*), OR2 (*Octopus vulgaris*); moreover, recently found elements BivaCORE-SINE1 (*Ruditapes decussatus*) and BivaCORE-SINE2 (*Mizuhopecten yessoensis*) have been added (Nishihara et al. 2016). The V domain dataset consisted of *Dana* (*D. rerio*), SINE2-1_XT (*X. tropicalis*), SINE2-1_NV (*N. vectensis*), and SINE2-2_Adi (*Adineta vaga*) and 5 RUDI V-SINEs from *Haliotis discus* (Gastropoda), *Ruditapes philippinarum*, *Tegillarca granosa*, *Yoldia limatula* (Bivalvia), and *Neomenia megatrapezata* (Solenogastres) (Luchetti et al. 2016). Finally, we also considered the V domain contained in the *H. sapiens* MER6 DNA transposon (Smit and Riggs 1996). The Meta domain dataset taken from Nishihara et al. (2016) was also analyzed but the phylogenetic tree was completely unresolved (Supplementary Figure S2): it was, therefore, not considered further.

An age *versus* divergence analysis was carried out on the widely distributed Nin, V, and CORE HCDs. Briefly, this analysis predicts a correlation between the taxa split ages and molecular divergence; the longer the time since 2 lineages split, the more divergence there should be between SINEs HCD. Hence, genetic distances between HCDs were plotted against the split ages of the species carrying the SINEs. If 2 sequences appear less divergent than expected on the basis of split age, a horizontal transfer event is suggested; in the scatterplot, the relative point falls below the regression line. On the contrary, a divergence higher than expected is likely to be due to a comparison between paralogous lineages. These are distinct SINE lineages sharing a common ancestry within the same host genome and that diverged before the host species split. In the plot, the point appears above the regression line. In order to identify potential comparisons significantly deviating from the regression model (i.e., significantly above or below the regression line), we used the method described in Biedler et al. (2015): studentized deleted residuals (RStudent) are calculated for each regression and absolute RStudent values >2.0 are considered as indicating significant outlier comparisons.

Species split ages were obtained from TimeTree database (accessed in July 2016; Hedges et al. 2006).

Results

Characterization of new SINEs

Sixteen new SINEs were isolated from 5 different phyla: 1 element from Cnidaria, 5 from Mollusca, 8 from Arthropoda, 1 from Echinodermata, and 1 from Chordata. SINE consensus sequences were obtained from multiple copies per species (from 4 to 14; 43 Roche 454 reads in the case of *Ante* element) and showed lengths ranging from 179 bp (*Elco* elements from *Elliptio complanata*; Mollusca, Bivalvia) to 388 bp (*Ante* element from *Antalis entalis*; Mollusca, Scaphopoda) (Table 2; Figure 1; Supplementary Figure S1).

We checked in the RepBase Update if there were already known SINEs similar to those retrieved in the present analysis: when identities were scored, they never involved the whole SINE sequence but they were only limited to regions of SINE sequences that contained HCDs or, in some instances, the RNA pol III promoters. These are obvious homologies as HCDs are similar by definition, and the RNA-related head is expected to be highly conserved due to functional constraints. Excluding these similarities, no other part (e.g., the part of the body excluding the HCD or the 3′-end tail) of retrieved elements were similar to any other known SINE. We, therefore, concluded that all 16 newly identified SINEs are novel elements.

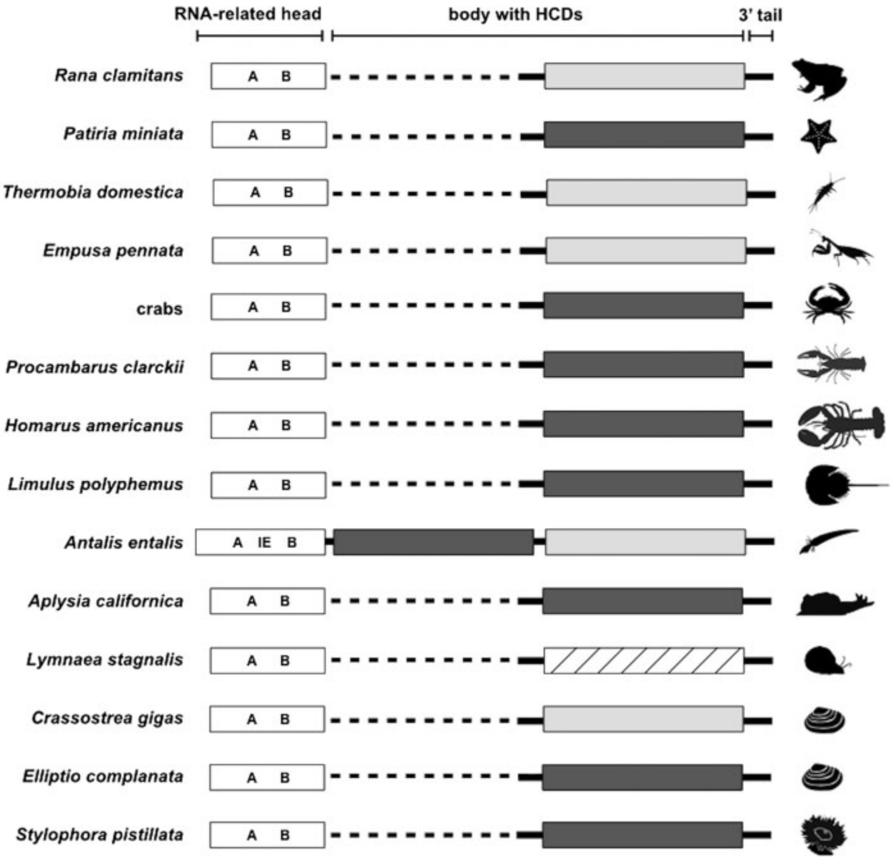

Figure 1. Schematic drawing of newly retrieved SINEs structures. Within the head module RNA pol III promoters are reported as: boxes A + B (tRNA-related head) or A + IE + B (5S rRNA-related head). Filled boxes represent highly conserved domains: CORE (dark gray), V (light gray), and Ceph (diagonal lines pattern). Drawings are not to scale. Crabs species are *Portunus trituberculatus*, *Petrolisthes cinctipes*, and *Gandalfus yunohana*.

All new SINEs had a tRNA-related head and conserved RNA pol III promoters (A and B boxes), with the only exception of the *Ante* element which has a 5S RNA-derived head (Figure 1; Supplementary Figure S1). Four out of the 16 elements retrieved showed the V domain, 10 showed the CORE domain, and 1 showed the Ceph domain. Moreover, the scaphopod SINE, *Ante*, showed both the CORE and the V domains, separated by a 35-bp spacer sequence (Figure 1; Supplementary Figure S1).

Furthermore, the mantis element (*Empe*) showed an extensive similarity with the termite SINE *Talub* in RepBase Update, spanning from the 5′-end to half of the sequence. This region corresponds to the tRNA-related head and part of the body: sequence alignment inspection indicated that *Talub* contains the V domain, a feature not previously recognized (Luchetti and Mantovani 2011). It was, therefore, added to the analysis.

The overall similarities of the newly retrieved V domains (also including the *Talub* one) was 61.5%, while the CORE domain showed an overall similarity of 63.2%.

Taxonomic distribution and evolution of HCD-SINEs in the animal kingdom

We tracked the ancestry of each HCD by plotting their emergence along the metazoan phylogeny, on the basis of the observed taxonomic distribution (Figure 2). In order to have a comprehensive analysis, beside the presently characterized elements, we included the taxonomic distribution of already described HCD-SINEs (Table 1; Figure 2).

On the whole, these data indicate a slightly different distributional pattern for the analyzed HCDs (Figure 2). While some domains (Pln, Ceph, and Deu) are restricted to phylogenetically related taxa, others (Nin, V, CORE, Meta) appear widely distributed and have apparently been established since the Radiata/Bilateria split.

Phylogeny and putative horizontal transfer events of Nin, V, and CORE domains

To explore the evolutionary relationships between HCDs from the same superfamily, we performed phylogenetic analyses on the nucleotide sequence of the widely distributed Nin, V, and CORE domains. As a general consideration, Maximum Likelihood trees obtained in these analyses showed low or no support at all at the deepest nodes and a weak congruence, mainly in restricted clades, with the host species phylogeny (Figure 3). On the other hand, some peculiar clustering emerged.

The tree based on the Nin domain dataset supports a relationship between sequences from bony fishes and tetrapods, although with a weak bootstrap value (Figure 3A). Moreover, these sequences are included in a wider clade including also HCDs from SINEs of the shark *Squalus acanthias*, the lancelet *B. floridae*, the sea urchin *S. purpuratus*, and the annelid *C. teleta*. It is to be noted that Nin domains from the lancelet (Bf1SINE1) and from the sea urchin (SINE2-3-SP) elements cluster with very high nodal support. The remaining relationships are unresolved.

The phylogenetic analysis of the V domain dataset shows a clustering pattern generally more congruent with the host species

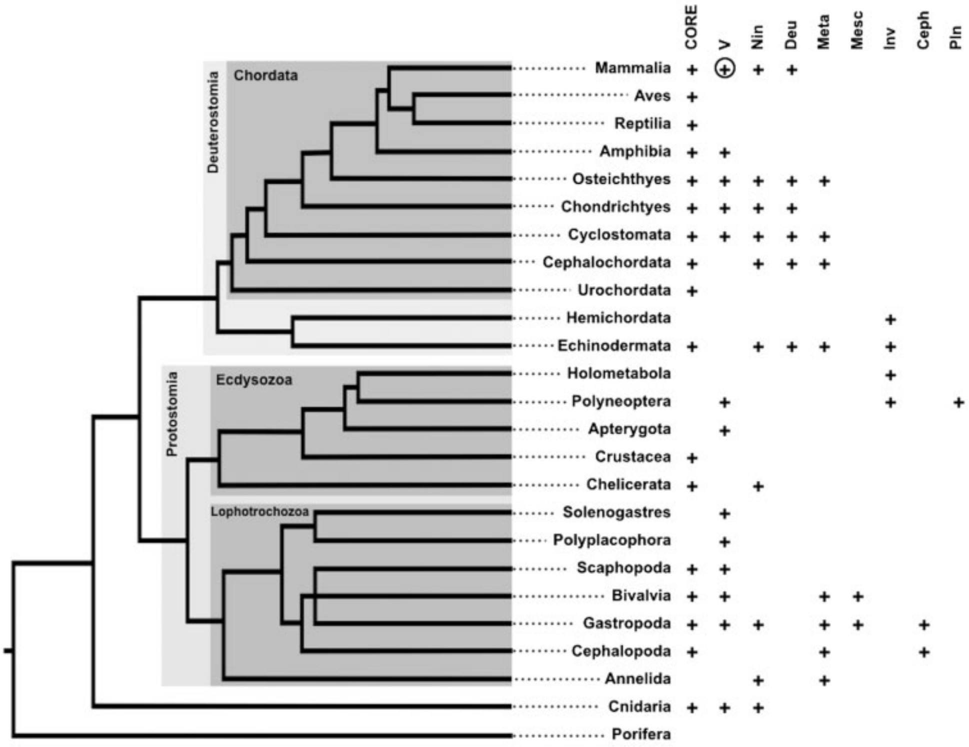

Figure 2. Metazoan phylogenetic tree with presence (+) of listed highly conserved domains in the relative taxa. The circle at mammalian V domain indicates that it is present within a DNA transposon instead of a SINE. Phylogenetic tree redrawn from Blair (2009).

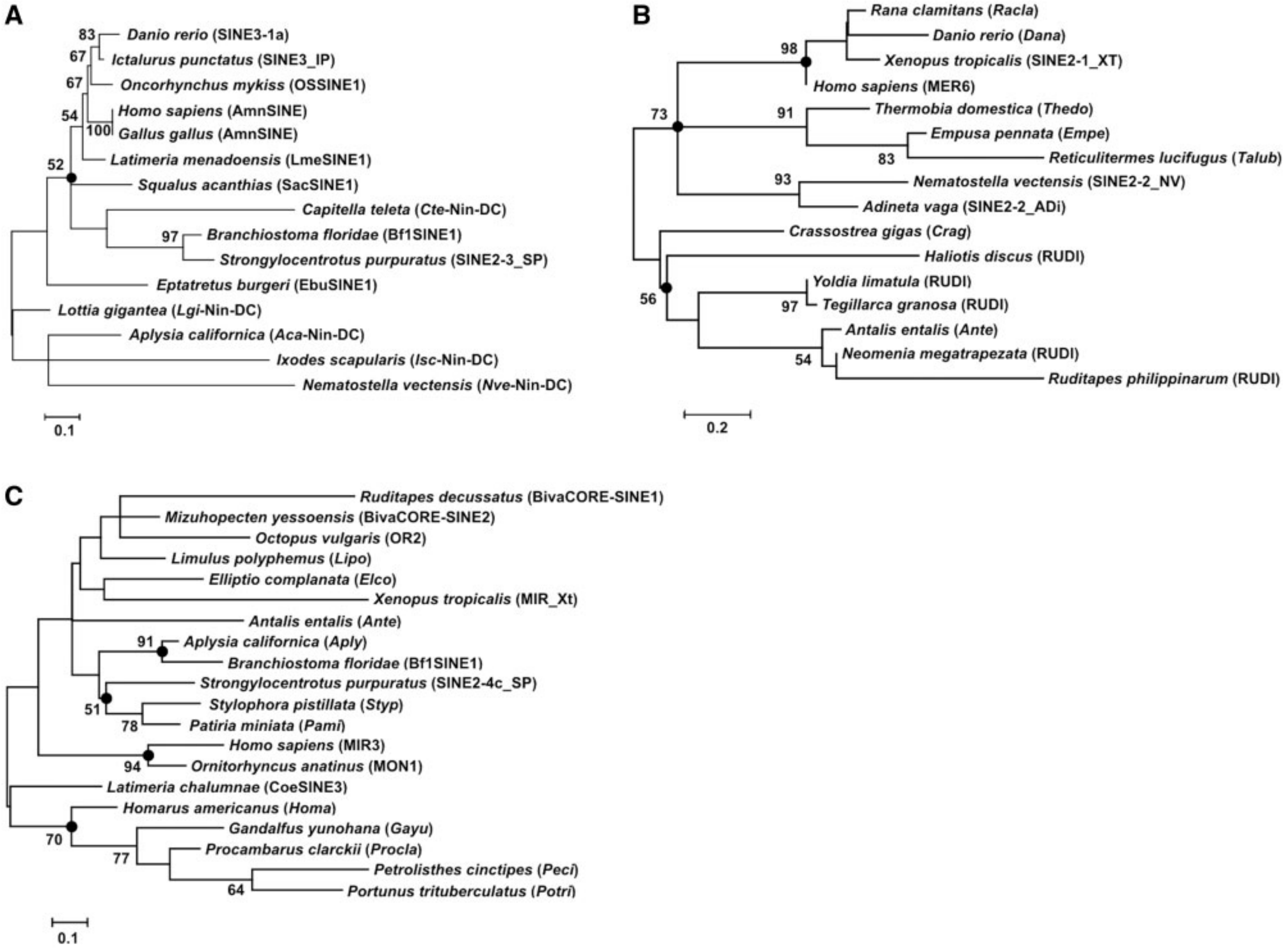
A
83
Danio rerio (SINE3-1a)
67
Ictalurus punctatus (SINE3_IP)
67
Oncorhynchus mykiss (OSSINE1)
54
Homo sapiens (AmnSINE)
100
Gallus gallus (AmnSINE)
Latimeria menadoensis (LmeSINE1)
52
Squalus acanthias (SacSINE1)
Capitella teleta (Cte-Nin-DC)
97
Branchiostoma floridae (Bf1SINE1)
Strongylocentrotus purpuratus (SINE2-3_SP)
Eptatretus burgeri (EbuSINE1)
Lottia gigantea (Lgi-Nin-DC)
Aplysia californica (Aca-Nin-DC)
Ixodes scapularis (Isc-Nin-DC)
Nematostella vectensis (Nve-Nin-DC)
0.1
B
Rana clamitans (Racla)
Danio rerio (Dana)
98
Xenopus tropicalis (SINE2-1_XT)
Homo sapiens (MER6)
73
91
Thermobia domestica (Thedo)
Empusa pennata (Empe)
83
Reticulitermes lucifugus (Talub)
93
Nematostella vectensis (SINE2-2_NV)
Adineta vaga (SINE2-2_ADi)
Crassostrea gigas (Crag)
Haliotis discus (RUDI)
56
Yoldia limatula (RUDI)
97
Tegillarca granosa (RUDI)
Antalis entalis (Ante)
54
Neomenia megatrapezata (RUDI)
Ruditapes philippinarum (RUDI)
0.2
C
Ruditapes decussatus (BivaCORE-SINE1)
Mizuhopecten yessoensis (BivaCORE-SINE2)
Octopus vulgaris (OR2)
Limulus polyphemus (Lipo)
Elliptio complanata (Elco)
Xenopus tropicalis (MIR_Xt)
Antalis entalis (Ante)
91
Aplysia californica (Aply)
Branchiostoma floridae (Bf1SINE1)
Strongylocentrotus purpuratus (SINE2-4c_SP)
51
Stylophora pistillata (Styp)
78
Patiria miniata (Pami)
Homo sapiens (MIR3)
94
Ornitorhyncus anatinus (MON1)
Latimeria chalumnae (CoeSINE3)
Homarus americanus (Homa)
70
Gandalfus yunohana (Gayu)
77
Procambarus clarckii (Procla)
Petrolisthes cinctipes (Peci)
64
Portunus trituberculatus (Potri)
0.1

Figure 3. Maximum Likelihood phylogenetic trees of Nin (**A**; $-\ln L = 1041.20$), V (**B**; $-\ln L = 1521.15$), and CORE (**C**; $-\ln L = 1341.16$) highly conserved domains. Nodes marked with black dots indicate clades analyzed in the age *versus* divergence analysis. Numbers at nodes represent bootstrap values >50%.

relationships (Figure 3B). In fact, we can observe a vertebrate-specific, an insect-specific, a cnidarian-specific, and a mollusc-specific clade. The only sequence unclearly allocated is the V domain from the oyster's *Crag* element. Within the insect clade, observed phylogenetic relationships are comparable with the host species phylogeny. On the contrary, within the mollusc clade, the cluster including the V domain of RUDI elements from the bivalve *R. philippinarum* and the solenogaster *N. megatrapezata*, and of the scaphopod SINE *Ante*, although only weakly supported, is in contrast with the species phylogeny.

In the phylogeny of the CORE domain dataset (Figure 3C), the following supported clusters can be observed: i) domains from lancelet (Bf1SINE1) and *A. californica* (*Aply*) elements; ii) domains of cnidarian (*Styp*) and echinoderm (*Pami* and SINE2-4c_SP) elements; iii) domains from human (MIR3) and platypus (MON1) SINEs; and iv) domains from all crustacean SINEs.

Given the Nin, V, and CORE HCDs wide taxonomic distribution, and that their phylogenetic relationships are often in disagreement with those of the host species, we performed an age *versus* divergence analysis in order to identify potential horizontal transfer events. Such horizontal transfers should be revealed by sequence divergences lower than that expected on the basis of the time since the host species split (Figure 4). However, in order to avoid inappropriate correlations, due to the rapid sequence evolution of SINEs, we have chosen a conservative approach and only considered comparisons between sequences falling within bootstrap-supported clades in the phylogenetic analysis.

Nin and V HCD datasets clearly exhibited a linear, positive relationship with the host species split age; on the other hand, the divergence of CORE sequences showed a slight trend for a negative correlation with host species split age, although the correlation is not significant. RStudent analysis identifies three comparisons that are potential outliers: crustacean *Gayu versus Potri* elements CORE domains (|RStudent| = 3.21), *Tegillarca granosa versus Yoldia limatula* RUDI elements V domain (|RStudent| = 2.02), termite *Talub versus* vertebrate elements V domains (|RStudent| = 2.30, 2.81, 2.17), and lancelet *versus* sea urchin SINEs Nin domain (|RStudent| = 5.20). Based on the position of comparisons with respect to the regression line (indicated by arrows in Figure 4), comparisons between lancelet/sea urchin Nin domain and *T. granosa*/*Y. limatula* RUDI's V domain resulted as less divergent than expected; therefore, they likely to indicate horizontal transfer events. On the contrary, all other outlier comparisons were more divergent than expected, indicating comparisons between paralogous lineages.

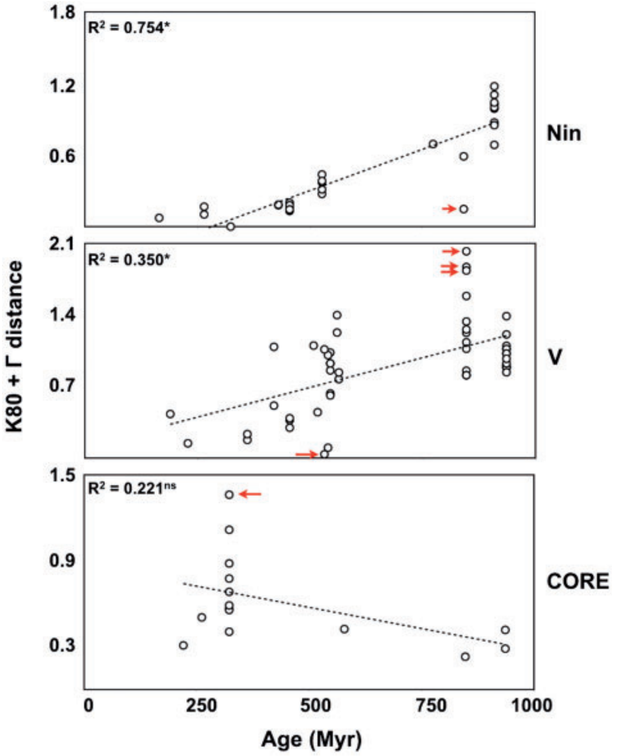

Figure 4. Age *versus* divergence analyses. In scatterplots, regression lines are dotted and arrows indicate outlier comparisons (|RStudent| >2). The correlation coefficient R^2 is given along with its significance (ns: not significant; $*P < 0.01$).

Discussion

We obtained new SINE elements belonging to 4 superfamilies, as determined by their HCD: 10 CORE-, 4 V-, and 1 Ceph-SINEs, and 1 element having both CORE and V domains. Sequence comparisons with already known SINEs indicated we found novel elements, and identified a previously known element (*Talub* from the termite *R. lucifugus*; Luchetti and Mantovani 2011) as a V-SINE. The new SINEs sequence structures perfectly match the canonical modular organization, with the exception of the *Ante* element that showed 2 distinct HCDs within its body: CORE + V. A similar arrangement has been found only in the Bf1SINE1 element from lancelet (CORE + Nin; Nishihara et al. 2006), and in the insect (cricket and blattarian) *Gbim* and *Taluc* elements (Inv + Pln; Luchetti and Mantovani 2013a).

Remarkably, we discovered CORE-SINEs in 2 phyla not previously known to harbor them: arthropods (horseshoe crab *L. polyphemus* and decapod crustaceans) and cnidarians (hood coral *S. pistillata*). The CORE domain outside vertebrates was only known previously in some molluscs (Gilbert and Labuda 1999) and in the sea urchin (Vassetzky and Kramerov 2013). We confirm these latter findings, having isolated CORE-SINEs from bivalves, gastropods, and scaphopods and in an echinoderm, the bat-star *P. miniata*. Similarly, we widened the known taxonomic distribution of the V domain, by finding V-SINEs in arthropods (three insect species: *Thermobia domestica*, *Empusa pennata*, and *Reticulitermes lucifugus*) and confirming their presence in molluscs and amphibians. It is to be noted that presently characterized CORE elements from gastropods and bivalves, and V-SINEs from the oyster *C. gigas* and insects have been also found by Matetovici et al. (2016) and Nishihara et al. (2016), who carried out the same search in parallel and independently from this study. Furthermore, we found the Ceph HCD, so far isolated only from cephalopods (Akasaki et al. 2010), in a gastropod: the snail *Lymnaea stagnalis*. We therefore propose to rename it as the "Cega" (cephalopod + gastropod) domain. Overall, our study enables a wider and more complete picture of HCD distribution than before (Gilbert and Labuda 1999, 2000; Ogiwara et al. 2002; Nishihara et al. 2006, 2016; Akasaki et al. 2010; Piskurek and Jackson 2011; Luchetti and Mantovani 2013a; Luchetti et al. 2016; Matetovici et al. 2016).

Tracking the emergence of each HCD on the animal phylogeny revealed an ancient origin for CORE, V, and Nin domains, together with the recently characterized Meta domain (Nishihara et al. 2016). In fact, if only vertical inheritance is considered, the origin of these HCDs would date back to the Radiata–Bilateria split within the Mid-Proterozoic (Hedges et al. 2006). Such an ancient origin has been hypothesized also for the non-long terminal repeats (LTR) element R2, a kingdom-wide retrotransposon found in animal genomes, able to specifically insert into the 28S rRNA gene (Kojima and Fujiwara 2005; Luchetti and Mantovani 2013b). R2 phylogeny does not always match that of the host species, as observed for other non-LTR elements (Malik et al. 1999), raising the question whether it has been vertically transmitted or not. However an age *versus* divergence analysis of R2 excluded the possibility of horizontal transfer, explaining the observed phylogenetic pattern as due to emergences and extinctions of paralogous lineages (Kojima and Fujiwara 2005).

Few examples of SINE horizontal transfer have yet been suggested: the SmaI family in coregonid and salmonid fishes (Hamada et al. 1997), a SINE shared by reptiles and mammals (Piskurek and Okada 2007), and a RUDI subfamily between distantly related bivalve species (Luchetti et al. 2016). We found that Nin and V domain divergence showed a positive correlation with host species split ages, with few deviations suggesting horizontal transfers and/or paralogous lineage comparisons. Horizontal transfers involved the Nin domain of SINEs from the lancelet and the sea urchin and the V domain from RUDI elements found in *T. granosa* and *Y. limatula*, an event already suggested on the basis of phylogenetic and divergence analyses of the whole SINE sequence (Luchetti et al. 2016). Instances of paralogous lineages comparisons of the V domain involved the termite elements *Talub* and sequences from *H. sapiens*, *X. tropicalis*, and *R. clamitans*. The higher divergence showed by the termite SINE lineage can also explain why it was not formerly identified as a V-SINE (Luchetti and Mantovani 2011). Beside these few outlier comparisons, the phylogenetic distribution of HCDs strongly confirms a pattern of vertical inheritance. However, the lack of a positive relationship between host species split ages and HCD divergences in the CORE domain dataset was unexpected. We suggest 2 non-exclusive possible explanations. First, among ancient CORE elements, nucleotide substitutions accumulated to a point that multiple, parallel mutations arose, hiding the true sequence divergence. Second, crustacean SINEs appeared to have experienced a higher substitution rate than other CORE elements.

From the analysis of Nin and V domains, 2 general conclusions can be drawn. First our data suggest that HCDs may undergo rare lateral transfer events. As HCDs are only part of a SINE, this implies that the SINE itself experienced the horizontal transfer. This can be exemplified by the case of the RUDI element in *T. granosa* and *Y. limatula*. Second, although the deepest nodes in the phylogenetic trees are mostly unresolved, there are some clades in which HCDs appear to have been conserved for extremely long periods of time. For example, the V domain shows a stable inheritance within insects, some molluscs (although note the *T. granosa*/*Y. limatula* RUDIs) and vertebrates (although in *H. sapiens* the HCD jumped into a DNA transposon; Smit and Riggs 1996). In the Nin domain analysis, such long-term inheritance is suggested by the monophyletic relationship between the HCD in the annelid *C. teleta* and those found in deuterostomes. Tracking the origins of each observed vertical transmission, the oldest estimates would date back to protostomes–deuterostomes split, about 850 Myr ago (Hedges et al. 2006).

Our analysis suggests that, beside a few lateral transfers and paralogous lineage turnovers, HCDs appeared to have undergone vertical inheritance and long-term conservation. Interestingly, although the apparently functionless HCD regions are highly conserved, other regions of SINE sequences, whether essential for retrotransposition or not, show no evidence of such long-term conservation (Gilbert and Labuda 1999, 2000; Ogiwara et al. 2002; Nishihara et al. 2006; Akasaki et al. 2010; Luchetti and Mantovani 2013a; this paper). It appears that the modular structure of SINEs allows the RNA-related head and the LINE-related tail to be switched. Thus, the same body module can be associated to different head and tails (Kramerov and Vassetzky 2011; Luchetti and Mantovani 2013a). These changes of head and tail may be vital for allowing the SINE to maintain retrotransposition competence despite potential selection on the rest of the genome to reduce mobilization of SINEs.

SINEs recombine frequently during the retrotransposition step, at the RNA level (Yadav et al. 2012). In the copy-choice model, recombination between viral RNAs seems to be promoted by sequence identity and secondary structure at breakpoints (Baird et al. 2006; Simon-Loriere and Holmes 2011). Thus, it is possible that the homology between the same HCDs in different SINEs may help the RNA–RNA pairing, allowing recombination and, eventually, promoting switches between modules. It is interesting to note that the Inv domain shows the ability to form conserved hairpin structures (Luchetti and Mantovani 2011): in this view, it may act as a signal for reverse transcriptase template switch (Baird et al. 2006). Interestingly, HCD-mediated recombination might maintain SINE retrotransposition ability also in the event of horizontal transfer: in fact, it may allow a newly transferred SINE to switch to a new tail by recombining with a resident SINE, eventually allowing the exploitation of resident LINE enzymes. Therefore, it is possible to hypothesize that the conservation of central SINE domains may cause the maintenance and promote diversity of active SINE lineages within genomes.

The conservation of specific domains within ubiquitous selfish genetic elements such as SINEs is likely to impact on host genome evolution. In fact, the very maintenance of SINE activity constitutes, *per se*, a possible cause of genome evolution: random insertions into genomic loci may lead to gene restructuring or changes in the expression profile (Gonzalez et al. 2010; Okada et al. 2010; Werren 2011; Deragon 2012; Zhang et al. 2015). Finally, HCDs may not only promote SINE–SINE recombination but may also represent a recombination hotspot for the host genome (Makałowski 2000; Werren 2011; Walters-Conte et al. 2014).

In conclusion, we have found evidence that the highly conserved regions (HCDs) within SINEs have been conserved for even longer than previously suggested, perhaps from the protostomes/deuterostomes split almost a billion years ago. However, we do find evidence for 2 horizontal transfers in SINEs. We suggest that this long-term conservation of apparently non-functional regions of SINEs may be due to advantages in interactions between SINEs, allowing elements to swap modules promoting their long-term persistence.

Supplementary Material

Supplementary material can be found at http://www.cz.oxfordjournals.org/.

Acknowledgments

We wish to thank Fabrizio Ghiselli for commenting on the manuscript. We are also grateful to Reviewers and Editors for their helpful and constructive criticisms and suggestions.

Funding

The work was funded by RFO-UNIBO and Canziani [to A.L. and B.M.].

References

Ågren JA, 2014. Evolutionary transitions in individuality: insights from transposable elements. *Trend Ecol Evol* **29**:90–96.

Ågren JA, Wright SI, 2015. Selfish genetic elements and plant genome size evolution. *Trend Plant Sci* **20**:195–196.

Akasaki T, Nikaido M, Nishihara H, Tsuchiya K, Segawa S et al., 2010. Characterization of a novel SINE superfamily from invertebrates: "Ceph-SINEs" from the genomes of squids and cuttlefish. *Gene* **454**:8–19.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–410.

Baird HA, Galetto R, Gao Y, Simon-Loriere E, Abreha M.A, 2006. Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. *Nucleic Acids Res* **34**:5203–5216.

Bao W, Kojima KK, Kohany O, 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**:11.

Biedler JK, Chen X, Tu Z, 2015. Horizontal transmission of an R4 clade non-long terminal repeat retrotransposon between the divergent *Aedes* and *Anopheles* mosquito genera. *Insect Mol Biol* **24**:331–337.

Blair JE, 2009. Animals (Metazoa). In: Hedges SB, Kumar S, editors. *The Timetree of Life*. Oxford: Oxford University Press, 223–230.

Brosius J, Gould SJ, 1992. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc Natl Acad Sci USA* **89**:10706–10710.

Deragon J-M, 2012. SINE exaptation as cellular regulators occurred numerous times during eukaryote evolution. In: Grandbastien M-A, Casacuberta JM, editors. *Plant Transposable Elements*. Berlin, Germany: Springer, 253–271.

Doolittle WF, Sapienza C, 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**:601–603.

Fedoroff NV, 2012. Transposable elements, epigenetics, and genome evolution. *Science* **338**:758–767.

Gilbert N, Labuda D, 1999. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci USA* **96**:2869–2874.

Gilbert N, Labuda D, 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. *J Mol Biol* **298**:365–377.

Gonzalez J, Karasov TL, Messer PW, Petrov DA, 2010. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet* **6**:e1000905.

Gould SJ, Vrba ES, 1982. Exaptation: a missing term in the science of form. *Paleobiology* **8**:4–15.

Gouy M, Guindon S, Gascuel O, 2010. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**:221–224.

Hamada M, Kido Y, Himberg M, Reist JD, Ying C et al., 1997. A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the *SmaI* family of repeats: possible evidence for the horizontal transfer of SINEs. *Genetics* **146**:355–367.

Hedges SB, Dudley J, Kumar S, 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**:2971–2972.

Hua-Van A, Le Rouzic A, Boutin T, Filée J, Capy P, 2011. The struggle for life of the genome's selfish architects. *Biol Direct* **6**:19.

Katoh K, Standley DM, 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.

Kojima KK, Fujiwara H, 2005. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* **22**:2157–2165.

Kramerov DA, Vassetzky NS, 2011. Origin and evolution of SINEs in eukaryotic genomes. *Heredity* **107**:487–495.

Luchetti A, Mantovani B, 2011. Molecular characterization, genomic distribution and evolutionary dynamics of Short INterspersed Elements in the termite genome. *Mol Genet Genomics* **285**:175–184.

Luchetti A, Mantovani B, 2013a. Conserved domains and SINE diversity during animal evolution. *Genomics* **102**:296–300.

Luchetti A, Mantovani B, 2013b. Non-LTR R2 element evolutionary patterns: phylogenetic incongruences, rapid radiation and the maintenance of multiple lineages. *PLoS ONE* **8**:e57076.

Luchetti A, Šatović E, Mantovani B, Plohl M, 2016. RUDI, a short interspersed element of the V-SINE superfamily widespread in molluscan genomes. *Mol Genet Genomics* **291**:1419–1429.

Malik HS, Burke WD, Eickbush TH, 1999. The age and evolution of non-LTR retrotransposons. *Mol Biol Evol* **16**:793–805.

Makałowski W, 2000. Genomic scrap yard: how genomes utilize all that junk. *Gene* **259**:61–67.

Matetovici I, Sajgo S, Ianc B, Ochis C, Bulzu P et al., 2016. Mobile element evolution playing jigsaw—SINEs in gastropod and bivalve mollusks. *Genome Biol Evol* **8**:253–270.

Nishihara H, Smit AF, Okada N, 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res* **16**:864–874.

Nishihara H, Plazzi F, Passamonti M, Okada N, 2016. MetaSINEs: broad distribution of a novel SINE superfamily in animals. *Genome Biol Evol* **8**:528–539.

Okada N, Sasaki T, Shimogori T, Nishihara H, 2010. Emergence of mammals by emergency: exaptation. *Genes Cells* **15**:801–812.

Ogiwara I, Miya M, Oshima K, Okada N, 2002. V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Res* **12**:316–324.

Orgel LE, Crick FHC, 1980. Selfish DNA: the ultimate parasite. *Nature* **284**:604–607.

Ohshima K, Okada N, 2005. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet Genome Res* **110**:475–490.

Piskurek O, Okada N, 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc Natl Acad Sci USA* **104**:12046–12051.

Piskurek O, Jackson DJ, 2011. Tracking the ancestry of a deeply conserved eumetazoan SINE domain. *Mol Biol Evol* **28**:2727–2730.

Richard GF, Kerrest A, Dujon B, 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* **72**:686–727.

Schaak S, Gilbert C, Feschotte C, 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* **25**:537–546.

Simon-Loriere E, Holmes EC, 2011. Why do RNA viruses recombine? *Nat Rev Microbiol* **9**:617–626.

Smit AFA, Riggs AD, 1996. Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci USA* **93**:1443–1448.

Takahashi N, Okada N, 2002. Mosaic structure and retropositional dynamics during evolution of subfamilies of short interspersed elements in African cichlids. *Mol Biol Evol* **19**:1303–1312.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**:2725–2729.

Vassetzky NS, Kramerov DA, 2013. SINEBase: a database and tool for SINE analysis. *Nucleic Acids Res* **41**:D83–D89.

Walters-Conte KB, Johnson DL, Johnson WE, O'brien SJ, Pecon-Slattery J, 2014. The dynamic proliferation of CanSINEs mirrors the complex evolution of Feliforms. *BMC Evol Biol* **14**:137.

Werren JH, 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* **108**:10863–10870.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P et al., 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**:973–982.

Yadav VP, Mandal PK, Bhattacharya A, Bhattacharya S, 2012. Recombinant SINEs are formed at high frequency during induced retrotransposition *in vivo*. *Nat Commun* **3**:854.

Yang G, Nagel DH, Feschotte C, Hancock CN, Wessler SR, 2009. Tuned for transposition: molecular determinants underlying the hyperactivity of a *Stowaway* MITE. *Science* **325**:1391–1394.

Zhang L, Chen J-G, Zhao Q, 2015. Regulatory roles of Alu transcript on gene expression. *Exp Cell Res* **338**:113–118.