MetaSINEs: Broad Distribution of a Novel SINE Superfamily in Animals

Hidenori Nishihara[1,†], Federico Plazzi[2,†], Marco Passamonti[2,*], and Norihiro Okada[3,4,*]

[1]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Midori-Ku, Yokohama, Kanagawa, Japan

[2]Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

[3]Department of Life Sciences, National Cheng Kung University, Tainan, Taiwan

[4]Foundation for Advancement of International Science, Tsukuba, Japan

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: nokada@fais.or.jp;marco.passamonti@unibo.it.

Accepted: February 8, 2016

Data deposition: The 298 sequences determined in this work has been deposited in DDBJ/EMBL/GenBank under the accession numbers LC122973-LC123270.

Abstract

SINEs (short interspersed elements) are transposable elements that typically originate independently in each taxonomic clade (order/family). However, some SINE families share a highly similar central sequence and are thus categorized as a SINE superfamily. Although only four SINE superfamilies (CORE-SINEs, V-SINEs, DeuSINEs, and Ceph-SINEs) have been reported so far, it is expected that new SINE superfamilies would be discovered by deep exploration of new SINEs in metazoan genomes. Here we describe 15 SINEs, among which 13 are novel, that have a similar 66-bp central region and therefore constitute a new SINE superfamily, MetaSINEs. MetaSINEs are distributed from fish to cnidarians, suggesting their common evolutionary origin at least 640 Ma. Because the 3′ tails of MetaSINEs are variable, these SINEs most likely survived by changing their partner long interspersed elements for retrotransposition during evolution. Furthermore, we examined the presence of members of other SINE superfamilies in bivalve genomes and characterized eight new SINEs belonging to the CORE-SINEs, V-SINEs, and DeuSINEs, in addition to the MetaSINEs. The broad distribution of bivalve SINEs suggests that at least three SINEs originated in the common ancestor of Bivalvia. Our comparative analysis of the central domains of the SINEs revealed that, in each superfamily, only a restricted region is shared among all of its members. Because the functions of the central domains of the SINE superfamilies remain unknown, such structural information of SINE superfamilies will be useful for future experimental and comparative analyses to reveal why they have been retained in metazoan genomes during evolution.

Key words: transposable elements, short interspersed elements, central domain, mollusks, bivalves, Eumetazoa.

Introduction

Recent accelerated genome sequencing of animals has revealed the diversity of their genomic sequence and composition and has provided us with a great deal of insights with respect to their evolution. These efforts have shed light on the reality that the genomes of eukaryotes, especially of metazoans, harbor various transposable elements (TEs) such as retroposons, which include SINEs (short interspersed elements), LINEs (long interspersed elements), and LTR retrotransposons, and DNA-type transposons (Smit 1999; Deininger et al. 2003). Retroposon is a general term for TEs that amplify their copies via an RNA intermediate using reverse transcriptase (RT) (Rogers 1985). Their wide distribution suggests that TEs have contributed to the diversification of eukaryotic genomes during their long evolutionary history (Kazazian 2004; Feschotte 2008; Cordaux and Batzer 2009). However, we are still far from fully understanding the variety of TEs, because many genome sequencing projects have focused on a restricted group of organisms and because deep investigations of genomes with respect to TEs have not kept pace with the exponential increase in available genomic data.

SINEs are one of the retroposons that borrow RT from a partner LINE for retrotransposition. Typical SINEs are derived from transfer RNA (tRNA) in the 5′ promoter region and transcribed by RNA polymerase III (Pol III) (Sakamoto and Okada 1985; Okada 1991a, 1991b). The 3′-terminal tail sequence of the SINE RNA is

recognized by RT encoded by a partner LINE and the cDNA is integrated into the genome (Kajikawa and Okada 2002; Hayashi et al. 2014). Many SINEs share 3′ tail sequences with their partner LINEs (Ohshima et al. 1996; Okada et al. 1997; Terai et al. 1998; Ohshima and Okada 2005; Kajikawa et al. 2005; Matveev et al. 2007). Through the propagation mechanism called retrotransposition, SINEs occupy a large fraction of vertebrate genomes; for example, over 1.5 million SINE copies occupy 14% of the human genome (Lander et al. 2001). It is an interesting observation that, unlike LINEs, distribution of a SINE family is generally restricted to a certain taxonomic group such as orders/families (Kramerov and Vassetzky 2005; Ohshima and Okada, 2005; Nishihara et al. 2007; Nishihara and Okada 2008), and currently ~200 SINE families/subfamilies are known in various clades in Metazoa, as reported in Repbase (Bao et al. 2015) and in SINEBase (Vassetzky and Kramerov 2013). Thus, SINEs are one of the major genetic elements that determine a clade-specific genomic composition.

Except for the 5′ promoters of the tRNA origin and the 3′ LINE-related tails, SINEs in general have a unique sequence in their central regions. However, in some cases, different SINE families that are distributed across a relatively wide variety of taxa share their central sequences. Such a group of SINE families that share a similar central sequence is called a SINE superfamily, and four SINE superfamilies have been reported so far. The first reported example is the CORE-SINE superfamily (Gilbert and Labuda 1999, 2000). The members of the CORE-SINEs, such as MIR (vertebrates) (Smit and Riggs 1995), Mar1 and Mar3 (marsupials) (Gilbert and Labuda 1999; Munemasa et al. 2008), CoeSINE2 and CoeSINE3 (coelacanth) (Nikaido et al. 2013), AFC (cichlids) (Takahashi et al. 1998; Brawand et al. 2014), BflSINE1 (lancelet) (Nishihara et al. 2006), and OR1 (octopus) (Ohshima and Okada 1994), share a homologous ~60-bp sequence in their central region. Similarly, the V-SINEs, which were originally considered to be confined in vertebrates, have an 80-bp central domain (Ogiwara et al. 2002) and have such family members as DANA (zebrafish), Ras1 (rasbora), Ac1 (fugu and medaka), Lun1 (lungfish), HE1 (shark and ray), and Lam1 (lamprey), as well as SINE2-2_Adi and SINE2-2_NV in cnidarians. The third SINE superfamily, the DeuSINEs, consists of >10 SINE families such as AmnSINE1 (amniotes), LmeSINE1 (coelacanth), SINE3-1 (zebrafish) (Kapitonov and Jurka 2003), OS-SINE1 (salmon), SacSINE1 (shark), EbuSINE1 (hagfish), BflSINE1 (lancelet; showing a hybrid structure of CORE and Deu), and SINE2-3_SP (sea urchin) (Nishihara et al. 2006). In addition, other domains which are widely distributed across metazoans were found to be derived members of DeuSINEs (Piskurek and Jackson 2011; Wang et al. 2012; Luchetti and Mantovani 2013). Therefore, the V-SINEs and DeuSINEs are distributed in Eumetazoa (from vertebrates to Cnidaria). The members of the other SINE superfamily, the Ceph-SINEs, are found only in squid and cuttlefish (Akasaki et al. 2010).

It is an interesting question to ask whether only these four SINE superfamilies exist in the animal genomes or whether there are many unknown superfamilies hidden in their genomes. Although a huge amount of genomic and expressed sequence tag (EST) data are now available, exploration for such SINEs has lagged because the available genomic data for a wide variety of metazoan species are limited and because detailed characterization of TEs has not caught up with the recent acceleration in the production of sequence data. To obtain a thorough understanding of the diversity of SINEs, particularly SINE superfamilies in metazoan genomes, it is necessary 1) to determine the variety of SINEs present in each phylogenetic clade and 2) to search for unknown SINE superfamilies based on the characterization of these SINE data. To address the first issue, it is essential to analyze the genomes of various metazoan groups in which SINEs are not known to be present.

Bivalves are large groups of mollusks. Because the origin of bivalves can be traced back to the Cambrian period (Plazzi and Passamonti 2010; Bieler et al. 2014), it is important to reveal an entire picture of the genomic diversity of bivalves with respect to TEs. Recently, large transcriptome data became available also for molluscs and the first phylogenetic reconstructions based on such data are being proposed (Kocot et al. 2011; Smith et al. 2011; González et al. 2015). As noted by Sharma et al. (2012), these reconstructions largely corroborated previous hypotheses on bivalve phylogeny and led to "an unprecedented stockpile of efficacious molecular loci," to be used as phylogenetic markers, but the large data are also expected to be used for many other genomic analyses. However, bivalve SINEs have not been isolated in spite of the availability of extensive genomic and EST sequence data from dozens of species as well as the draft genomes of two species (Zhang et al. 2012; Takeuchi et al. 2012).

In this study, we discovered that 15 SINEs, among which 13 are newly characterized, constitute a novel SINE superfamily, the MetaSINEs. This new SINE superfamily is found across a variety of metazoan species from fish to cnidarians. In addition, we thoroughly searched for members of the five SINE superfamilies, including the MetaSINEs, in bivalves and characterized eight new SINE members belonging to the CORE-SINEs, V-SINEs, DeuSINEs, and MetaSINEs. These data showed that SINEs from the CORE-SINEs, V-SINEs, and Deu-SINEs are unexpectedly widely distributed among metazoan species. The current data will help us to carry out further deep characterization of metazoan genomes to unveil new SINE superfamilies, with the goal of understanding the comprehensive evolutionary history of genomes inhabited by SINE retroposons.

Materials and Methods

Characterization of Novel Members of Three SINE Superfamilies in Bivalves

Initially, we searched for the members of known SINE superfamilies such as the CORE-SINEs, V-SINEs, and DeuSINEs in bivalves with an NCBI BLAST search (Johnson et al. 2008).

The central sequences of MIR, DANA, and AmnSINE1 were used as queries for representatives of CORE-SINE, V-SINE, and DeuSINE superfamilies, respectively, and the BLAST search was conducted against all nucleotide (nr/nt) and EST sequence data by specifying the target organisms as Bivalvia. The V-SINE search produced dozens of hits in *Mizuhopecten yessoensis*, and all hit sequences were collected. By using the *M. yessoensis* sequences as queries, we again conducted an NCBI BLAST search against the same nucleotide and EST data. In the second-round search, hit sequences were obtained from *M. yessoensis*, *Chlamys farreri*, *Hyriopsis cumingii*, and *Tegillarca granosa*. The sequences were aligned with MEGA6 software (Tamura et al. 2013), and consensus sequences were reconstructed for each species. Through this procedure, we characterized two novel SINE families belonging to the V-SINE superfamily, BivaV-SINE1 and BivaV-SINE2. We further conducted an NCBI BLAST search using the consensus sequence from each of the two SINE families as queries, and the hit sequences were again used to reconstruct consensus sequences of additional novel SINEs, which were then used for the next-round BLAST search. We repeated this procedure for at least five rounds and reconstructed the SINE consensus sequences from the GenBank data of *M. yessoensis*, *Ch. farreri*, *H. cumingii*, *T. granosa*, *Ruditapes philippinarum*, *Crassostrea virginica*, *Spisula solidissima*, and *Coelomactra antiquate*. Through this process, new SINE members of the CORE-SINE and DeuSINE superfamilies were found in bivalves based on the similarity of their 3′-terminal sequences with those of BivaV-SINE1 and BivaV-SINE2. Eventually, we characterized eight novel SINE families from the GenBank sequence data of bivalves.

Characterization of MetaSINEs

We again performed an NCBI BLAST search against all the nucleotide and EST data by using the eight novel bivalve SINEs from the previous step as queries and found that, in the genomes of multiple animals, there are many sequences that are similar to the central region of the two bivalve SINEs (BivaMeta-SINE1 and BivaMD-SINE1). For each species, the hit sequences were collected from GenBank and aligned with MEGA6 to reconstruct a consensus sequence of the SINEs. For medaka and lancelet SINEs, we collected the SINE sequences from the UCSC Genome Browser database (Rosenbloom et al. 2015). After collecting the consensus sequences of the related SINEs, another NCBI BLAST search was conducted including the consensus sequences of the newly characterized SINEs as queries under a slightly sensitive condition (gap open penalty = −2). We repeated this procedure for at least 10 rounds and finally characterized 11 novel SINE families. In addition, we found that two SINE families of sea urchins in Repbase (Bao et al. 2015), SINE2-1_SP and SINE2-2_SP, as well as SINE2-1_PL, a subfamily of SINE2-1_SP characterized from the EST data of *Paracentrotus lividus*, are also members of the novel superfamily.

Distribution Analysis of the SINE Families among Bivalves

We performed a polymerase chain reaction (PCR) analysis and an additional NCBI BLAST search to determine the presence of the eight new SINEs in other bivalve species. PCR primers were designed within each of the SINE sequences to identify the presence of the SINE family in the genomes examined (supplementary table S1, Supplementary Material online). Genomic DNA samples from a broadly selected group of bivalve species were used for the PCR analysis (supplementary table S2, Supplementary Material online). PCR was performed with a first denaturation step at 95 °C for 3 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 50 or 55 °C for 30 s, and extension at 72 °C for 1 min, with the use of Ex Taq kit reagents (TaKaRa BIO, Shiga, Japan). The obtained PCR products were cloned with the pGEM-T Vector System (Promega), and at least five cloned DNAs per product were sequenced with an ABI PRISM 3100 DNA sequencer (Applied Biosystems). The sequences from the PCR products have been deposited at DDBJ/EMBL/GenBank under the accessions LC122973-LC123270. The sequences were aligned using MEGA6, and the consensus sequences were constructed (supplementary fig. S1, Supplementary Material online). The SINE consensus sequences were compared with one another and with known SINE superfamilies in vertebrates by using MEGA6 and Genetyx (Genetyx Corporation, Tokyo, Japan). All the SINE consensus sequences have been deposited at Repbase (Bao et al. 2015; http://www.girinst.org/repbase/, last accessed February 19, 2016).

Results and Discussion

Broad Distribution of a Novel SINE Superfamily, the MetaSINEs

Based on a comprehensive database search, we characterized 13 novel SINEs from various metazoan species such as medaka (OrySINE1 from *Oryzias latipes*), salmon (SalSINE1 from *Salmo salar*), hagfish (EptSINE1 from *Eptatretus burgeri*), lancelet (BflSINE2 from *Branchiostoma floridae*), cuttlefish (SepiaSINE2 from *Sepia officinalis*), bivalves (BivaMeta-SINE1 and BivaMD-SINE1), gastropods (LitSINE1 from *Littorina saxatilis* and HalSINE1 from *Haliotis discus*), parchment worm (ChaetoSINE1 from *Chaetopterus variopedatus*), brachiopods (LinSINE1 from *Lingula anatina*), and cnidarians (ClySINE1 from *Clytia hemisphaerica* and HydSINE1 from *Hydractinia symbiolongicarpus*) (fig. 1). The consensus sequence of each of these 13 novel SINEs described above was deduced from multiple members (at least 5 copies) of the SINE family in the species (fig. 2). They, together with two additional SINE families of sea urchin (SINE2-1_SP and SINE2-2_SP in *Strongylocentrotus purpuratus*), were found to be members

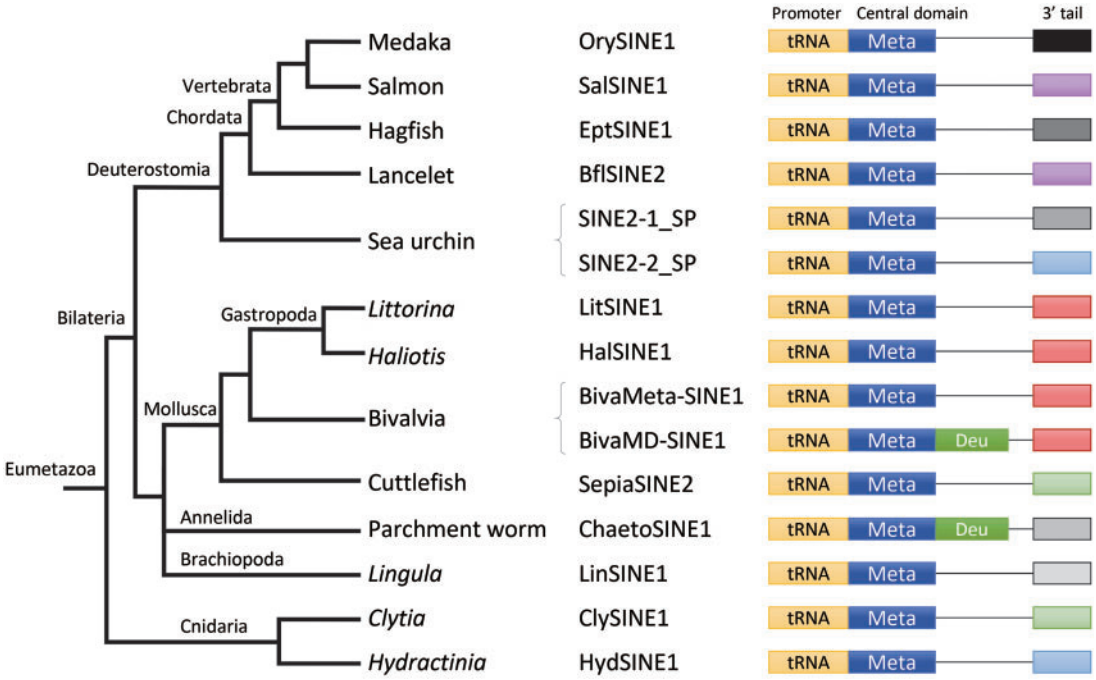


Fig. 1.—Schematic representation of the MetaSINE families and the phylogeny of the host species. All but the sea urchin SINEs are newly characterized in this study. All the MetaSINEs begin with the tRNA-derived promoter (yellow) immediately followed by a similar central sequence, the Meta-domain (blue) (fig. 2). BivaMD-SINE1 and ChaetoSINE1 also have Deu-domain sequences (green) after the Meta-domain. A similar 3′ tail sequence is shared between SINE2-2_SP and HydSINE1 (light blue), between SepiaSINE2 and ClySINE1 (light green), and between SalSINE1 and BflSINE2 (purple), of which the alignments are shown in figure 2*C*(a–c), respectively. In addition, 3′ sequences are shared among LitSINE1, HalSINE1, BivaMeta-SINE1, and BivaMD-SINE1 (red) (fig. 5*E*). OrySINE1 has a 3′ tail sequence similar to that of SINE2-5_SSa and AvaIII SINEs in salmon (fig. 2*C*(d)). The gray boxes indicate unique 3′ sequences.

of a novel SINE superfamily. As shown in the comparison of these consensus sequences in figure 2, all 15 SINE sequences have a 66-bp region consisting of a common central sequence (61–86% identity compared with the consensus sequence) (fig. 2*B*), suggesting that these central domains have a common origin. Based on the characterization, we designated the novel metazoan SINE superfamily as "MetaSINE" and the common central domain as the "Meta-domain." Because the MetaSINE members are widely distributed—from fish to cnidarians—the origin of the MetaSINE superfamily can be traced back to at least 640 Ma (dos Reis et al. 2015), and perhaps as long as 780 Ma (Parfrey et al. 2011).

The promoters of all the MetaSINEs characterized in this study are derived from tRNA. RepeatMasker (http://www.repeatmasker.org, last accessed February 19, 2016) indicated that the promoter sequence of the MetaSINEs shows the highest similarity to the vertebrate tRNA (Thr). Because MetaSINEs originated long ago and the sequences of cognate tRNAs between vertebrates and invertebrates are generally different, it is possible that this sequence similarity does not reflect the real origin of tRNA species of MetaSINEs.

It was of interest to find that the sequence similarity among tRNA-derived sequences of the MetaSINEs was very high across individual members (69–89% identity for the region from the A-box to the end of the promoter region), to an extent that was higher than the similarity between tRNA and each SINE (fig. 2*A*). This observation suggests the possibility that the tRNA-derived regions might not have originated independently among the SINEs but were retained, along with the Meta-domain, from the common ancestor of the MetaSINEs. Beyond the tRNA-derived region and the Meta-domain, the members of MetaSINEs have variable 3′ tails (figs. 1 and 2*C*). Several SINEs share a similar 3′ sequence, such as those between sea urchin SINE2-2_SP and *Hydractinia* HydSINE1 (light blue box in figs. 1 and 2*C*(a)), between cuttlefish SepiaSINE2 and *Clytia* ClySINE1 (light green box in figs. 1 and 2*C*(b)), and between salmon SalSINE1 and lancelet BflSINE2 (purple box in figs. 1 and 2*C*(c)). The 3′ tail of medaka OrySINE1 was similar to that of AvaIII and SINE2-5_SSa (Bao and Jurka 2015) in salmon (fig. 2*C*(d)). In addition, the 3′ regions of three bivalve MetaSINEs resemble that of gastropod MetaSINEs (red box in fig. 1). These data suggest that in each case a particular partner LINE is involved in retrotransposition of a group of SINEs with the same 3′ tail. Although the identity of such partner LINEs of the MetaSINEs is not known at present, each of the SINE groups

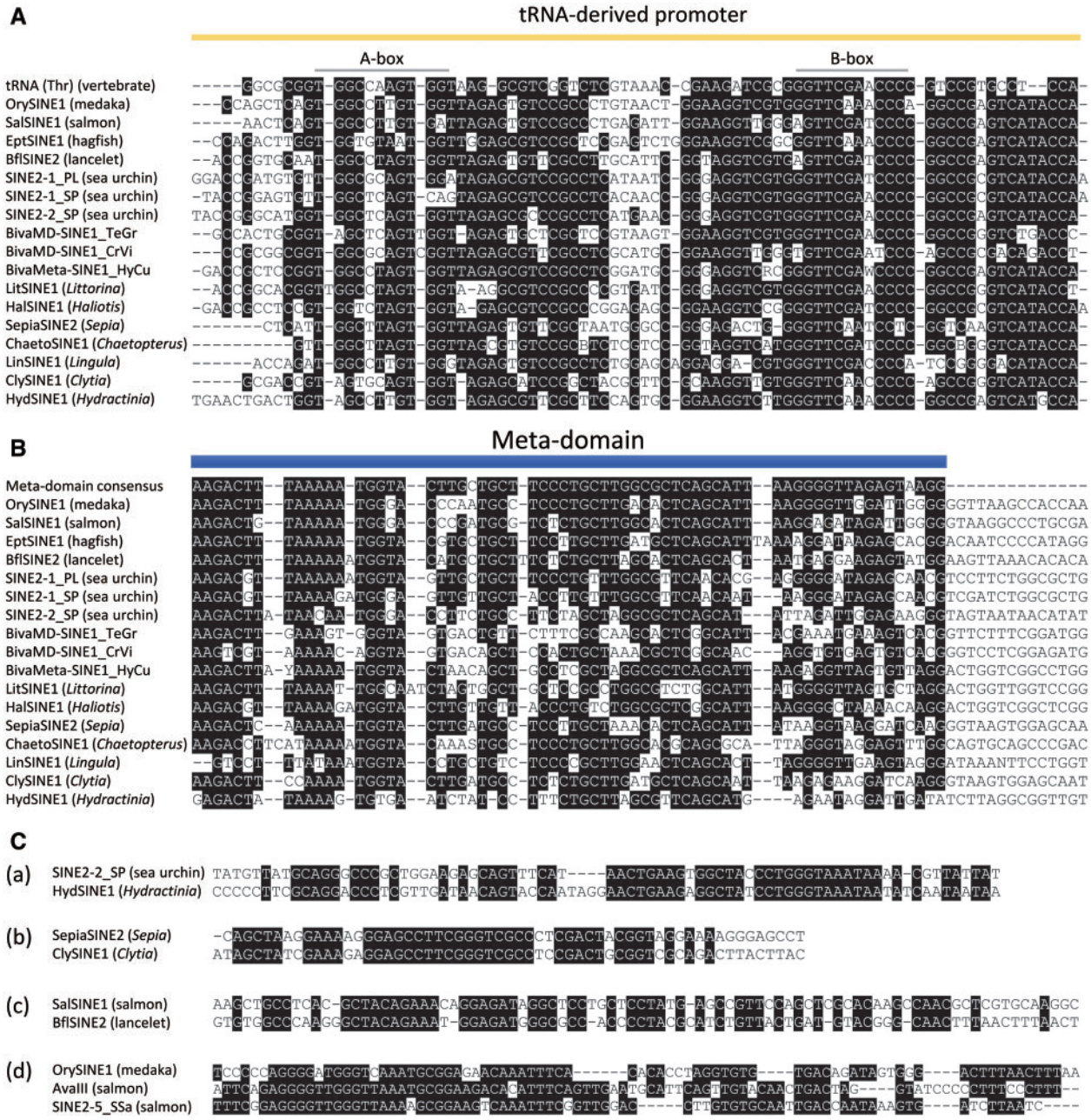


Fig. 2.—Sequence alignments of MetaSINEs. All SINE sequences are the consensus of at least five copies. (*A*) Alignment of the tRNA-derived region among MetaSINEs as well as the vertebrate tRNA (Thr). The A-box and B-box are the internal promoters of RNA Pol III. (*B*) Alignment of the Meta-domain region of MetaSINEs, including the consensus sequence of the Meta-domain. (*C*) Sequence similarities of the 3′ tail sequences between (a) sea urchin SINE2-2_SP and *Hydractinia* HydSINE1, (b) *Sepia* SepiaSINE2 and *Clytia* ClySINE1, (c) salmon SalSINE1 and lancelet BflSINE2, and (d) medaka OrySINE1, salmon AvaIII, and salmon SINE2-5_SSa.

with the same 3′ tail is considered to use a cognate RT for retrotransposition encoded by the same type of LINE (i.e., from a single clade) (Hayashi et al. 2014).

It remains unknown how SINE families with different 3′ tail sequences evolved. One likely possibility is that an RT encoded by a LINE switches its template from the RNA of a SINE/LINE to another SINE during retrotransposition (Buzdin et al. 2003; Nishihara et al. 2006). For example, the promoter regions of a few SINE families in vertebrates have a composite structure of 5S ribosomal RNA (rRNA) and partial tRNA, suggesting a possibility that they were generated by a template switch from a SINE to 5S rRNA (Nishihara et al. 2006). Likewise, the hybrid

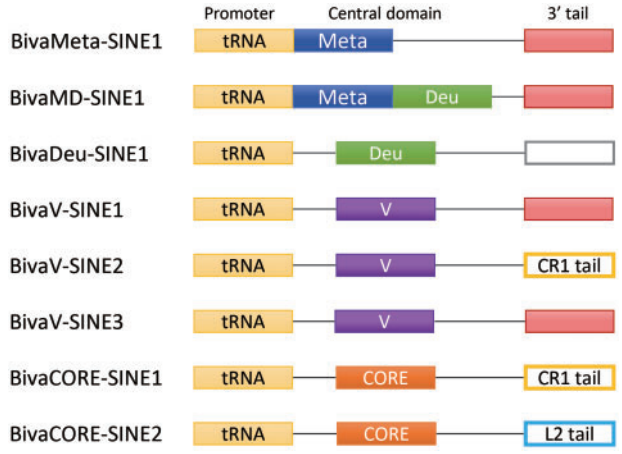


Fig. 3.—Schematic representation of the eight novel SINE families in bivalves. Each SINE contains a tRNA-derived promoter. BivaMeta-SINE1 and BivaDeu-SINE1 contain the Meta-domain and Deu-domain, respectively. BivaMD-SINE1 contains both the Meta- and Deu-domain sequences. Diagnostic central sequences of V-SINEs and CORE-SINEs are present in three and two other SINE families, respectively. The 3′ tail sequences of BivaV-SINE2 and BivaCORE-SINE1 are similar to that of the CR1 LINE, whereas the 3′ sequence of BivaCORE-SINE2 is similar to that of the L2 LINE (fig. 5*A*–*D*). Red boxes represent homologous sequences shared among the four SINEs as well as the gastropod MetaSINEs (fig. 5*E*).

SINEs, such as BivaMD-SINE1, that have both a Meta-domain and a partial Deu-domain were presumably produced via an inter-SINE template switching event. Among vertebrates, the most abundant LINE family in a particular host species differs, such as L1 in mammals (Lander et al. 2001), CR1 in birds (Hillier et al. 2004), and L2 in cichlids (Brawand et al. 2014). This suggests that the 3′ tail of the retrotranspositionally active SINE is also likely to differ among metazoan species depending on the particular LINE family that is most prevalent. Therefore, repeated template switching events that lead to changes in the 3′ tail that are related to an active LINE family might be an advantage for SINE survival. The long-time survival of MetaSINEs during evolution may be explained by the ability of the members of this SINE superfamily to change their 3′ tail according to a dominant LINE in the species.

Eight Novel SINEs in Bivalves Have a Variety of Structures

To understand whether a variety of SINEs belonging to SINE superfamilies is present in a certain metazoan clade, we searched for members of the MetaSINEs, DeuSINEs, V-SINEs, and CORE-SINEs in Bivalvia, because no such SINEs have been reported thus far from this group. Based on a comprehensive BLAST search of genomic and EST data from GenBank, we characterized eight new SINE families (fig. 3). Therefore, the four major SINE superfamilies are present in the bivalve genomes. Except for the two bivalve MetaSINEs, BivaMeta-SINE1, and BivaMD-SINE1 (fig. 2*A*), the tRNA regions of the bivalve SINEs do not exhibit high similarity to one another beyond the A- and B-box promoters of Pol III, suggesting that they may have originated from different tRNAs.

BivaMeta-SINE1 and BivaMD-SINE1, both of which contain the Meta-domain, belong to the MetaSINE superfamily (figs. 1 and 2). It is interesting that BivaMD-SINE1 has both a Meta- and Deu-domain in its sequence, which was presumably generated via template switching. Another member of the DeuSINE superfamily, BivaDeu-SINE1, was also found. The Deu-domain consensus sequences, which were based on BivaMD-SINE1 and BivaDeu-SINE1 from various bivalve species, were determined and compared with other known DeuSINEs (fig. 4*A*). The vertebrate DeuSINEs share a 330-bp Deu-domain (green bar in fig. 4*A*; Nishihara et al. 2006), but only a part of the Deu-domain (~120 bp, referred to as the Nin-domain in Piskurek and Jackson 2011) is conserved among the corresponding SINEs in other animals such as Cnidaria, Arthropoda, Annelida, and Gastropoda (blue bar in fig. 4*A*). One of the newly characterized DeuSINEs in bivalves, BivaMD-SINE1, contains 110 bp of the Deu-domain having the Nin-domain (120-bp) with a partial deletion (blue bar in fig. 4*A*), whereas the other SINE family, BivaDeu-SINE1, has an even shorter region of this domain (70 bp, light blue bar in fig. 4*A*), as does that of *Nve*-Nin-DC-SINE in *Nematostella* (Piskurek and Jackson 2011). Therefore, it is likely that the central domain of DeuSINEs was originally short, and additional conserved domains were added in the common ancestor of jawed vertebrates during their evolution.

Three families within the V-SINE superfamily, BivaV-SINE1, BivaV-SINE2, and BivaV-SINE3, have central sequences similar to those of other V-SINE members. Comparison of the V-domains revealed that BivaV-SINE1 and BivaV-SINE2 share an entire sequence of the vertebrate V-domain (80 bp; purple bar in fig. 4*B*), whereas BivaV-SINE3 has only the 5′ half (~40 bp) of the vertebrate V-domain (pink bar in fig. 4*B*).

Two SINE families (BivaCORE-SINE1 and BivaCORE-SINE2) belong to the CORE-SINE superfamily. The CORE sequence of BivaCORE-SINE2 is 65 bp in length and has 66–70% identity with other CORE-SINEs (orange bar in fig. 4*C*). The CORE sequence of BivaCORE-SINE1 has a higher similarity with SINE_AFC in cichlid fishes than other CORE-SINEs across an ~80-bp region (yellow bar in fig. 4*C*). These two types of CORE-SINEs share only a 45-bp region (highlighted by black in fig. 4*C*), suggesting the possibility that there are two types of CORE-domain sequences. Therefore, we hereafter designate the latter type of CORE-domain as CORE2 (yellow bar in fig. 4*C*). It should be determined in the near future how widely CORE-SINEs with CORE2 are distributed among metazoan genomes.

L2 and CR1 LINEs Have Contributed to the Survival of SINE Superfamilies in Bivalves

To reveal the partner LINEs of the bivalve SINE families, we compared the 3′ tail sequences of the SINEs with those of all

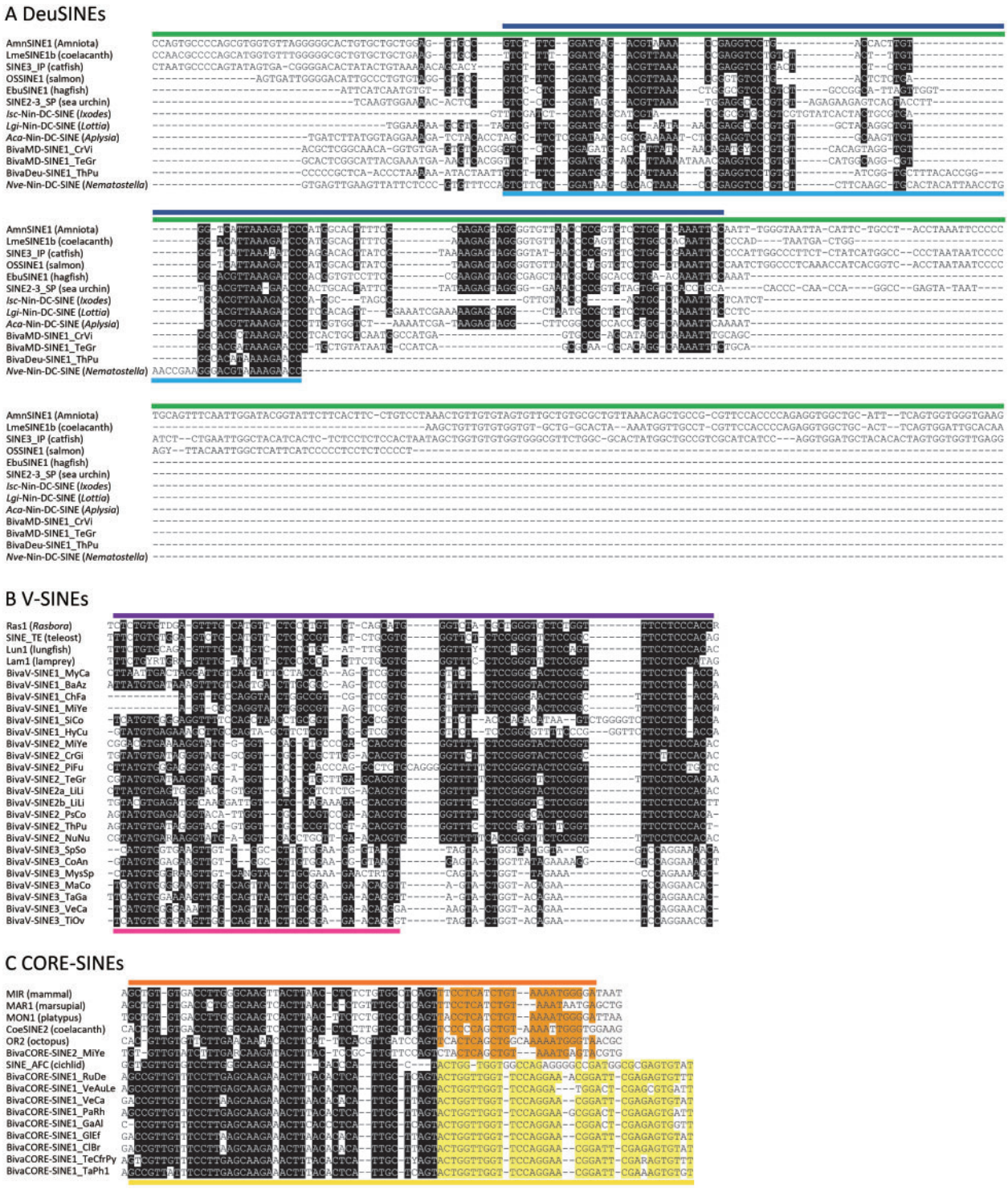


Fig. 4.—Alignments of the central sequences of the three SINE superfamilies (DeuSINEs, V-SINEs, and CORE-SINEs). All sequences are the consensus of the SINEs reconstructed for each species. Refer to supplementary table S2, Supplementary Material online, for the abbreviations of the species names of bivalves. (*A*) Alignment of the Deu-domain sequences of DeuSINE families from bivalves and other animals. Deu-domain sequences of the vertebrate DeuSINE families (330 bp), BivaMD-SINE1 (110 bp), and BivaDeu-SINE1 (70 bp) were shown by green, blue, and light blue lines, respectively. (*B*) Alignment of V-SINEs from vertebrates and bivalves. The homologous regions of BivaV-SINE3 (bottom pink bar) are shorter than those of BivaV-SINE1 and BivaV-SINE2 (top purple bar). (*C*) Alignments of CORE-SINE families from bivalves and other animals. The more 5′ 45-bp CORE sequence is conserved among all the CORE-SINEs (white on black). The more 3′ CORE sequence of BivaCORE-SINE2 is similar to that of CORE-SINEs in mammals, coelacanth, and octopus (orange), whereas this region of BivaCORE-SINE1 is similar to that of cichlid SINE_AFC (referred to as CORE2; yellow).

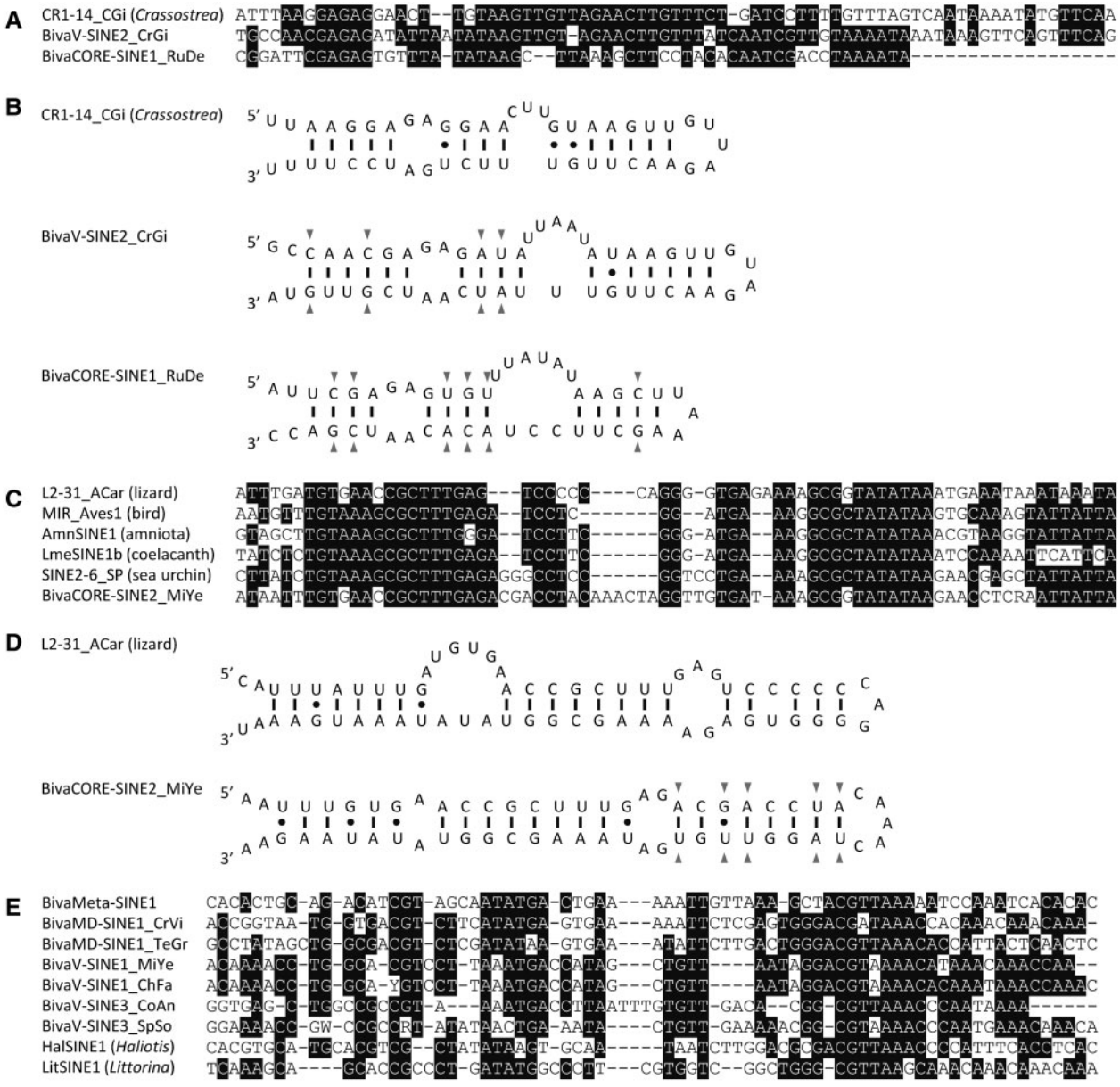
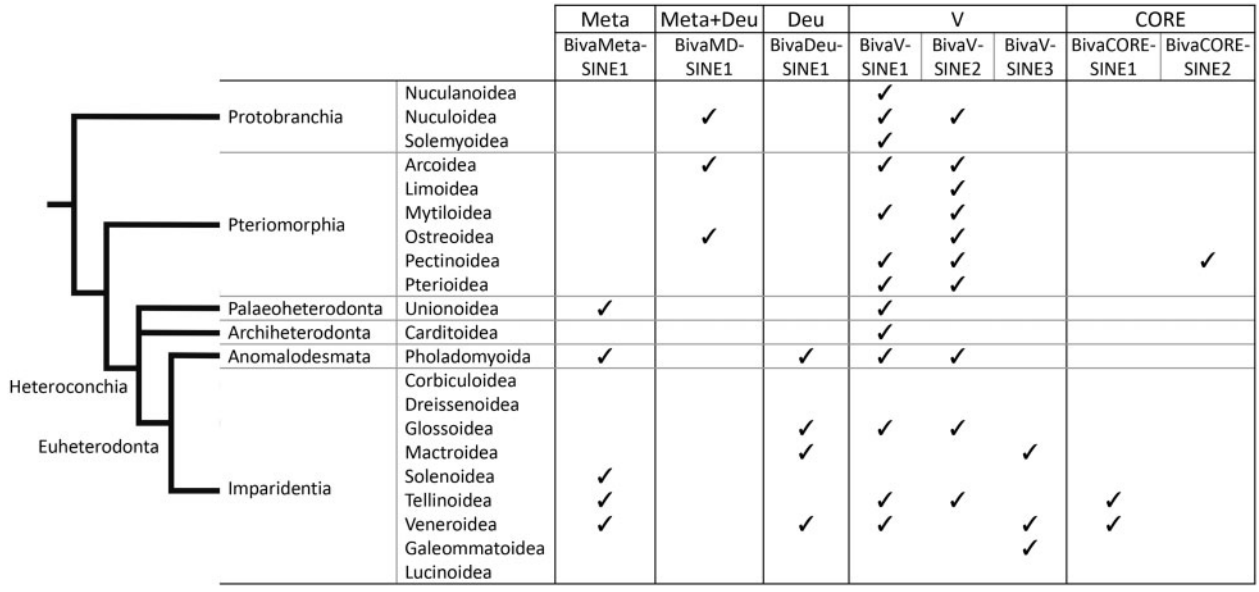


Fig. 5.—Shared 3′ tail sequences among the consensus sequences of the bivalve SINEs. (*A*) CR1-related 3′ tail sequences of BivaV-SINE2 and BivaCORE-SINE1 compared with CR1-14_CGi in *Crassostrea gigas*. (*B*) Putative secondary structures with the conserved stem loops of the 3′ tails of CR1-14_CGi LINE in *Cr. gigas* compared with BivaV-SINE2 (*Cr. gigas*), and BivaCORE-SINE1 (*Ruditapes decussatus*). Hydrogen bonds between paired residues and G-U wobble base pairs are shown as lines and dots, respectively. Compensatory mutations in the SINEs in comparison with corresponding sites of the LINE are denoted by gray arrowheads. (*C*) Comparison of the 3′ tail sequences of an L2 LINE (L2-31_ACar) with BivaCORE-SINE2 and other L2-related SINEs (MIR_Aves1, AmnSINE1, LmeSINE1b, SINE2-6_SP). (*D*) Putative stem loop structures of the 3′ tails of L2-31_ACar LINE in lizard and BivaCORE-SINE2 (*Mizuhopecten yessoensis*). Hydrogen bonds between paired residues and G-U wobble base pairs are shown as lines and dots, respectively. Compensatory mutations in the SINE are denoted by gray arrowheads. (*E*) BivaMeta-SINE1, BivaMD-SINE1, BivaV-SINE1, and BivaV-SINE3 share a similar 3′ tail sequence, suggesting that their retrotransposition may depend on the same clade of LINE.

known LINEs of bivalves and vertebrates obtained from Repbase (Bao et al. 2015). BivaV-SINE2 and BivaCORE-SINE1 share their 3′-terminal sequences with that of CR1-14_CGi in *Crassostrea gigas* (fig. 5*A*). The similarity of the 3′ sequences of these SINEs to that of this LINE family was supported by the formation of possible secondary structures with the conserved stem loops of these 3′ tails. In addition, the presence of several compensatory mutations may validate the possible stable secondary structures observed in the SINEs and the LINE (gray arrowheads in fig. 5*B*). These facts suggest that the two



		Meta	Meta+Deu	Deu	V			CORE	
		BivaMeta-SINE1	BivaMD-SINE1	BivaDeu-SINE1	BivaV-SINE1	BivaV-SINE2	BivaV-SINE3	BivaCORE-SINE1	BivaCORE-SINE2
Protobranchia	Nuculanoidea				✓				
	Nuculoidea		✓		✓	✓			
	Solemyoidea				✓				
Pteriomorphia	Arcoidea		✓		✓	✓			
	Limoidea					✓			
	Mytiloidea				✓	✓			
	Ostreoidea		✓			✓			
	Pectinoidea				✓	✓			✓
	Pterioidea				✓	✓			
Palaeoheterodonta	Unionoidea	✓			✓				
Archiheterodonta	Carditoidea				✓				
Anomalodesmata	Pholadomyoida	✓		✓	✓	✓			
Imparidentia	Corbiculoidea								
	Dreissenoidea								
	Glossoidea			✓	✓	✓			
	Mactroidea			✓			✓		
	Solenoidea	✓							
	Tellinoidea	✓			✓	✓		✓	
	Veneroidea	✓		✓	✓		✓	✓	
	Galeommatoidea						✓		
	Lucinoidea								

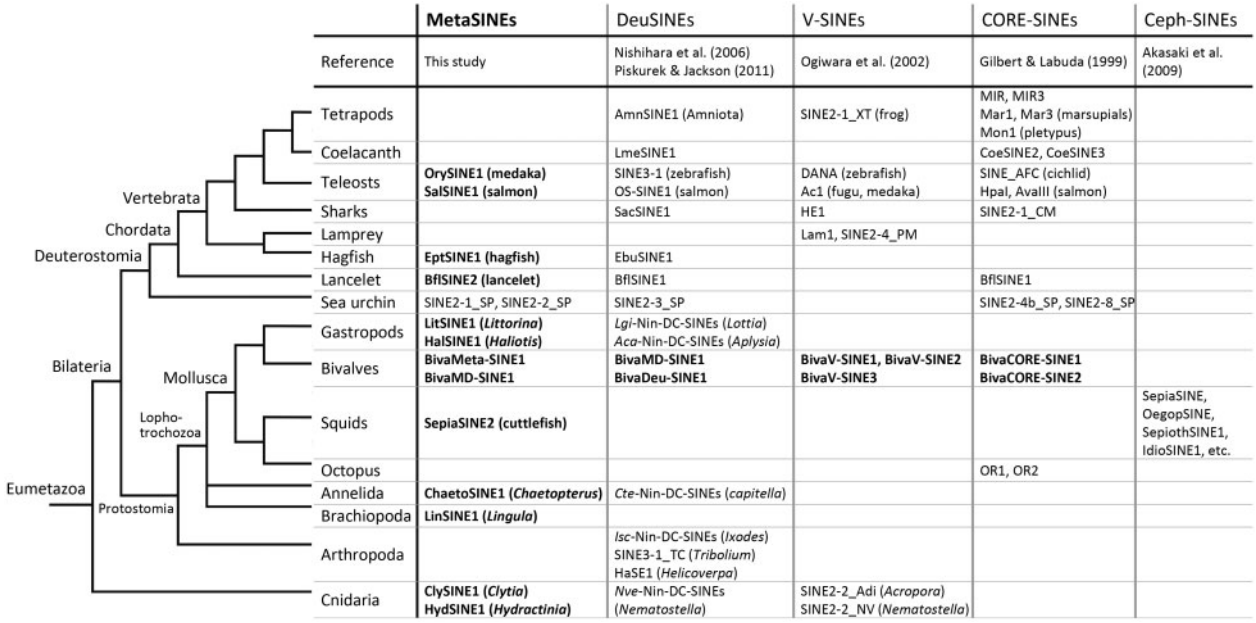
Fig. 6.—Distribution of the eight SINEs in Bivalvia. The presence of each SINE was tested by PCR and by a GenBank search, and the taxonomic clades to which the detected species belonged were confirmed. The phylogenetic relationships among the bivalve subclades are shown to the left (Bieler et al. 2014). Detailed results from the distribution analysis are available in supplementary table S2, Supplementary Material online.

SINE families depend on an RT encoded by CR1 elements for their retrotransposition (Kajikawa and Okada 2002; Ohshima and Okada 2005). As another case, we found that the 3′-terminal sequence of BivaCORE-SINE2 is highly similar to that of L2-31_ACar in lizard (72% identity) as well as other L2-related SINEs such as MIR_Aves1 (bird), AmnSINE1 (Amniota), LmeSINE1b (coelacanth), and SINE2-6_SP (sea urchin) (fig. 5*C*). In addition, the 3′ sequence of BivaCORE-SINE2 was shown to form a stable stem loop structure with possible compensatory mutations in comparison with that of L2 LINE (fig. 5*D*). Therefore, it is most likely that L2 LINE is a partner of BivaCORE-SINE2 for its retrotransposition. Among the other bivalve SINEs, BivaV-SINE1, BivaV-SINE3, BivaMD-SINE1, and BivaMeta-SINE1 in bivalves as well as HalSINE1 and LitSINE1, the newly characterized gastropod MetaSINEs, share a homologous 3′ tail (red boxes in figs. 1 and 3, and fig. 5*E*). This observation suggests that they depend on the same clade of LINEs for their propagation. Therefore, the bivalve SINE superfamilies, including the MetaSINEs, have survived by relying on the retrotranspositional machinery of multiple types of LINEs.

At Least Three SINE Families Originated in the Common Ancestor of Bivalvia

To examine how widely the eight new SINE families are distributed in taxonomic groups of bivalves, we performed PCR and sequencing using the genomic DNA of broadly sampled bivalve species (supplementary tables S1 and S2, Supplementary Material online). In addition, we also performed a BLAST search against the nucleotide and EST data of GenBank. Bivalve systematics is still debated, both in terms of taxonomy and phylogenetic relationships (Carter et al. 2011; Plazzi et al. 2011; Sharma et al. 2012; Bieler et al. 2014; González et al. 2015; and reference therein). However, notwithstanding the assigned taxonomic rank, five main subclades are generally accepted: Protobranchia, Pteriomorphia, Palaeoheterodonta, Anomalodesmata, and Imparidentia *sensu* (Bieler et al. 2014). Our distribution analysis revealed that, among these main subclades of Bivalvia, BivaMD-SINE1, BivaV-SINE1, and BivaV-SINE2 are present in multiple subclades including Protobranchia (fig. 6). Because Protobranchia diverged first among bivalves (Plazzi and Passamonti 2010; Kocot et al. 2011; Plazzi et al. 2011; Smith et al. 2011; Bieler et al. 2014; González et al. 2015), the three SINE families originated in a common ancestor of Bivalvia before their divergence, which is estimated to be 530 Ma (Plazzi and Passamonti 2010; Bieler et al. 2014).

In addition, BivaMeta-SINE1 was detected in Heteroconchia, a wider taxon encompassing Palaeoheterodonta, Archiheterodonta, Anomalodesmata, and Imparidentia. Although this cluster is still questioned (Plazzi et al. 2011; Sharma et al. 2012; and reference therein), Bieler et al. (2014) estimated its age to the Middle Cambrian (around 500 Ma). Therefore, the origin of the BivaMeta-SINE1 can be traced back at least to this point or even deeper, if Palaeoheterodonta were actually the first diverging bivalve clade after the Protobranchia split (Doucet-Beaupré et al. 2010; Plazzi et al. 2011). BivaDeu-SINE1 is found in various



	MetaSINEs	DeuSINEs	V-SINEs	CORE-SINEs	Ceph-SINEs
Reference	This study	Nishihara et al. (2006) Piskurek & Jackson (2011)	Ogiwara et al. (2002)	Gilbert & Labuda (1999)	Akasaki et al. (2009)
Tetrapods		AmnSINE1 (Amniota)	SINE2-1_XT (frog)	MIR, MIR3 Mar1, Mar3 (marsupials) Mon1 (pletypus)	
Coelacanth		LmeSINE1		CoeSINE2, CoeSINE3	
Teleosts	**OrySINE1 (medaka)** **SalSINE1 (salmon)**	SINE3-1 (zebrafish) OS-SINE1 (salmon)	DANA (zebrafish) Ac1 (fugu, medaka)	SINE_AFC (cichlid) HpaI, AvaIII (salmon)	
Sharks		SacSINE1	HE1	SINE2-1_CM	
Lamprey			Lam1, SINE2-4_PM		
Hagfish	**EptSINE1 (hagfish)**	EbuSINE1			
Lancelet	**BflSINE2 (lancelet)**	BflSINE1		BflSINE1	
Sea urchin	SINE2-1_SP, SINE2-2_SP	SINE2-3_SP		SINE2-4b_SP, SINE2-8_SP	
Gastropods	**LitSINE1 (*Littorina*)** **HalSINE1 (*Haliotis*)**	*Lgi*-Nin-DC-SINEs (*Lottia*) *Aca*-Nin-DC-SINEs (*Aplysia*)			
Bivalves	**BivaMeta-SINE1** **BivaMD-SINE1**	**BivaMD-SINE1** **BivaDeu-SINE1**	**BivaV-SINE1, BivaV-SINE2** **BivaV-SINE3**	**BivaCORE-SINE1** **BivaCORE-SINE2**	
Squids	**SepiaSINE2 (cuttlefish)**				SepiaSINE, OegopSINE, SepiothSINE1, IdioSINE1, etc.
Octopus				OR1, OR2	
Annelida	**ChaetoSINE1 (*Chaetopterus*)**	*Cte*-Nin-DC-SINEs (*capitella*)			
Brachiopoda	**LinSINE1 (*Lingula*)**				
Arthropoda		*Isc*-Nin-DC-SINEs (*Ixodes*) SINE3-1_TC (*Tribolium*) HaSE1 (*Helicoverpa*)			
Cnidaria	**ClySINE1 (*Clytia*)** **HydSINE1 (*Hydractinia*)**	*Nve*-Nin-DC-SINEs (*Nematostella*)	SINE2-2_Adi (*Acropora*) SINE2-2_NV (*Nematostella*)		

Fig. 7.—Representative members of the five SINE superfamilies and the phylogenetic relationships of the host species (Kocot et al. 2011; Telford et al. 2015), showing a widespread distribution of the four major SINE superfamilies (the MetaSINEs, DeuSINEs, V-SINEs, and CORE-SINEs) in Eumetazoa including bivalves. The newly characterized SINEs from this study are shown in bold.

orders/superfamilies of Imparidentia and Anomalodesmata (Pholadomyoida), again suggesting an ancient origin of the SINE family, given the estimated split age of these two groups (490 Ma; Bieler et al. 2014). The distributions of the other three SINE families, BivaV-SINE3, BivaCORE-SINE1, and BivaCORE-SINE2, are restricted to only one or a few families. Overall, at least three members of the SINE superfamilies have been retained in the genomes of bivalves for a long evolutionary time (at least 530 Myr).

Possible Function of the Central Domains of the SINE Superfamilies

Although it is still unknown why these SINEs share such similar sequences in the central region, such a broad distribution and high degree of sequence conservation during evolution suggest that this region might have provided an advantageous function for SINEs as well as for the host species. For example, it is possible that the central domains of the SINEs may increase the transcriptional activity or posttranscriptional RNA stability of a SINE during retrotransposition or that the SINE central sequence may be subjected to epigenetic regulation for transcription by the host such that SINE retrotransposition can be spatiotemporally controlled. Another possibility is that the central domain provides a beneficial function to the host genome during evolution. For example, hundreds of AmnSINE1 (DeuSINE) copies are under purifying selection in mammals, and some of these copies act as a distal enhancer of developmental genes (Nishihara et al. 2006; Sasaki et al 2008; Hirakawa et al. 2009; Okada et al. 2010; Tashiro et al. 2011; Nakanishi et al. 2012). In addition, a CORE-SINE sequence has an enhancer function in mammals (Santangelo et al. 2007). Further characterization of other members of the SINE superfamilies and a comparative sequence analysis as well as functional analyses of the individual members may reveal their hidden roles, which could help to explain why eukaryotes have allowed the existence of such a variety of SINEs during evolution.

The areas of conservation along the length of the central domain in nonvertebrate SINEs (fig. 4) may be useful for exploring the function of this region. The length of the central sequence was originally considered to be ~330, 80, and 65 bp for DeuSINEs (Nishihara et al. 2006), V-SINEs (Ogiwara et al. 2002), and CORE-SINEs (Gilbert and Labuda 1999), respectively, whereas the corresponding bivalve SINEs, BivaDeuSINE1, BivaV-SINE3, and BivaCORE-SINE1 and 2, share only ~70, 40, and 45 bp, respectively, of this central domain (fig. 4). Therefore, a functional analysis focusing on this more restricted region of the bivalve sequence may reveal the potential role of the central domains of the SINE superfamilies.

A Diversity of SINEs in Animal Genomes

The MetaSINEs characterized here represent the fifth SINE superfamily distributed across metazoan genomes, in addition to the CORE-SINEs, V-SINEs, DeuSINEs, and Ceph-SINEs (fig. 7). The members of the SINE superfamilies have contributed to diversification of their host genomes during their long

evolution (640 Myr; dos Reis et al. 2015). Given the discovery of this novel SINE superfamily comprising 15 SINEs, it is expected that there are still unknown SINE superfamilies hidden in metazoan genomes. Therefore, further exploration of various animal genomes for novel SINEs is important to reveal what kinds of TEs have been generated and retained during metazoan evolution. It is also important to understand the diversity of TEs within the genomes of each taxonomic group of Metazoa. In this study, we found that eight SINE families, representing the major superfamilies (the MetaSINEs, DeuSINEs, V-SINEs, and CORE-SINEs) are present in the bivalve genomes (figs. 6 and 7). Given the widespread distribution of the SINE superfamilies in bivalve genomes, we expect that a variety of SINE members of the superfamilies will be discovered in other eumetazoan phyla/classes.

For the time being, the entire picture of the characterization and distribution of TEs will remain obscured because of the biased knowledge of metazoan genomes. For example, in the NCBI database, genomic assemblies of 560 metazoan species are available, comprising 257 chordate species (45.9%), 210 arthropods (37.5%), 47 roundworms (8.40%), 12 flatworms (2.1%), 8 cnidarians (1.4%), and 26 other metazoan species (only 4.6%). Therefore, to elucidate the enigma of why such a variety of SINE superfamilies have originated and been retained during animal evolution, broad and deep genomic analyses of both major and minor metazoan phylogenetic taxa will be important. In addition to such efforts, future studies to elucidate the function of the central domain will provide new insights concerning the strategies of retroposons for their long-term survival during evolution.

After submission of our work, Matetovici et al (2016) has independently published a similar work regarding analysis of mollusk SINEs. The sequence of the "MESC-SINE" superfamily, which they characterized from mollusks, is basically the same as that of MetaSINEs described here, but their distribution (bivalve and gastropod) covers only a part of the phylogenetic groups that the SINE members of MetaSINEs cover (from fish to cnidarians). In addition, the central "MESC-domain" (100 bp) is broader than that of Meta-domain (66 bp), due to their characterization of SINEs from more restricted phylogenetic groups (fig. 1). Matetovici et al (2016) also discovered the same SINEs corresponding to BivaV-SINE1, BivaV-SINE2, and BivaCORE-SINE2 in this study.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 21227002, 26291075, 23128503, 25125719 to N.O. as well as by JSPS KAKENHI 23123504, 26840117, 26106004 to H.N. and the Naito Foundation Natural Science Scholarship to H.N. This work was also supported by "Marco Polo" grant to F.P., and Canziani Bequest to M.P.

Supplementary Material

Supplementary tables S1 and S2 and figure S1 are available at *Genome Biology and Evolution online* (http://www.gbe.oxfordjournals.org/).

Literature Cited

Akasaki T, et al. 2010. Characterization of a novel SINE superfamily from invertebrates: "Ceph-SINEs" from the genomes of squids and cuttlefish. Gene 454:8–19.

Bao W, Jurka J. 2015. Non-LTR retrotransposons from the Atlantic salmon. Repbase Rep. 15:749.

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

Bieler AN, et al. 2014. Investigating the Bivalve Tree of Life—an exemplar-based approach combining molecular and novel morphological characters. Invertebrate Systematics 28:32–115.

Brawand D, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

Buzdin A, et al. 2003. The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. Nucleic Acids Res. 31:4385–4390.

Carter JG, et al. 2011. A synoptical classification of the Bivalvia (Mollusca). Paleontol Contrib. 4:1–47.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 10:691–703.

Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. 2003. Mobile elements and mammalian genome evolution. Curr Opin Genet Dev. 13:651–658.

dos Reis M, et al. 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr Biol. 25:2939–2950.

Doucet-Beaupré H, et al. 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evol Biol. 10:50.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nat Rev Genet. 9:397–405.

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.

González VL, et al. 2015. A phylogenetic backbone for Bivalvia: an RNA-seq approach. Proc Biol Sci. 282:20142332

Hayashi Y, Kajikawa M, Matsumoto T, Okada N. 2014. Mechanism by which a LINE protein recognizes its 3′ tail RNA. Nucleic Acids Res. 42:10605–10617.

Hillier LW, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716.

Hirakawa M, Nishihara H, Kanehisa M, Okada N. 2009. Characterization and evolutionary landscape of AmnSINE1 in Amniota genomes. Gene 441:100–110.

Johnson M, et al. 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36:W5–W9.

Kajikawa M, Ichiyanagi K, Tanaka N, Okada N. 2005. Isolation and characterization of active LINE and SINEs from the eel. Mol Biol Evol. 22:673–682.

Kajikawa M, Okada N. 2002. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 111:433–444.

Kapitonov VV, Jurka J. 2003. A novel class of SINE elements derived from 5S rRNA. Mol Biol Evol. 20:694–702.

Kazazian HH Jr. 2004. Mobile elements: drivers of genome evolution. Science 303:1626–1632.
Kocot KM, et al. 2011. Phylogenomics reveals deep molluscan relationships. Nature 477:452–456.
Kramerov DA, Vassetzky NS. 2005. Short retroposons in eukaryotic genomes. Int Rev Cytol. 247:165–221.
Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.
Luchetti A, Mantovani B. 2013. Conserved domains and SINE diversity during animal evolution. Genomics 102:296–300.
Matetovici I, et al. 2016. Mobile element evolution playing jigsaw—SINEs in gastropod and bivalve mollusks. Genome Biol Evol. 8:253–270.
Matveev V, Nishihara H, Okada N. 2007. Novel SINE families from salmons validate Parahucho (Salmonidae) as a distinct genus and give evidence that SINEs can incorporate LINE-related 3′-tails of other SINEs. Mol Biol Evol. 24:1656–1666.
Munemasa M, et al. 2008. Newly discovered young CORE-SINEs in marsupial genomes. Gene 407:176–185.
Nakanishi A, et al. 2012. A SINE-derived element constitutes a unique modular enhancer for mammalian diencephalic Fgf8. PLoS One 7:e43785
Nikaido M, et al. 2013. Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome Res. 23:1740–1748.
Nishihara H, Kuno S, Nikaido M, Okada N. 2007. MyrSINEs: a novel SINE family in the anteater genomes. Gene 400:98–103.
Nishihara H, Okada N. 2008. Retroposons: genetic footprints on the evolutionary paths of life. Methods Mol Biol. 422:201–225.
Nishihara H, Smit AF, Okada N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. 16:864–874.
Ogiwara I, Miya M, Ohshima 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.
Ohshima K, Hamada M, Terai Y, Okada N. 1996. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. Mol Cell Biol. 16:3756–3764.
Ohshima K, Okada N. 1994. Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposons in the octopus. J Mol Biol. 243:25–37.
Ohshima K, Okada N. 2005. SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. Cytogenet Genome Res. 110:475–490.
Okada N. 1991a. SINEs. Curr Opin Genet Dev. 1:498–504.
Okada N. 1991b. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol Evol. 6:358–361.
Okada N, Hamada M, Ogiwara I, Ohshima K. 1997. SINEs and LINEs share common 3′ sequences: a review. Gene 205:229–243.
Okada N, Sasaki T, Shimogori T, Nishihara H. 2010. Emergence of mammals by emergency: exaptation. Genes Cells 15:801–812.
Parfrey LW, Lahr DJ, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A. 108:13624–13629.
Piskurek O, Jackson DJ. 2011. Tracking the ancestry of a deeply conserved eumetazoan SINE domain. Mol Biol Evol. 28:2727–2730.
Plazzi F, Ceregato A, Taviani M, Passamonti M. 2011. A molecular phylogeny of bivalve mollusks: ancient radiations and divergences as revealed by mitochondrial genes. PLoS One 6:e27147
Plazzi F, Passamonti M. 2010. Towards a molecular phylogeny of Mollusks: bivalves' early evolution as revealed by mitochondrial genes. Mol Phylogenet Evol. 57:641–657.
Rogers JH. 1985. The origin and evolution of retroposons. Int Rev Cytol. 93:187–279.
Rosenbloom KR, et al. 2015. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 43:D670–D681.
Sakamoto K, Okada N. 1985. Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J Mol Evol. 22:134–140.
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-specific brain formation. Proc Natl Acad Sci U S A. 105:4220–4225.
Sharma PP, et al. 2012. Phylogenetic analysis of four nuclear protein-encoding genes largely corroborates the traditional classification of Bivalvia (Mollusca). Mol Phylogenet Evol. 65:64–74.
Smit AF. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev. 9:657–663.
Smit AF, Riggs AD. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res. 23:98–102.
Smith SA, et al. 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480:364–367.
Takahashi K, Terai Y, Nishida M, Okada N. 1998. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. Mol Biol Evol. 15:391–407.
Takeuchi T, et al. 2012. Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. DNA Res. 19:117–130.
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.
Tashiro K, et al. 2011. A mammalian conserved element derived from SINE displays enhancer properties recapitulating Satb2 expression in early-born callosal projection neurons. PLoS One 6:e28497
Telford MJ, Budd GE, Philippe H. 2015. Phylogenomic insights into animal evolution. Curr Biol. 25:R876–R887.
Terai Y, Takahashi K, Okada N. 1998. SINE cousins: the 3′-end tails of the two oldest and distantly related families of SINEs are descended from the 3′ ends of LINEs with the same genealogical origin. Mol Biol Evol. 15:1460–1471.
Vassetzky NS, Kramerov DA. 2013. SINEBase: a database and tool for SINE analysis. Nucleic Acids Res. 41:D83–D89.
Wang J, et al. 2012. Characterization of three novel SINE families with unusual features in *Helicoverpa armigera*. PLoS One 7:e31355
Zhang G, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490:49–54..

Associate editor: Ellen Pritham