

Evolution of Two Short Interspersed Elements in *Callorhynchus milii* (Chondrichthyes, Holocephali) and Related Elements in Sharks and the Coelacanth

Andrea Luchetti*, Federico Plazzi, and Barbara Mantovani

Dipartimento di Scienze Biologiche, Geologiche e Ambientali - Università di Bologna, Italy

*Corresponding author: E-mail: andrea.luchetti@unibo.it.

Accepted: May 11, 2017

Abstract

Short interspersed elements (SINEs) are non-autonomous retrotransposons. Although they usually show fast evolutionary rates, in some instances highly conserved domains (HCDs) have been observed in elements with otherwise divergent sequences and from distantly related species. Here, we document the life history of two HCD-SINE families in the elephant shark *Callorhynchus milii*, one specific to the holocephalan lineage (CmiSINEs) and another one (SacSINE1-CM) with homologous elements in sharks and the coelacanth (SacSINE1s, LmeSINE1s). The analyses of their relationships indicated that these elements share the same 3'-tail, which would have allowed both elements to rise to high copy number by exploiting the *C. milii* L2-2_CM long interspersed element (LINE) enzymes. Molecular clock analysis on SINE activity in *C. milii* genome evidenced two replication bursts occurring right after two major events in the holocephalan evolution: the end-Permian mass extinction and the radiation of modern Holocephali. Accordingly, the same analysis on the coelacanth homologous elements, LmeSINE1, identified a replication wave close to the split age of the two extant *Latimeria* species. The genomic distribution of the studied SINEs pointed out contrasting results: some elements were preferentially sorted out from gene regions, but accumulated in flanking regions, while others appear more conserved within genes. Moreover, data from the *C. milii* transcriptome suggest that these SINEs could be involved in miRNA biogenesis and may be targets for miRNA-based regulation.

Key words: CORE domain, Deu domain, miRNA, SINEs, SINE highly conserved domain, SINE replication waves.

Introduction

Transposable elements (TEs) are DNA sequences that multiply either by producing copies that will insert in different genomic locations or by simply repositioning themselves elsewhere in the genome. TEs can be grouped into two major classes: Class I includes all elements replicating via an RNA intermediate (retrotransposons), and Class II comprises elements that move via DNA intermediates (transposons; Wicker et al. 2007). In both classes, autonomous and non-autonomous elements occur. The autonomous elements encode all the enzymes that are required for their retrotransposition, while non-autonomous elements are basically molecular parasites. In fact, for their mobilization, they exploit the enzymatic machinery encoded by autonomous TEs (Ohshima and Okada 2005; Yang et al. 2009).

The best-known representatives of such quasi-parasitic relationships are the non-autonomous short interspersed elements (SINEs) and their partners, the long interspersed

elements (LINEs; Ohshima and Okada 2005). LINEs are ubiquitous Class I elements which code for a polyprotein with a reverse transcriptase (RTase) domain, responsible for the reverse transcription of the RNA intermediates and the integration of cDNA into chromosomal locations. SINEs are small elements with a well-defined modular structure: a small RNA-derived head, an anonymous body (i.e., a sequence without any relationship with known protein-coding domains) and a tail terminating with an oligo-(A) or a microsatellite motif (Kramerov and Vassetzky 2011). In SINEs, the functionally relevant modules are the head and the tail. The former contains the promoters for the RNA polymerase III (*pol-III*) and is thus responsible for the transcription of the SINE-derived RNA that will constitute the template for the reverse transcription. The tail module, on the other hand, is homologous to the tail of the partner LINE. It is this region of the SINE RNA that binds the LINE-encoded protein (Ohshima and Okada 2005). The replication rate and survival of a SINE is

therefore strictly dependent on the partner LINE activity, so that if the latter undergoes silencing, the former will also stop replicating. A way for SINEs to escape silencing is to change the mobilizer LINE; this can be achieved by switching the tail module through recombination with a different LINE or with another SINE carrying a tail matching that of an active LINE.

It is well-known that distinct SINE families have exchanged functional modules (Takahashi and Okada 2002; Deragon and Zhang 2006). This led to the development of the “module exchange” model which predicts that recombination between different SINEs and/or other DNA elements may promote the switch of functional modules of the element, assuring the retrotransposition competency and, eventually, leading to SINEs diversity (Kramerov and Vassetzky 2011; Luchetti and Mantovani 2013).

As mentioned above, the body sequence of a SINE is apparently not functional for retrotransposition. Nonetheless, some SINE families carry highly conserved domains (HCDs) within the body sequence; this has been observed for widely different SINE families, isolated from a variety of host species (Gilbert and Labuda 1999; Ogiwara et al. 2002; Nishihara et al. 2006; Piskurek and Jackson 2011; Luchetti and Mantovani 2013; Matetovici et al. 2016; Nishihara et al. 2016; Luchetti and Mantovani 2016). So far, the discovery of SINE domains conserved and vertically inherited for long periods (up to 800 Ma; Luchetti and Mantovani 2016) remains to be explained. Three models have been proposed. Gilbert and Labuda (1999) suggested that HCDs may help in maintaining the interaction of a SINE with the partner LINE. Alternatively, HCDs may serve as recombination hot spots allowing the generation of new SINEs, in line with the module exchange model (Gilbert and Labuda 1999; Luchetti and Mantovani 2013). A third option points to a possible functional role of HCDs conferring some selective advantage to the host genome (Nishihara et al. 2006; Deragon 2012).

SINE insertions may have a deep impact on the host genome by inducing structural genomic variation and/or modifying gene expression profiles (reviewed in Schmitz 2012). Remarkably, some SINE insertions appear to have triggered major evolutionary changes through exaptation, being included as functional components into the host genome (reviewed in Deragon 2012). In general, TE activity is considered one of the major facilitator of genome evolution, suggesting a link with species diversity and evolution as well (TE-Thrust hypothesis; Oliver and Greene 2011).

In a wide attempt to understand the evolution of HCD-SINEs, we collected several elements (either previously isolated or newly characterized) looking for possible internal homologies with known conserved domains. We here present the results for the elephant shark *Callorhinchus milii* genome (Venkatesh et al. 2014), which shows a SINE family with the CORE HCD. Moreover, we analyzed the relationships with coexisting SINEs and the potential impact of detected insertions on the host genome evolution. The data helped to

depict the life history of two SINE families in the *C. milii* genome and contribute to the knowledge on internal domain conservation.

Materials and Methods

Elephant shark’s SINE2-1_CM and SINE2-1B_CM element consensus sequences, found in the *Callorhinchus milii* genome assembly 6.1.3 (Venkatesh et al. 2014), were downloaded from Repbase (Bao et al. 2015). Presence/absence of SINE copies in the other genomes has been assayed by BLAST (Altschul et al. 1990), selecting hits with an e -value $\leq 1e^{-10}$.

Genome-wide detection of SINE insertions was performed using RepeatMasker v. 4.0 (Smit et al. 2013–2015). We excluded from the analysis all those insertions that could not be unequivocally attributed to a single SINE family, due to the close sequence similarity. These amounted to ~5% of total scored insertions, except the LmeSINE1c family, for which 41% of scored insertions could not be unequivocally attributed.

The age analysis of SINE insertions was calculated using two approaches. First, transposition in transposition (TinT; Churakov et al. 2010) analysis was used to evaluate the relative age of activity within genomes. For this analysis, a RepeatMasker outfile was generated using a collection of TEs downloaded from the RepBase databank (accessed on December 2016), which included elements both specific to *C. milii* and ancestral ones (i.e., shared with other taxa). We, then, added to this library the consensus sequence of the newly found SacSINE1-CM. A second approach was based on sequence divergences. The age of SINE copies was estimated from their divergence from the relative consensus sequence using the Jukes-Cantor substitution model, which accounts for multiple nucleotide substitutions.

The SINE genomic distribution was tested in *C. milii* and in *L. chalumnae*, the only species, among presently examined ones, with well-annotated genomes (obtained from NCBI database, accessed on March 2016; Amemiya et al. 2013; Venkatesh et al. 2014). The overlap between SINE insertions and annotated genes was evaluated with BEDTools v. 2.17 (Quinlan and Hall 2010) using the *intersectBed* function. To assess the significance of observed SINE distribution we performed a simulation. The genomic locations of SINE insertions were randomly permuted 10,000 times using the BEDTools *shuffleBed* function (excluding genome gaps); their genomic distribution was then determined with *intersectBed* and considered as a null distribution against which observed data were compared. Finally, we calculated an empirical P value of obtaining an equal or greater number of overlaps in observed data.

Results

CmiSINE Sequence Characterization

SINE2-1_CM and SINE2B-1_CM heads have a clear tRNA origin, being very similar to *C. milii* Alanine tRNA

(similarity = 81.9–90.3%, respectively); they can also be folded into the typical tRNA clover-leaf secondary structure and retain conserved RNA *pol-III* promoters, the A and B boxes (fig. 1). The two SINEs differ for the 5.1%; due to their close similarity, and for reasons of clarity, they will be collectively referred to as CmiSINEs.

We then checked whether the body aligns to one of the known HCDs (Vassetzky and Kramerov 2013). It emerged that CmiSINEs belong to the CORE-SINE superfamily, having

a body 66.78–68.5% similar to 58 out of 65 bp of the CORE consensus sequence (fig. 1).

As SINEs exploit the partner LINE enzymatic machinery by sharing homology at the 3' end of the tail, we also searched for known *C. millii* LINEs whose tail matches those of CmiSINEs. The analysis of RepBase non-LTRs collection evidenced a 78.6% similarity with the tail of the L2-2_CM LINE (fig. 1; supplementary fig. S1, Supplementary Material online).

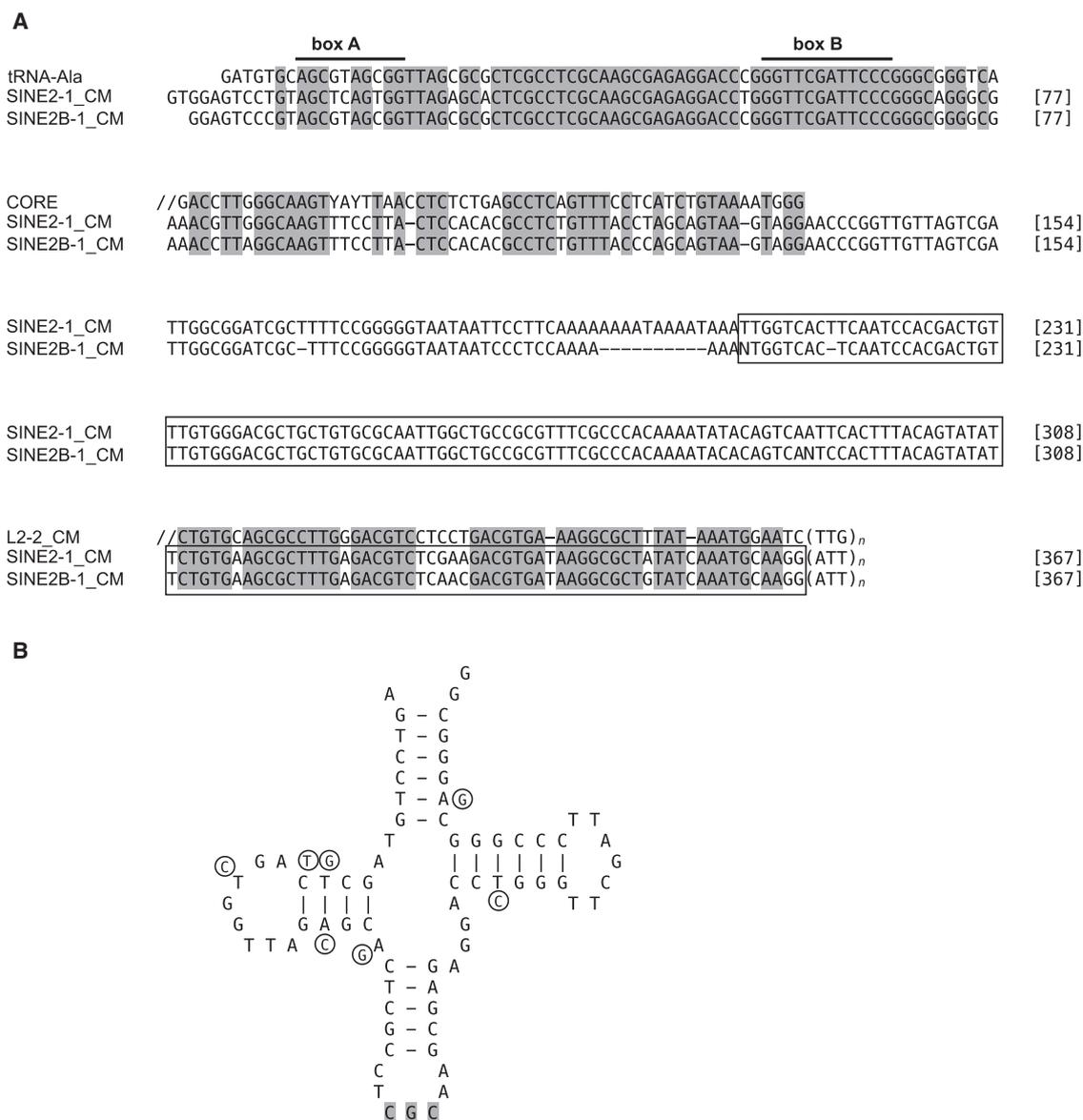


Fig. 1.—CmiSINEs sequence characterization. (A) In the head module, A and B boxes of the RNA *pol-III* promoter are indicated by black lines. Above SINE sequences, the *C. millii* Alanine tRNA, the CORE HCD consensus sequence (taken from Vassetzky and Kramerov 2013) and the L2-2_CM tail are aligned. Identical residues are shaded in grey; the “//” symbol indicates that part of the sequence is omitted as non-homologous. The region shared with the SacSINE1-CM family is boxed. (B) The cloverleaf secondary structure of the SINE2-1_CM tRNA-like head; nucleotide substitutions observed in SINE2B-1_CM are reported circled. The anticodon is shaded in grey.

Search of CmiSINEs in Related Genomes and Relationship with Other Elements

We searched CmiSINEs in other cartilaginous fishes: *Squalus acanthias* (whole genome shotgun clone traces archive, N = 12,873; EST library, N = 32,562), *Rhincodon typus* (assembly Rhincodon v1; Read et al. 2015) and *Leucoraja erinacea* (assembly LER_WGS_1; King et al. 2011). Representatives of bony fishes were also checked: *Danio rerio* (assembly GRCz10; Howe et al. 2013), *Takifugu rubripes* (assembly FUGU5; Aparicio et al. 2002) and *Latimeria chalumnae* (assembly LatCha1; Amemiya et al. 2013). We also searched in the cyclostome *Petromyzon marinus* (assembly 7.0; Smith et al. 2013).

These searches evidenced the absence of CmiSINEs in the assayed genomes; however, the 3' end of CmiSINEs aligned significantly with clones from *Squalus acanthias* (~140 bp, similarity = 77.4–75.5%). Alignment of positive clone fragments obtained from *S. acanthias* indicated that they were copies of another SINE family with a tRNA-related head: when searched in Repbase, it matched with a known element carrying the Deu HCD and isolated in *S. acanthias*: SacSINE1 (Nishihara et al. 2006). The latter element was, then, searched and found in the *C. millii* genome (henceforth, it will be called SacSINE1-CM), with copies showing 74.1% average identity with the SacSINE1 consensus sequence. The comparison between SacSINE1-CM and CmiSINEs confirmed 94.3–97.5% similarity only at their 3' end (~160 bp), including the region homologous to the 3' end of L2-2_CM LINE (fig. 2A; supplementary fig. S1, Supplementary Material online).

We found SacSINE1 also in the *R. typus* genome, where detected copies showed lengths ranging from 97 to 463 bp and a similarity score to the SacSINE1 consensus of 75.4%. The *L. erinacea* genome also harbors SacSINE1-like elements, with length spanning from 185 to 425 bp and an average identity to the SacSINE1 consensus of 73.7%.

Apart from *T. rubripes*, SacSINE1 showed significant matches with cyclostome and the other bony fish genomes (table 1). The similarity with *P. marinus* occurred in a region homologous to the SacSINE1 Deu domain but, apparently, it does not belong to any known repeat, nor it was possible to isolate a new one. Cyclostomes harbor Deu-SINEs (Nishihara et al. 2006), therefore, it is possible that these similarities involved fragments of an extinct SINE lineage. In *D. rerio* positive hits are located within the Deu HCD, that is part of the 5S rDNA-derived element SINE3-1 (Kapitonov and Jurka 2003; Nishihara et al. 2006). Hits on the *L. chalumnae* genome, though, span from the 5' end to the Deu domain, thus including the tRNA-like head and the 5' half of the body. These hits correspond to the coelacanth LmeSINE1 family (Nishihara et al. 2006), comprising five different lineages (LmeSINE1a-e). The alignment of consensus sequences revealed that LmeSINE1s, SacSINE1, and SacSINE1-CM share similarity across the tRNA-like head, the Deu domain and the tail. Only LmeSINE1a and LmeSINE1d lineages showed a large

deletion in the body as a main structural difference (fig. 2A; supplementary fig. S2, Supplementary Material online).

Timing of the SINE Activity

The age analysis performed with the TinT method clearly indicated that CmiSINEs experienced replication waves more recently than SacSINE1-CM, with an almost contemporary activity of SINE2-1_CM and SINE2B-1_CM (fig. 3). The inferred activity time of the LINE L2-2 overlaps to that of analyzed SINEs.

The same analysis was, then, performed to check the relative time of activity of LmeSINE1 elements, evidencing that replication waves of the five lineages occurred with a slightly different timing (fig. 3), LmeSINE1a and LmeSINE1d being mobilized more recently.

We also performed an age analysis by comparing the extent of similarity of each copy to the relative consensus sequence, based on the principle that more recently produced copies show fewer substitutions than those that were produced more distantly in the past. CmiSINEs and SacSINE1-CM consensus sequences were therefore searched in the *C. millii* genome. We found 302,861 CmiSINE insertions (175,998 for SINE2-1_CM and 126,863 for SINE2B-1_CM). Copies divergence ranged from 0.0% to 42.4% (average = 12.6%) and from 0.0% to 47.7% (average = 17.2%), respectively. Moreover, we found 218,294 SacSINE1-CM insertions, with copies diverging from the consensus by 0.0–49.2% (average = 22.4%). The histogram plot clearly suggests a more recent replication wave of CmiSINEs and an older activity of SacSINE1-CM (fig. 4), in full agreement with the TinT analysis.

We, then, attempted to put estimated SINEs activity in an absolute timeframe. Considering the presence of CmiSINEs in the holocephalan lineage and its absence in the elasmobranch one, we used the split age of these two taxa (421 Ma; Inoue et al. 2010) as the maximum age for CmiSINEs. Therefore, the oldest CmiSINE copies, that are the most divergent from the consensus (42.4–47.7%), were considered as being produced 421 Ma and those completely identical to the consensus as being very recently produced elements (0 Ma; fig. 4). Assuming a strict molecular clock, the substitution rate experienced by the oldest CmiSINE copies ranges between 1.0×10^{-3} and 1.1×10^{-3} substitutions/site per million year. Based on these estimates, and assuming that each SINE copy likely accumulated substitutions at the same rate, the maximum activity seems to have occurred ~105–118.2 Ma for SINE2-1_CM, and ~144.7–162.8 Ma for SINE2B-1_CM. SacSINE1-CM, on the other hand, appears to have originated about 434.2–488.5 Ma, with a peak in activity around 201.2–226.4 Ma.

Using the elasmobranch substitution rate calculated on nuclear gene synonymous substitutions, 1.6×10^{-4} substitutions/site per million year (Martin 1999), we obtained widely different ranges of activity: CmiSINEs reached their

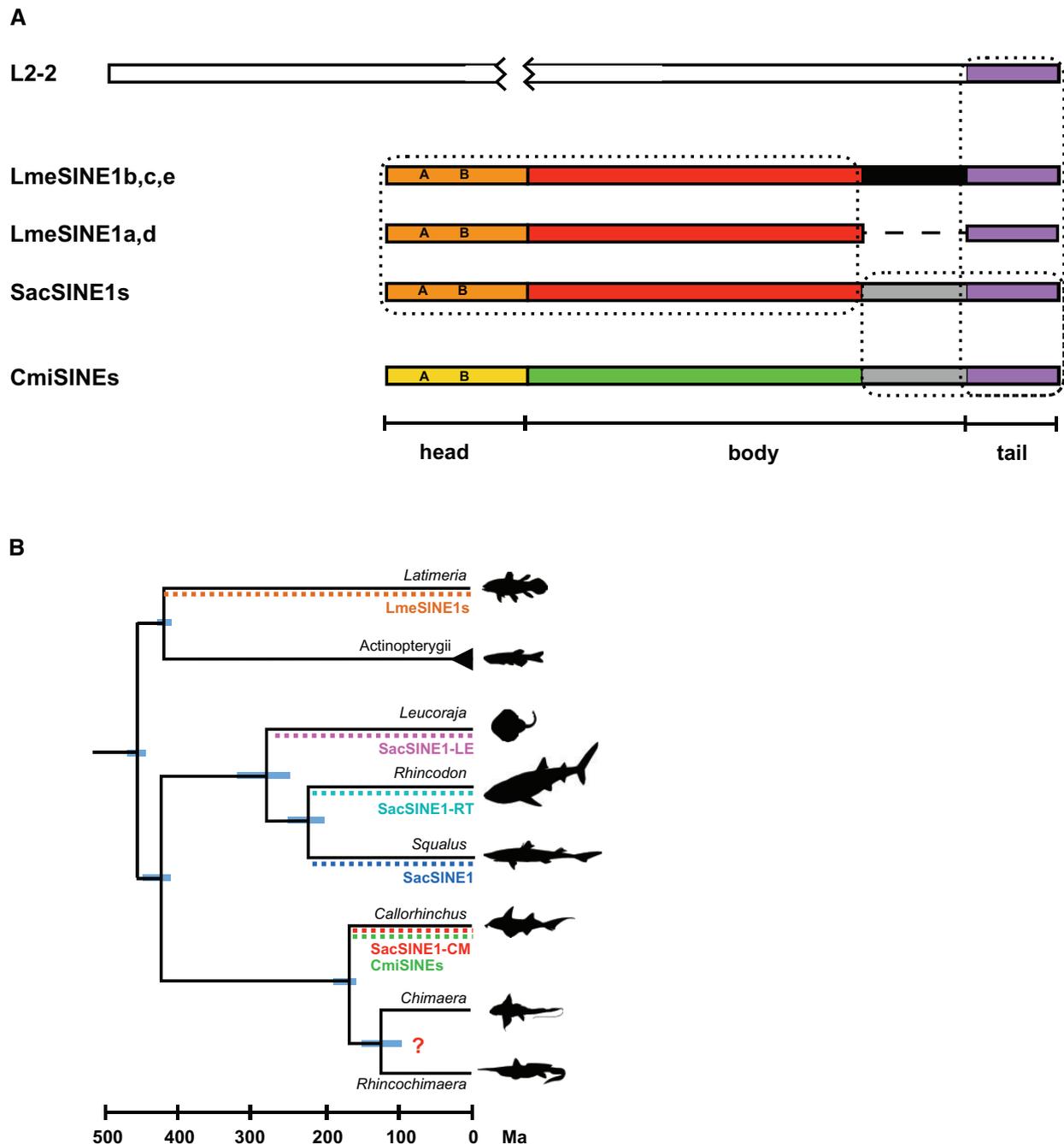


Fig. 2.—Schematic drawing of analyzed SINE structures and taxonomic distribution. (A) Homologous regions are boxed with dotted lines; A and B in the SINE head indicate the A and B boxes of the RNA *pol-III* promoter. CmiSINEs: SINE2-1_CM and SINE2B-1_CM. SacSINEs: SacSINE1, SacSINE1-CM, SacSINE1-LE (*L. erinacea*), and SacSINE1-RT (*R. typus*). The dashed line in the LmeSINE1a,d sequence indicates the deletion with respect to other LmeSINE1 and SacSINE1 elements. Above, the L2-2 LINE has been represented to show the tail homology; the interruption marks indicate a sequence portion omitted for graphical reasons. SINE sizes are representative and not in scale. (B) Taxonomic distribution of analyzed SINEs superimposed to the fish phylogeny obtained from Inoue et al. (2010) and Betancur-R. et al. (2013). Bars at nodes represent error associated to age estimates; timescale is reported at bottom (Ma, million years ago). The question mark indicates the lack of data for other holocephalan species.

peak of activity around 787 and 1,075 Ma, while SacSINE1-CM would have reached it about 1,400 Ma.

We then applied the newly calculated substitution rate to coelacanth LmeSINE1s in order to date their activity. The

search for insertions of the five LmeSINE1 lineages in the *L. chalumnae* genome resulted in 73,075 (LmeSINE1a), 32,791 (LmeSINE1b), 500 (LmeSINE1c), 70,400 (LmeSINE1d), and 35,616 (LmeSINE1e) hits. Overall,

divergence values ranged from 0.0% to 54.2%. Again, the divergence plot fully agrees with TinT relative time analysis (fig. 3). The age analysis indicated that the replication of the oldest lineage (LmeSINE1e) started at 478.4–538.2 Ma and two main replication waves, corresponding to LmeSINE1d and LmeSINE1a activity, occurred 122.8–138.1 Ma and 57.1–64.2 Ma (fig. 4).

SINEs Genomic Distribution

We checked the distribution of retrieved CmiSINEs and SacSINE1-CM copies within the *C. milii* genome and of LmeSINE1 insertions within the *L. chalumnae* genome, to verify whether insertions are differentially distributed with respect to coding regions. The analysis showed that the two CmiSINEs have about 76.7% (SINE2-1_CM) and 78.1% (SINE2B-1_CM) of insertions distributed within genes or in flanking regions

Table 1

First Ten Best Hits of SacSINE1 BLAST Search in Cyclostome and Bony Fish Genomes

Species	Identity (%)	Homologous Region Span ^a
<i>Petromyzon marinus</i> ^b	70.9–73.5	134–257
<i>Danio rerio</i>	78.7–81.6	124–258
<i>Latimeria chalumnae</i>	76.6–77.5	7–255

^aRefers to nucleotide position in the SacSINE1 consensus sequence.

^bThese results have been obtained when the *e-value* threshold has been lowered to $\leq 1e^{-5}$; only three hits have been scored.

(± 10 kbp), while the third element, SacSINE1-CM, showed only 69.3% insertions in these regions (table 2). Comparisons with simulated data indicated that while SacSINE1-CM observed insertions are significantly less common within both genes and flanking regions, CmiSINEs insertions are significantly less common within genes but significantly more represented in flanking regions ($P < 0.05$) (table 2).

Concerning *L. chalumnae* element LmeSINE1s, the proportion of insertions within genes and flanking regions ranged from 55.0% (LmeSINE1a) to 59.6% (LmeSINE1c). Comparisons with simulated insertions profiles indicated quite different patterns (table 2). LmeSINE1b and LmeSINE1c observed insertions are generally not significantly different from simulated ones, except for LmeSINE1b in the upstream 10 kbp (table 2). On the other hand, LmeSINE1d and LmeSINE1e are both significantly more represented than expected in genes and flanking regions, except for LmeSINE1e in downstream flanking regions (table 2). Finally, LmeSINE1a turned out to be significantly less represented in both genes and flanking regions (table 2).

SINE Distribution in *C. milii* Transcriptome

BLAST comparison with *C. milii* microRNAs (miRNAs) collection revealed significant similarities with both CmiSINEs and SacSINE1-CM. CmiSINEs showed significant similarity with the precursor of eshark_NOVEL_91 (acc. no. JX994806; identity = 82.6–85.9%) and SacSINE1-CM resulted similar to

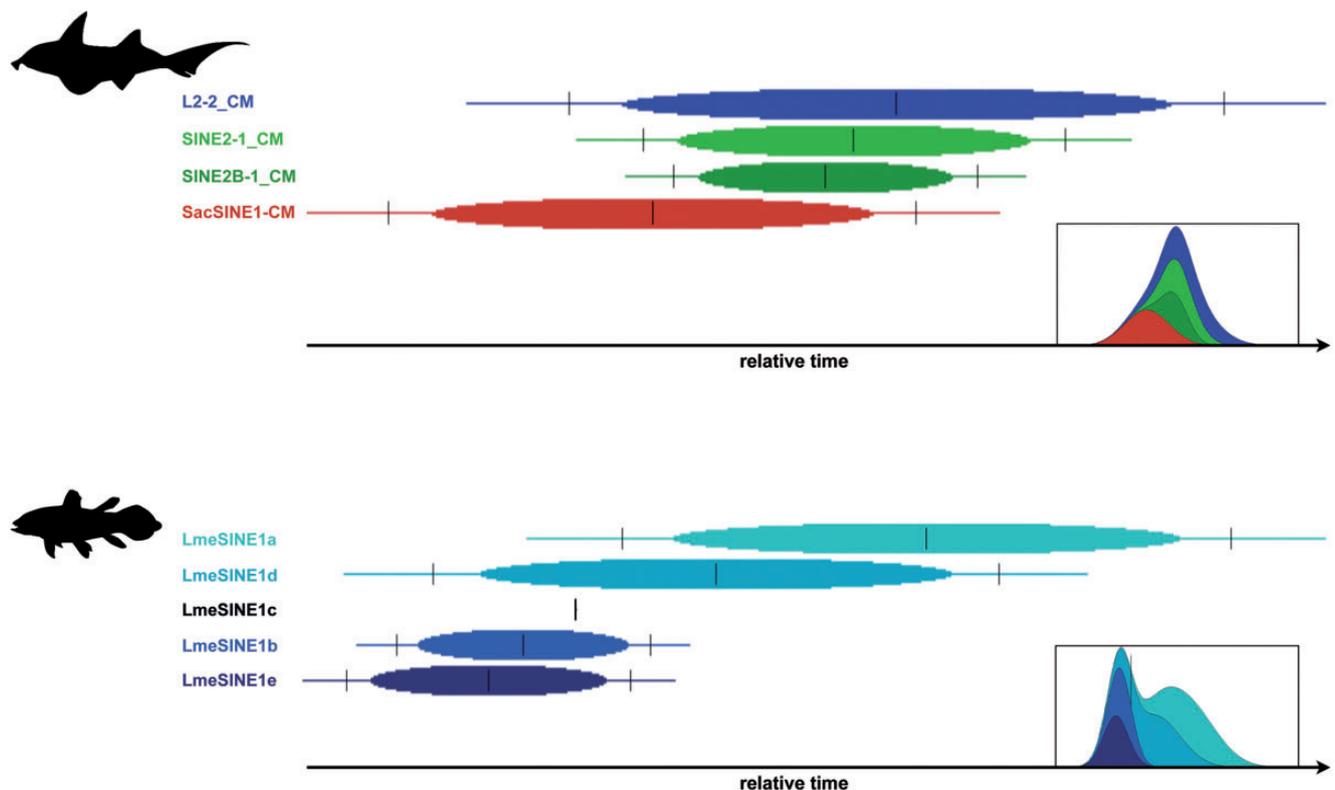


Fig. 3.—TinT analysis performed on *C. milii* and *L. chalumnae* genomes. In the lower right inset, the cumulative distribution of insertions is also reported.

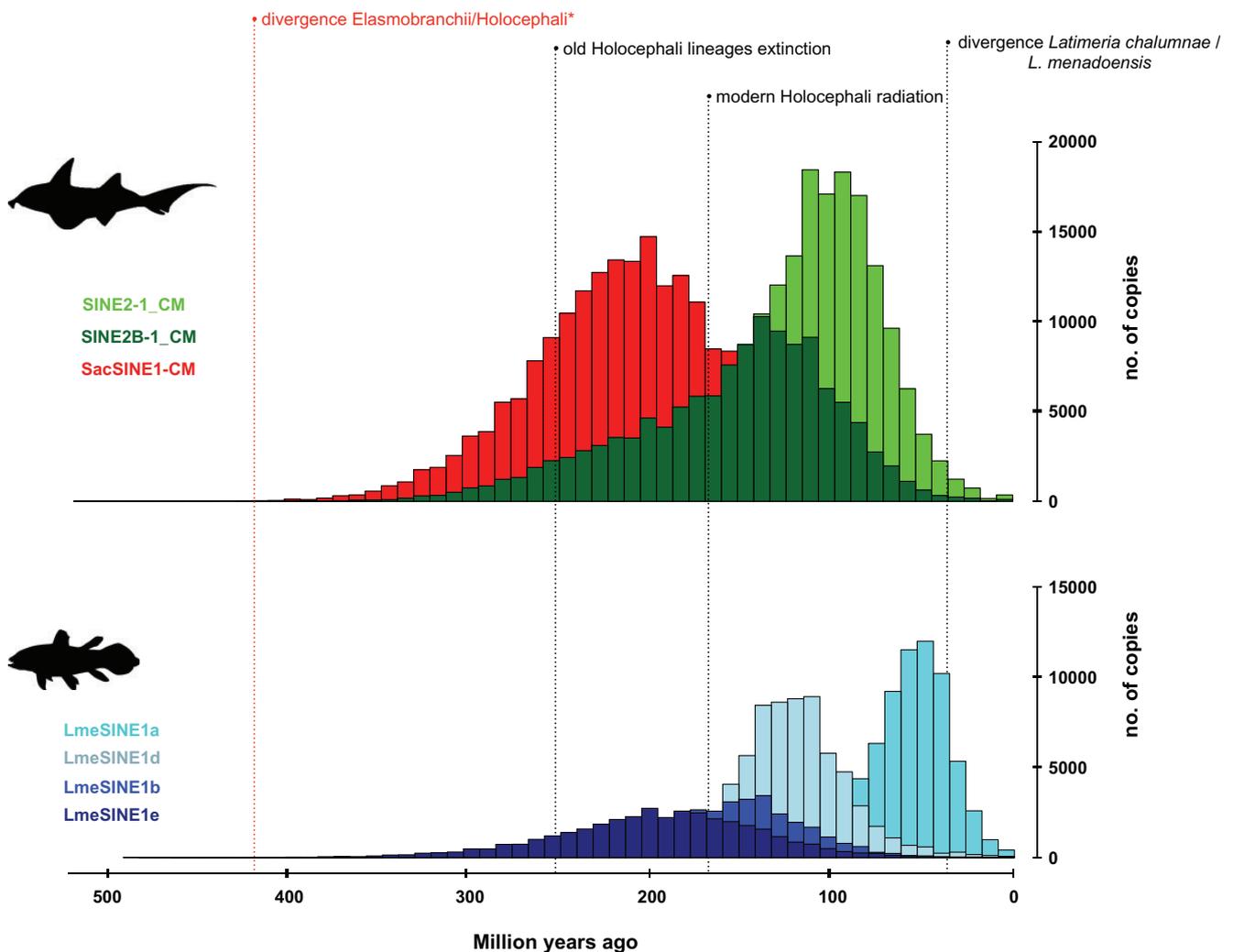


Fig. 4.—Divergence analysis of SINEs in *C. milii* and *L. chalumnae* genomes. Each bin represents the 0.5% of Jukes-Cantor divergence from the consensus sequence. Age of species divergences and radiations are taken from Inoue et al. (2005, 2010) and Betancur-R. et al. (2013); the asterisk marks the cladogenetic event used to calibrate the CmiSINEs' evolutionary rate, that has been applied also to the divergence of other SINEs.

precursors of eshark_NOVEL_211 and eshark_NOVEL_234 (acc. nos. JX994610, JX994511; identity = 76.5% and 78.9%, respectively) (fig. 5). The extent of similarity includes the SINE HCD (CORE in the CmiSINE and Deu in SacSINE1-CM) and, in two instances, part of the tRNA-related head. The predicted mature miRNA sequence resulted homologous to a fragment either of the tRNA-related head or of the HCD itself (fig. 5).

We then checked the presence of SINEs within *C. milii* messenger RNAs (mRNAs). BLAST search retrieved 104 positive hits for CmiSINE and 34 for SacSINE1-CM: all hits are located within the 3' UTR (supplementary tables S1 and S2, Supplementary Material online).

Discussion

Short interspersed elements are almost ubiquitous, fast evolving components of the eukaryotic genome that may have a

profound impact on the host genome (reviewed in Schmitz 2012). In the present analysis, we describe the life-cycle and genome biology of two SINEs (SINE2-1_CM and SINE2B-1_CM, here collectively referred to as CmiSINEs) within the genome of the elephant shark *Callorhynchus milii*, and their relationships with pre-existing SINEs.

Evolutionary History and Genome Invasion of CmiSINEs

We found CmiSINEs only in the *C. milii* genome, coexisting with the newly discovered element SacSINE1-CM. This SINE has homologs distributed in the genome of other cartilaginous fishes (SacSINE1 family in *S. acanthias*, *R. typus*, and *L. erinacea*) and, although with a quite divergent sequence, in the coelacanth (LmeSINE1 family).

We discovered that CmiSINEs belong to the CORE-SINE superfamily, as they show similarity with the CORE domain,

Table 2Genomic Distribution of Observed and Simulated SINE Insertions (\pm Standard Deviation)

SINE	Location	Observed Insertions	Simulated Insertions (\pm S.D.)	Results of Significant Comparisons ^a
SINE2-1_CM	Upstream genes	29,556	27,591 \pm 162	>
	Within genes	76,000	81,194 \pm 213	<
	Downstream genes	29,487	27,570 \pm 162	>
SINE2B-1_CM	Upstream genes	22,418	19,825 \pm 136	>
	Within genes	54,015	58,364 \pm 176	<
	Downstream genes	22,641	19,807 \pm 135	>
SacSINE1-CM	Upstream genes	30,323	34,271 \pm 177	<
	Within genes	90,622	100,836 \pm 233	<
	Downstream genes	30,258	34,242 \pm 179	<
LmeSINE1a	Upstream genes	5,156	5,429 \pm 71	<
	Within genes	29,903	30,778 \pm 143	<
	Downstream genes	5,146	5,415 \pm 71	<
LmeSINE1b	Upstream genes	2,709	2,524 \pm 49	>
	Within genes	13,918	13,796 \pm 89	>
	Downstream genes	2,589	2,516 \pm 49	>
LmeSINE1c	Upstream genes	39	38 \pm 6	>
	Within genes	227	209 \pm 11	>
	Downstream genes	32	37 \pm 6	>
LmeSINE1d	Upstream genes	5,585	5,429 \pm 71	>
	Within genes	30,584	29,641 \pm 132	>
	Downstream genes	5,575	5,415 \pm 71	>
LmeSINE1e	Upstream genes	2,917	2,726 \pm 50	>
	Within genes	15,468	14,962 \pm 93	>
	Downstream genes	2,713	2,720 \pm 50	>

^aSymbols "<" and ">" indicate comparisons where observed insertions are significantly less or more ($P < 0.05$), respectively, than simulated ones.

a HCD sequence retrieved among several SINE families isolated in widely different host species (Gilbert and Labuda 1999, 2000; Vassetzky and Kramerov 2013; Nishihara et al. 2016). The CORE domain is assumed to be conserved since the Radiata-Bilateria split, being therefore more than 800 million years old (Vassetzky and Kramerov 2013). The CmiSINEs 3' end, though, is homologous to that of SacSINE1 and LmeSINE1 families, and to the 3' end of a L2 LINE. As the SacSINE1 family harbors the Deu domain, the two SINEs appear to have been originated independently. Yet, their activities appear linked because they share the exploitation of the same LINE.

The *de novo* origin of a SINE is a chimera between a small RNA gene, such as tRNA, 7SL RNA or 5S RNA, and other sequences (Kramerov and Vassetzky 2011), most likely occurring upon reverse transcriptase template switch. This process is also known to occur in retrovirus recombination (Negroni and Buc 2001) and produced chimeric transcripts in *Homo sapiens* (Buzdin et al. 2002). This allows RNAs from different origins to merge into a single molecule that can be, then, reintegrated. For example, in the platypus genome, small nucleolar RNAs (snoRNA) formed a chimera with a LINE-related RTE retrotransposon tail; the resulting element was, then, replicated up to 40,000 copies through a SINE-like duplication (Schmitz et al. 2008). Furthermore, it is well known that the

highly dynamic nature of SINEs is also due to the frequent recombination with other elements, exchanging sequence modules (the head, the body and the tail; Takahashi and Okada 2002; Deragon and Zhang 2006; Kramerov and Vassetzky 2011; Luchetti and Mantovani 2013). As the CORE domain is considered an ancestral component of some SINE families, we can speculate that an undetected/extinct CORE-SINE switched the head with a new tRNA after the holocephalan lineage divergence. Similarly, a proto-CmiSINE acquired a new 3' end, containing the region homologous to the tail of a functional LINE, through recombination with a SacSINE1-CM element. The possibility of recombination or template switching with the L2 LINE element can be ruled out as the region homologous to SacSINE1-CM is larger than that homologous to the L2 tail (fig. 1; supplementary fig. S1, Supplementary Material online). The 3' end switch would have conferred to CmiSINEs the retrotransposition competency and, therefore, the ability to successfully invade the host genome.

As evidenced by the age analysis, the copy number of CmiSINEs increased when SacSINE1-CM started to decrease, reaching the peak of activity more recently (figs. 3 and 4). It is possible that a competition between SacSINE1-CM and CmiSINEs was established. What may have determined the success of CmiSINEs over the SacSINE1-CM? SacSINE1-CM

appears quite unlikely, we decided to disregard results from this substitution rate. The substitution rate based on presently analyzed data, further, gave age estimates more consistent with the evolutionary history of the considered species.

Since the origin of the holocephalan lineage, about 421 Ma, we detected three replication waves: the SacSINE1-CM one, ~200–220 Ma, and the CmiSINE ones, occurring ~140–160 Ma and ~100–120 Ma. Although having a potentially large error associated with the assumptions we made for the age calculation (i.e., strict molecular clock and constancy of SINE copies substitution rate), this dating suggests that replication waves occurred right after two major events in the evolution of Holocephali: first, the end-Permian extinction about 250 Ma, when, as suggested by fossil records, it seems that a single lineage survived by habitat shifting and gave rise to modern Holocephali (Grogan and Lund 2004); second, the diversification of modern Holocephali, which occurred about 167 Ma (Inoue et al. 2010).

The LmeSINE1 family in the coelacanth genome showed a pattern with two main replication waves, ~120 and ~50 Ma, corresponding to the activity of two elements: LmeSINE1d and LmeSINE1a, respectively. This double-peak activity profile is in line with the analysis carried out on the full TE complement (Chalopin et al. 2014). Interestingly, the most recent replication wave occurred close to the divergence time of the two coelacanth species, *L. chalumnae* and *L. menadoensis* (30–40 Ma; Inoue et al. 2005).

On the whole, it appears that some SINE mobilization bursts in *C. milii* and *L. chalumnae* occurred relatively close to major events in the host taxon evolution. It is thus intriguing to ask whether this correlation provides some causal relationships. TE bursts are explosive replication waves of elements that may occur under certain conditions. For example, environmental stresses acting on host species (Capy et al. 2000; Grandbastien et al. 2005; Zeh et al. 2009) or genomic “shocks”, such as host lineage hybridization, may induce TE bursts (O’Neill et al. 1998; Labrador et al. 1999; Wang et al. 2010). Increases of TE copy number may trigger genomic restructuring and it is considered a generator of diversity, possibly underlying speciation events (the TE-Thrust hypothesis; Rebollo et al. 2010; Oliver and Greene 2011; Belyayev 2014). Furthermore, some TE insertions may also have adaptive significance (Gonzalez et al. 2010; Kofler et al. 2015), supporting their role in species diversity and evolution. Examples of correlations between TE expansion and speciation events have been given in mammals (Pace and Feschotte 2007; Ray et al. 2008; Jurka et al. 2011) as well as in bony fishes (De Boer et al. 2007; Symonova et al. 2013). Thus, it could be hypothesized a possible role of SINE insertions in species diversification and adaptation.

Most of the SINE insertions in both *C. milii* and *L. chalumnae* occurred within genes and their flanking regions. In the elephant shark, though, CmiSINEs have accumulated in gene flanking regions, but they appear as preferentially sorted out

from genes. On the other hand, SacSINE1-CM resulted significantly less represented in the whole genic regions (i.e., genes and flanking regions). Conversely, the picture in the coelacanth genome is more complex: the most recent active element, LmeSINE1a, appears as preferentially sorted out from the whole genic regions, at variance of those elements that burst in the past. In *L. chalumnae* both past and recent TE activity has been taken as an indication of a dynamic genome, in contrast to the morphological stasis (Chalopin et al. 2014); moreover, it has been suggested that TE activity affected postspeciation divergence of the two *Latimeria* species (Naville et al. 2014). Although direct evidence cannot be drawn from this analysis, our data on LmeSINE1 elements genomic distribution appear partially in line with the role proposed for TEs in genome evolution. On the other hand, the replication wave of CmiSINEs in the holocephalan genome occurred right after the main lineage diversification, indicating that these elements might be not directly implicated in the modern Holocephali radiation. Though, the preferential insertions accumulation in gene flanking regions could hint at possible roles in the *C. milii* lineage evolution.

TE replication waves may be not causative but rather caused by host’s evolutionary history, and may have affected genome evolution subsequently. For example, TE massive burst occurred during the domestication of rice, thus TE insertions were not causative of cultivar evolution even if some insertions became part of the new regulatory network (Naito et al. 2009). In the “Carrier Subpopulation” hypothesis, Jurka et al. (2011) linked TE burst with species fragmentation in subpopulations. As subpopulations can lead to speciation, hence TE replication waves correlate with species diversification, being not causative but rather accompanying the process.

Obviously, this does not exclude that new insertions may trigger species-specific genomic changes (Jurka et al., 2011). From data presented here, we can infer a possible role of *C. milii* SINEs in the genome functionality and, thus, in its evolution. In fact, both CmiSINEs and SacSINE1-CM sequences match miRNA sequences. These are small RNAs that act as post-transcriptional regulators of gene expression by binding target mRNAs at their 3’ UTR (He and Hannon, 2009). Accordingly, we found mRNAs transcribed from 138 genes carrying CmiSINEs or SacSINE1-CM insertions within the 3’ UTR. Therefore, our data point out that SINEs can be transcribed as regulatory RNAs and that they may regulate the expression of these 138 genes. Post-transcriptional gene expression regulation by SINE-derived miRNA have been found for primate *Alus* and fish V-SINEs: in both instances SINEs located at mRNAs 3’ UTR act as binding sites for regulatory miRNAs (Daskalova et al. 2006; Scarpato et al. 2015). Moreover, a possible role for the regulation of SINEs activity cannot be excluded, with SINE-derived miRNA binding retrotransposing SINE transcripts. Overall, data presented here reinforce the hypothesis of a functional role for these elements.

SINEs Genomic Distribution and Highly Conserved Domains

HCD-SINEs represent an apparent paradox as they show a long-term conserved element, the HCD in the body module, in an otherwise fast evolving sequence, the SINE itself. Many attempts have been made to explain the conservation of such regions and two main hypotheses, not necessarily mutually exclusive, appear as more reliable: i) HCD are selectively conserved because they may serve as SINE-SINE recombination hotspot, or ii) HCD are selectively conserved because they provide some advantage to the host genome, having some still undetermined, albeit important, function (reviewed by Deragon 2012). Some highly conserved SINE domains have been implicated in exaptation events, being domesticated as enhancers (Bejerano et al. 2006; Santangelo et al. 2007; Sasaki et al. 2008; Nakanishi et al. 2012). If HCDs have some function exploitable by the host genome for gene functionality, one might expect to find them more frequently associated to coding regions. In this study, we provide evidence that two different HCD-SINEs, CORE-(CmiSINEs) and Deu-SINEs (SacSINE1-CM), may have different genomic distribution in the same genome. In fact, while both HCD-SINEs are preferentially sorted out from gene regions, CmiSINEs appear to accumulate in the flanking regions.

Data on SacSINE1-CM are in line with data on another Deu-SINE, the amniote-wide AmnSINE (Hirakawa et al. 2009). On the other hand, elements of the LmeSINE1 lineage—a Deu-SINE family homologous to SacSINE1-CM—may show preferential occupancy within genes and their flanking regions, a random distribution or even a significant depletion from gene regions. Thus, it is likely that HCD-SINEs may affect genomes in different ways.

Present data do not allow us to suggest a specific function for the Deu domain of LmeSINE1, but a possible model for both CmiSINEs and SacSINE1-CM can be envisaged. In these instances, HCDs of the two SINEs match almost entirely to miRNA precursors, a form that is not simply an intermediate but may have regulatory activity (Roy-Chaudhuri et al. 2014). Therefore, the possibility to be transcribed as miRNA and to act also as a target for mRNA post-transcriptional regulation would imply an active function for their HCD. This could be the first evidence of a specific function for these paradoxical highly conserved domains. Other HCD-SINEs have been found to act as miRNA target (V-SINEs); on the other hand, it was not specified whether they are also transcribed in miRNA precursors nor which part of the SINE is included in the transcript (Scarpato et al. 2015). It would be, therefore, interesting to check if other HCD-SINEs are actively transcribed as miRNA and if the HCD itself is involved in the transcription.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We wish to thank Andrea Pozzi for reading and commenting the manuscript. This work was supported by RFO funding from the University of Bologna (no. [RFO2015_2016](#)) to A.L. and B.M. We would like also to thank two anonymous reviewers, whose comments and suggestions helped to substantially improve the manuscript.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Amemiya CT, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* 496:311–316.
- Aparicio S, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–1310.
- Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11.
- Bejerano G, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441:87–90.
- Belyayev A. 2014. Bursts of transposable elements as an evolutionary driving force. *J Evol Biol.* 27:2573–2584.
- Betancur-R R, et al. 2013. The tree of life and a new classification of bony fishes. *PLoS Curr.* 18:5.
- Blair JE. 2009. Animals (Metazoa). In: Hedges SB, Kumar S editors. *The timetree of life*. Oxford: Oxford University Press, p. 223–230.
- Buzdin A, et al. 2002. A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of L1. *Genomics* 80:402–406.
- Capy P, Gasperi G, Biémont C, Bazin C. 2000. Stress and transposable elements: co-evolution or useful parasites?. *Heredity* 85:101–106.
- Chalopin D, et al. 2014. Evolutionary active transposable elements in the genome of the coelacanth. *J Exp Zool (Mol Dev Evol)* 322B:322–333.
- Churakov G, et al. 2010. A novel web-based TinT application and the chronology of the primate Alu retroposon activity. *BMC Evol Biol* 10:376.
- Daskalova E, Baev V, Rusinov V, Minkov I. 2006. 3'UTR-located *Alu* elements: donors of potential miRNA target sites and mediators of network miRNA-based regulatory interactions. *Evol Bioinform Online* 2:103–120.
- De Boer JG, Yazawa R, William S, Davidson WS, Ben Koop BF. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8:422.
- 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 Heidelberg: Springer-Verlag. p. 253–271.
- Deragon J-M, Zhang X. 2006. Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. *Syst Biol.* 55:949–956.
- Gilbert N, Labuda D. 1999. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci U S A.* 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.
- Grandbastien M-A, et al. 2005. Stress activation and genomic impact of *Tnt1* retrotransposons in Solanaceae. *Cytogenet Genome Res.* 110:229–241.

- Grogan D, Lund R. 2004. The origin and relationships of early Chondrichthyes. In: Carrier JC, Musick JA, Heithanus MR, editors. *Biology of sharks and their relatives*. London: CRC Press. p. 3–31.
- He L, Hannon GJ. 2009. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*. 5:522–531.
- Howe K, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496:498–503.
- Inoue JG, Miya M, Venkatesh B, Nishida M. 2005. The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanth. *Gene* 349:227–235.
- Inoue JG, et al. 2010. Evolutionary origin and phylogeny of the modern holoccephalan (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. *Mol Biol Evol*. 27:2576–2586.
- Jurka J, Bao W, Kojima KK. 2011. Families of transposable elements, population structure and the origin of species. *Biol Direct*. 6:44.
- Kapitonov VV, Jurka J. 2003. A novel class of SINE elements derived from 5S rRNA. *Mol Biol Evol*. 20:694–702.
- King BL, Gillis JA, Carlisle HR, Dahn RD. 2011. A natural deletion of the HoxC cluster in elasmobranch fishes. *Science* 334:1517.
- Kofler R, Nolte V, Schlötterer C. 2015. Tempo and mode of transposable element activity in *Drosophila*. *PLoS Genet*. 11:e1005406.
- Kramerov DA, Vassetzky NS. 2011. Origin and evolution of SINEs in eukaryotic genomes. *Heredity* 107:487–495.
- Labrador M, Farre M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of *Osvaldo*. *Mol Biol Evol*. 16:931–937.
- Luchetti A, Mantovani B. 2013. Conserved domains and SINE diversity during animal evolution. *Genomics* 102:296–300.
- Luchetti A, Mantovani B. 2016. Rare horizontal transmission does not hide long-term inheritance of SINE highly conserved domains in the meta-zoan evolution. *Curr Zool*. 62:667–674.
- Martin AP. 1999. Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). *Mol Biol Evol*. 16:996–1002.
- Matetovici I, et al. 2016. Mobile element evolution playing jigsaw: SINEs in gastropod and bivalve mollusks. *Genome Biol Evol*. 8:253–270.
- Nakanishi A, et al. 2012. A SINE-derived element constitutes a unique modular enhancer for mammalian diencephalic *Fgf8*. *PLoS One* 7:e43785.
- Naito K, et al. 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461:1130–1134.
- Naville M, Chalopin D, Volff J-N. 2014. Interspecies insertion polymorphism analysis reveals recent activity of transposable elements in extant coelacanth. *PLoS One* 9:e114382.
- Negrioni M, Buc H. 2001. Mechanisms of retroviral recombination. *Annu Rev Genet*. 35:275–302.
- 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.
- 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.
- Oshima K, Okada N. 2005. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet Genome Res*. 110:475–490.
- Oliver KR, Greene WK. 2011. Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. *Mob DNA* 2:8.
- O'Neill RJ, O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393:68–72.
- Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res*. 17:422–432.
- Piskurek O, Jackson DJ. 2011. Tracking the ancestry of a deeply conserved eumetazoan SINE domain. *Mol Biol Evol*. 28:2727–2730.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- Ray DA, et al. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res*. 18:717–728.
- Read TD, et al. 2015. Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828. *PeerJ PrePrints* 3: e837v1 [last accessed on July 2016]. Available from: <https://doi.org/10.7287/peerj.preprints.837v1>.
- Rebollo R, Horard B, Hubert B, Vieira C. 2010. Jumping genes and epigenetics: towards new species. *Gene* 454:1–7.
- Roberston HM. 2002. Evolution of DNA transposons in eukaryotes. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. *Mobile DNA II*. Washington, D.C.: ASM Press. p. 1093–1110.
- Roy-Chaudhuri B, et al. 2014. Regulation of miRNA-mediated gene silencing by miRNA precursors. *Nat Struct Mol Biol*. 21:825–832.
- Santangelo AM, et al. 2007. Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. *PLoS Genet*. 3:1813–1826.
- Sasaki T, et al. 2008. Possible involvement of SINEs in mammalian brain formation. *Proc Natl Acad Sci U S A*. 105:4220–4225.
- Scarpato M, et al. 2015. Short interspersed DNA elements and miRNAs: a novel hidden gene regulation layer in zebrafish?. *Chromosome Res*. 23:533–544.
- Schmitz J, et al. 2008. Retroposed SNOfall: a mammalian-wide comparison of platypus snoRNAs. *Genome Res*. 18:1005–1010.
- Schmitz J. 2012. SINEs as driving forces in genome evolution. *Genome Dyn*. 7:92–107.
- Smalheiser NR, Torvik VI. 2006. Alu elements within human mRNAs are probable microRNA targets. *Trends Genet*. 22:532–536.
- Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>.
- Smith JJ, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet*. 45:415–421.
- Symonova R, et al. 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol Biol*. 13:42.
- 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.
- Varshney D, Vavrova-Anderson J, Oler AJ, Cairns BR, White RJ. 2015. Selective repression of SINE transcription by RNA polymerase III. *Mob Genet Elements*. 5:86–91.
- Vassetzky NS, Kramerov DA. 2013. SINEBase: a database and tool for SINE analysis. *Nucleic Acids Res*. 41:D83–D89.
- Venkatesh B, et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505:174–179.
- Wang N, et al. 2010. Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with *Zizania latifolia*. *BMC Plant Biol*. 10:190.
- Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 8:973–982.
- 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.
- Zeh DW, Zeh JA, Ishida Y. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays* 31:715–726.

Associate editor: Michelle Meyer