# LINEs Contribute to the Origins of Middle Bodies of SINEs besides 3′ Tails

Kenji K. Kojima[1,2,*]

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

[2]Genetic Information Research Institute, Mountain View, California

*Corresponding author: E-mail: kojimakk@mail.ncku.edu.tw.

**Accepted:** January 8, 2018

## Abstract

Short interspersed elements (SINEs), which are nonautonomous transposable elements, require the transposition machinery of long interspersed elements (LINEs) to mobilize. SINEs are composed of two or more independently originating parts. The 5′ region is called the "head" and is derived mainly from small RNAs, and the 3′ region ("tail") originates from the 3′ region of LINEs and is responsible for being recognized by counterpart LINE proteins. The origin of the middle "body" of SINEs is enigmatic, although significant sequence similarities among SINEs from very diverse species have been observed. Here, a systematic analysis of the similarities among SINEs and LINEs deposited on Repbase, a comprehensive database of eukaryotic repeat sequences was performed. Three primary findings are described: 1) The 5′ regions of only two clades of LINEs, *RTE* and *Vingi*, were revealed to have contributed to the middle parts of SINEs; 2) The linkage of the 5′ and 3′ parts of LINEs can be lost due to occasional tail exchange of SINEs; and 3) The previously proposed Ceph-domain was revealed to be a fusion of a CORE-domain and a 5′ part of *RTE* clade of LINE. Based on these findings, a hypothesis that the 5′ parts of bipartite nonautonomous LINEs, which possess only the 5′ and 3′ regions of the original LINEs, can contribute to the undefined middle part of SINEs is proposed.

**Key words:** SINE, LINE, nonautonomous, internal deletion, RTE, Ceph-domain.

## Introduction

Short interspersed elements (SINEs) are composite mobile elements that can mobilize dependent on the help of counterpart long interspersed elements (LINEs), also called non-long terminal repeat (non-LTR) retrotransposons (Kajikawa and Okada 2002; Dewannieux et al. 2003). SINEs are composed of several independent parts: a head, body, and tail.

The heads of SINEs typically originate from noncoding RNAs such as 7SL RNA, tRNA, and 5S rRNA, which are the key for one classification scheme of SINEs (Kapitonov and Jurka 2003). SINEs with 7SL RNA-derived heads are called SINE1 and are only found in Euarchontoglires (primates, tree shrews, and rodents) (Kriegs et al. 2007). SINEs with tRNA-derived heads are the most widely distributed among eukaryotes and are called SINE2 (Bao et al. 2015). SINEs with 5S rRNA-derived heads are called SINE3 (Kapitonov and Jurka 2003). These SINEs are transcribed by RNA polymerase III depending on the activity of internal promoters inside of these SINE heads. Recently, a new group of SINEs with U1 or U2 snRNA-derived heads was proposed and designated SINEU (Kojima 2015). SINEs with 28S rRNA-derived sequences (SINE28) and with GC-rich sequences of unknown origins have also been proposed (Longo et al. 2015; Suh et al. 2016). High copy numbers of these newly proposed SINEs with nearly identical structures suggest that they are retrotransposition units, and not chimeric copies derived from two or more RNA templates, although the transcription mechanism for these SINEs are not yet demonstrated.

The 3′ termini of SINEs are called tails and are responsible for the mobilization of SINEs. Tails often exhibit sequence similarity to 3′ regions of LINEs, and the secondary structure of the tail is recognized by the proteins encoded by LINEs (Ohshima et al. 1996; Kajikawa and Okada 2002). However, many SINEs, represented by *Alu* elements from primates, do not have 3′ tail shared by their counterpart LINEs. In the case of *Alu*, the counterpart LINE, *LINE-1* (*L1*), can mobilize any RNA with 3′ polyA tail, even mRNAs (Dewannieux et al. 2003). Such "relaxed" recognition of RNA by *L1* proteins is likely the key of the success of many mammalian SINEs without specific 3′ tail sequences (Okada et al. 1997; Ohshima 2012).

The simplest SINEs, such as *B1* and *ID* from rodents, have only a head and a 3′ tail. Some SINEs contain additional sequences unrelated to either LINEs or small RNA genes between the head and the tail. Among the three parts constituting SINEs, the middle part (i.e., the body) is the most enigmatic. This region rarely exhibits any sequence similarity to anything but SINEs. It is of interest because SINEs from very divergent animals sometimes exhibit significant similarity in the body region. Based on this similarity, several groups of SINEs, such as CORE-SINE, V-SINE, and Ceph-SINE, have been proposed (Gilbert and Labuda 1999; Ogiwara et al. 2002; Nishihara et al. 2006, 2016; Akasaki et al. 2010; Piskurek and Jackson 2011; Luchetti and Mantovani 2013). Even these body parts can be composite; sometimes SINEs share just the 5′ half of the body (Piskurek and Jackson 2011; Luchetti and Mantovani 2013; Nishihara et al. 2016).

Although the origins of widely conserved SINE bodies are completely unknown, the middle regions of some narrowly distributed SINEs have been characterized. One major group exhibits a bipartite structure of sequences that originate from LINEs. The bipartite structures often originate from 5′- and 3′-UTR of *RTE*-type LINEs. Examples include *Bov-tA*, *Mar-1*, *AfroSINE*, *Ped-1*, *Ped-2*, *BuceSINE*, *GymnSINE*, *ManaSINE*, and *MeloSINE* (Okada and Hamada 1997; Gilbert and Labuda 2000; Nikaido et al. 2003; Gogolevsky et al. 2008; Suh et al. 2016). The 3′ part originates from the extreme 3′ end including a 3′ polyA or microsatellite tail. The 5′ part is either the extreme 5′ end or an internal sequence inside of the 5′-UTR. We previously reported another type of LINE that can contribute to the bipartite structures of SINEs; the middle and 3′ terminal regions of *SINE2-1_ACar* and *SINE2-1B_ACar* exhibit similarities with the 5′ and 3′ of *Vingi-2_ACar* (Kojima et al. 2011).

Several nonautonomous LINEs possessing only the 5′ parts and 3′ parts of autonomous LINEs have been reported (Bringaud et al. 2003, 2009; Kojima et al. 2011). Their representatives are *RIME* derived from *Ingi* and *NARTc* derived from *L1Tc* (Bringaud et al. 2003, 2009). *Vingi-1_EE* have many nonautonomous derivatives generated due to internal deletion (Kojima et al. 2011). A proposed ancestral retrotransposition unit *Bov-A*, which is the shared part between *Bov-A2* and *Bov-tA*, is an internally deleted derivative of the *Bov-B* LINE (Okada and Hamada 1997). *Bov-A2* is a dimer of two *Bov-A* units, and *Bov-tA* is a combination of a tRNA-derived head and *Bov-A*. These observations—that is, the presence of nonautonomous LINEs with a bipartite structure and SINEs with a bipartite structure plus a 5′ RNA-derived head—raised the possibility that the middle parts of SINEs can originate from a part of LINEs. Here, this hypothesis is expanded to indicate the body of SINEs can be originated by bipartite LINEs even if SINEs do not have bipartite structures.

In this study, systematic analysis of the similarity between SINEs and LINEs and in-between is performed. Several new examples of bipartite structure of *RTE*-type LINEs in SINEs were found. A fragment of an *RTE*-derived sequence contributes to the latter half of the proposed Ceph-domain of SINEs, supporting the hypothesis that the conserved bodies of SINEs can be generated by a part of LINEs.

## Materials and Methods

### Repeat Detection and Classification

Multicopy sequences in published eukaryotic genomes were screened using approaches similar to those described previously in the literature (Bao and Eddy 2002). Screening for low-copy-number repeat sequences was also performed by Censor search (Kohany et al. 2006) with the protein sequences of well-characterized repeat sequences deposited in Repbase (Bao et al. 2015) (http://www.girinst.org/repbase). Classification is based on the similarity to known repeat sequences deposited in Repbase with Censor (Kohany et al. 2006). RTclass1 (Kapitonov et al. 2009) was used to further classify LINEs. All of the repeat sequences detected here have been deposited in Repbase. The similarity between LINEs and SINEs were analyzed with Censor and was confirmed via manual inspection. Sequence alignment was performed using MAFFT (Katoh et al. 2005) and MUSCLE (Edgar 2004) and was visualized using Jalview (Waterhouse et al. 2009) and UGENE (Okonechnikov et al. 2012).

## Results

### The Contribution of Bipartite LINEs to SINEs

The similarity between LINEs and SINEs and between different SINEs was analyzed using Censor with redundant option (Kohany et al. 2006). Censor used BLAST to compare the SINE sequences extracted from Repbase to the LINE sequences or the SINE sequences extracted also from Repbase. All LINE–SINE pairs and SINE–SINE pairs showing sequence similarity detected by Censor were extracted and inspected manually to remove accidental hits. First, the hits on the complementary strand were all removed. Several accidental hits were observed when a LINE had a low-complexity sequence (e.g., the sequence 6897–6977 of *L1-10_Pl*). The presence of a tRNA-like sequence in *RTE-1_DAn* and its relatives results in hits between these LINEs and many SINE2 elements. After removing these hits, the remaining LINE–SINE pairs were analyzed to determine whether the LINE-derived sequences were present in the counterpart SINE besides the 3′ terminus. Because the similarity between LINEs and SINEs at their 3′ termini is common if the SINE is dependent on the transposition machinery of the LINE, this step is essential. Finally, SINEs that have been already reported to possess bipartite LINE structure (*Bov-tA*, *Mar-1*, *AfroSINE*, *Ped-1*, *Ped-2*, *PlatSINE1*, *Plat_RTE1_SINE*, *BuceSINE*, *GymnSINE*, *ManaSINE*, and *MeloSINE*) were removed (Okada and Hamada 1997; Gilbert and Labuda 2000;

Nikaido et al. 2003; Gogolevsky et al. 2008; Bao et al. 2015; Suh et al. 2016). Goat *NLA* repeat is likely a member of *Bov-tA*. The structure, sequence, and distribution of *SINE2-1_Laf* from the African elephant *Loxodonta africana* and *SINE2-1_Pca* from the rock hyrax *Procavia capensis* suggest that they are members of AfroSINEs. *RTESINE1* and *RTESINE2* are both bipartite *RTE*-type nonautonomous LINEs. The final candidates for new bipartite LINE-derived regions seen in SINEs are shown in figure 1 and listed in table 1. The sequences of these SINEs along with information of their composite structure appear in supplementary figure S1, Supplementary Material online, and the alignments between LINEs and SINEs appear in supplementary figure S2, Supplementary Material online.

### *CoeSINE4 and CoeSINE5*

Two coelacanth SINE families, *CoeSINE4* and *CoeSINE5*, have similar 3′ sequences (table 1). These sequences correspond to the 5′- and 3′-UTR of *RTE*-type LINEs. *CoeSINE4* has a tRNA-derived head, and *CoeSINE5* has a 5S rRNA-derived head.

### *HaSE1, HaSE2_DP, SINE2-1_PXu, and SINE2-1_PPo*

*HaSE1* and *HaSE2* were reported from a lepidopteran insect *Helicoverpa armigera* by Wang et al. (Wang et al. 2012). *HaSE2_DP* is a *HaSE2*-related SINE from another lepidopteran insect, the monarch butterfly *Danaus plexippus.* The 5′ ∼130-bp sequences of *HaSE1* and *HaSE2_DP* are 78% identical, and this region corresponds to the 5′ tRNA-derived head and "conserved central domain" reported by Wang et al. (Wang et al. 2012). *SINE2-4_NV* from sea anemone exhibits similarity to both 5′ regions of *HaSE1* and *HaSE2_DP*. Furthermore, *HaSE2_DP* exhibits sequence similarity to two butterfly SINEs (*SINE2-1_PXu* and *SINE2-1_PPo*) with the exception of the 5′ half of the tRNA-derived region. The alignment of these SINEs with *SINE2-5_NV*, which is also similar to *SINE2-4_NV*, reveals the strong similarity among *HaSE2_DP*, *SINE2-1_PXu*, and *SINE2-1_PPo* starting around nucleotide 130 of *HaSE2_DP* (data not shown). In contrast, the 3′ region of *HaSE1* is similar to *SINE2-5_NV*. However, the "conserved central domain" does not exhibit strong conservation among these six SINE families.

The sequence 255–311 of *HaSE1* exhibits similarity with the 5′-UTR of the autonomous *RTE*-type LINE from the monarch, *RTE-2_DPl* (table 1). The 3′ end of *HaSE1* was reported to be similar to the 3′ end of *RTE-3_BM* from the domestic silkworm *Bombyx mori* (Wang et al. 2012). A Censor search with Repbase yields a more similar sequence in *RTE-N1_ATr* from a plant *Amborella trichopoda*, but the sequence similarity is restricted to the ∼40-bp 3′ end (supplementary fig. S1, Supplementary Material online).

The 3′ regions of *HaSE2_DP* exhibit similarity to *RTE-N2_Lch* from coelacanths (table 1). *RTE-N2_Lch* is an internally deleted derivative of *RTE-4_Lch*. *RTE-N2_Lch* corresponds to the 5′ 214 bp and the 3′ 70 bp of *RTE-4_Lch*. Therefore, the similarity of *HaSE2_DP* to the region 151–186 of *RTE-N2_Lch* indicates that *HaSE2_DP* contains a sequence originating from the

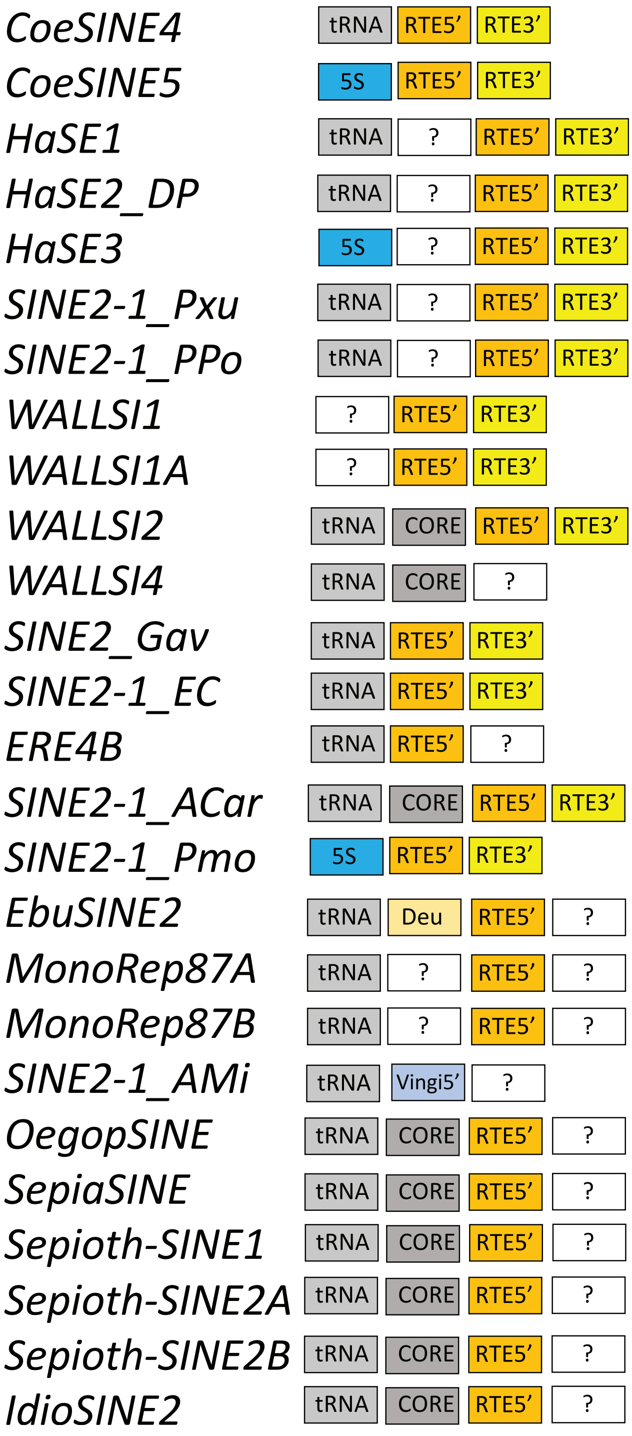

CoeSINE4 tRNA RTE5′ RTE3′
CoeSINE5 5S RTE5′ RTE3′
HaSE1 tRNA ? RTE5′ RTE3′
HaSE2_DP tRNA ? RTE5′ RTE3′
HaSE3 5S ? RTE5′ RTE3′
SINE2-1_Pxu tRNA ? RTE5′ RTE3′
SINE2-1_PPo tRNA ? RTE5′ RTE3′
WALLSI1 ? RTE5′ RTE3′
WALLSI1A ? RTE5′ RTE3′
WALLSI2 tRNA CORE RTE5′ RTE3′
WALLSI4 tRNA CORE ?
SINE2_Gav tRNA RTE5′ RTE3′
SINE2-1_EC tRNA RTE5′ RTE3′
ERE4B tRNA RTE5′ ?
SINE2-1_ACar tRNA CORE RTE5′ RTE3′
SINE2-1_Pmo 5S RTE5′ RTE3′
EbuSINE2 tRNA Deu RTE5′ ?
MonoRep87A tRNA ? RTE5′ ?
MonoRep87B tRNA ? RTE5′ ?
SINE2-1_AMi tRNA Vingi5′ ?
OegopSINE tRNA CORE RTE5′ ?
SepiaSINE tRNA CORE RTE5′ ?
Sepioth-SINE1 tRNA CORE RTE5′ ?
Sepioth-SINE2A tRNA CORE RTE5′ ?
Sepioth-SINE2B tRNA CORE RTE5′ ?
IdioSINE2 tRNA CORE RTE5′ ?

**Fig. 1.**—Schematic representation of SINE structures. The origins of head (tRNA or 5S rRNA), body (CORE), and LINE-derived parts (*RTE* 5′-UTR, *RTE* 3′-UTR, *Vingi* 5′-UTR) are indicated. Regions whose origins are unknown are indicated by "?."

**Table 1**

SINEs Whose Two Parts of Sequences Show Similarity to LINEs

| SINE | Region | LINE | Region | Identity |
|---|---|---|---|---|
| *CoeSINE4* (201) | 84–134 | *RTE-4_PPo* (3095) | 53–107 | 81% |
| | 141–194 | *RTE-4_PPo* | 3040–3093 | 76% |
| *CoeSINE5* (225) | 117–168 | *RTE-2_MMa* (3176) | 115–166 | 83% |
| | 173–216 | *RTE-2_LVa* (5082) | 5034–5077 | 76% |
| *HaSE1* (385) | 255–311 | *RTE-2_DPl* (3242) | 188–243 | 77% |
| | 341–385 | *RTE-N1_ATr* (195) | 144–186 | 91% |
| *HaSE2_DP* (299) | 194–229 | *RTE-N2_Lch* (286) | 151–186 | 83% |
| | 242–299 | *RTE-3_PXu* (565) | 504–558 | 84% |
| *HaSE3* (349) | 235–288 | *RTE-2_DPl* (3242) | 188–239 | 80% |
| *SINE2_Gav* (271) | 78–223 | *RTE-11_AMi* (3664) | 295–439 | 87% |
| | 223–267 | *RTE-11_AMi* | 3616–3659 | 89% |
| *SINE2-1_EC* (407) | 85–302 | *RTE-1_OAf* (3275) | 1–221 | 88% |
| | 317–406 | *RTE-1_OAf* | 3182–3269 | 83% |
| *SINE2-1_PPo* | 199–277 | *RTE-N1_Lch* (262) | 152–233 | 78% |
| *SINE2-1_PXu* | 197–296 | *RTE-N1_Lch* (262) | 152–250 | 73% |
| *SINE2-2_ACar* (239) | 10–220 | *MAR1*[a] (250) | 6–246 | 69% |
| *SINE-1_Pmo* (241) | 16–107 | *5S-Sauria*[a] (348) | 29–119 | 87% |
| | 134–239 | *BOVA2*[a] (269) | 1–122 | 74% |
| *WALLSI1* (387) | 170–335 | *RTE-1_PSi* (3769) | 27–196 | 69% |
| | 338–378 | *RTE-1_PSi* | 3726–3767 | 86% |
| *WALLSI1A* (610) | 415–564 | *RTE-3_AMi* (3899) | 147–292 | 74% |
| | 570–609 | *RTE-3_AMi* | 3854–3893 | 83% |
| *WALLSI2* (321) | 134–266 | *RTE-3_AMi* (3899) | 159–289 | 74% |
| | 275–317 | *RTE-3_AMi* | 3854–3895 | 88% |

NOTE.—If the same region of SINE hits several different LINEs, only the LINE with the highest CENSOR score is shown. The length of LINE/SINE is shown in parenthesis.

[a] SINEs originated by the internal deletion of LINEs.

5′-UTR of *RTE*. It is noteworthy that the 3′ region of *HaSE2* has been reported to be from a *Mariner* DNA transposon (Wang et al. 2012). However, the presence of a sequence similar to *RTE* indicates that *HaSE2* is also a canonical SINE whose 3′ region originates from a LINE. *SINE2-1_PXu* and *SINE2-1_PPo* also contain the sequence of bipartite *RTE* (table 1).

*HaSE3* and *HaSE1* share 3′ sequences but are different in their 5′ regions. Instead of a tRNA-derived head and conserved central domain of *HaSE1*, *HaSE3* has a 5S rRNA-derived head. The 3′ end of *HaSE3* is similar to that of *HaSE1*, and they share a common origin of the 3′ end of *RTE* (table 1).

### *WALLSI*

Five *WALLSI* subfamilies (*WALLSI1*, *WALLSI1A*, *WALLSI2*, *WALLSI3*, and *WALLSI4*) have been reported from the tammar wallaby *Macropus eugenii. WALLSI* subfamilies other than *WALLSI2* have also been found in the Tasmanian devil (Nilsson et al. 2012). The 3′ half of *WALLSI1* is similar to that of *MAR4_MD*, a bipartite nonautonomous *RTE* from the opossum *Monodelphis domestica*. *WALLSI1*, *WALLSI1A*, *WALLSI2*, and *WALLSI3* share very similar 3′ halves that exhibit strong similarity to the 5′- and 3′-UTRs of *RTE* (table 1). *WALLSI3* has been revealed to be a bipartite nonautonomous *RTE* and is very similar to *RTESINE2*, an older bipartite nonautonomous *RTE* family, which is also found in the genome of the opossum *M. domestica* (Nilsson et al. 2010). The 5′ ∼130 bp of *WALLSI2* is similar to the corresponding regions of the *MIR* and *THER1* families. Therefore, *WALLSI2* is composed of a tRNA-derived head (roughly 1–80), CORE (roughly 80–133), 5′ part of *RTE* (134–266), and 3′ end of *RTE* (275–317) (supplementary fig. S1, Supplementary Material online). The 5′ regions of *WALLSI1* and *WALLSI1A* do not exhibit any similarities with other transposable elements (TEs), tRNAs or 5S rRNA. *WALLSI4* does not exhibit sequence similarity with any other *WALLSI* SINEs in its 3′ region, but its 5′ region is similar to that of *WALLSI2*. This finding suggests that *WALLSI2* was generated by the fusion of a 5′ region of *WALLSI4* and a 3′ region of *WALLSI1*, *WALLSI1A*, or *WALLSI3*. *RTESINE2* and *WALLSI3* are very similar, and *RTESINE2* is older than any *WALLSI* subfamilies, which indicates that *WALLSI3* is the direct descendant of *RTESINE2* in the wallaby lineage and that *WALLSI1* and *WALLSI1A* are the derivatives of *WALLSI3* with swapped 5′ regions.

### *SINE2_Gav*

*SINE2_Gav* from crocodilians is 271 bp in length. It is composed of a $\text{tRNA}^{\text{Gly}}$-like head (roughly 1–70), a middle sequence (78–223) similar to the 5′-UTR of *RTE-11_AMi* and a tail (223–267) similar to the 3′-UTR of *RTE-11_AMi* (supplementary fig. S1, Supplementary Material online).

### *SINE2-2_ACar*

*SINE2-2_ACar* is 239 bp in length. It is composed of the 5′ tRNA-derived head, a CORE-like middle sequence and two regions derived from the 5′- and the 3′-UTRs of an *RTE*-type LINE (supplementary fig. S1, Supplementary Material online). The downstream sequence from the CORE shows no similarity to known LINEs, but it is similar to *RTE*-derived regions of some SINEs including *AFROSINE3* and *MAR1*. Many *SINE2-2_ACar* copies are roughly 85% identical to the consensus. The structure of *SINE2-2_ACar* is identical to that of *MAR1*, and it therefore may be a distant relative of *MAR1*.

### *SINE-1_Pmo*

*SINE-1_Pmo* is a SINE3 family from the python *Python molurus*. Although the 3′ end (188–241) of *SINE-1_Pmo* has no closely related LINEs, it exhibits similarity with *BovA2* (table 1). A comparison between *SINE-1_Pmo* and *BovB* (a family of *RTE* and the counterpart LINE of *BovA2*) revealed that *SINE-1_Pmo* includes the sequences originating from the 5′- and 3′-UTRs of *RTE*.

### *SINE2-1_EC* and Its Descendants

*SINE2-1_EC*, which originated from the horse *Equus caballus*, is 407 bp in length and has a 3′ region (85–406) exhibiting >80% sequence identity to the 5′- and 3′-UTRs of *RTE-1_OAf* from the aardvark *Orycteropus afer* (table 1). Therefore, the structure of *SINE2-1_EC* resembles that of *AfroSINEs* even though horses are not Afrotherians. Upstream of this sequence (6–77) is a tRNA$^{Glu}$-derived head based on the result of tRNAscan-SE (http://lowelab.ucsc.edu/cgi-bin/tRNAscan-SE2.cgi) and a Censor search in Repbase. Interestingly, the 5′ 115-bp sequence of *SINE2-1_EC* is almost identical to that of *ERE4B*, another SINE from the horse. As a consequence, the downstream sequence of the tRNA-derived head of *ERE4B* exhibited a pronounced similarity to the 5′ ends of *RTE*. The entire length of *ERE4B* is similar to *ERE4*, *ERE1*, *ERE1B*, and *ERE1C*—all from the horse—as well as *CERE1* from the white rhinoceros *Ceratotherium simum*. These SINEs may have a chimeric origin between a *SINE2-1_EC*-like sequence contributing to the 5′ half and another LINE or SINE contributing to the 3′ half. There are no clues in terms of the counterpart LINE for *ERE1*, *ERE1B*, *ERE1C*, *ERE4*, or *ERE4B*.

### Solo LINE-Derived Sequences in the Middle of SINEs

Ancient bipartite LINE-derived sequences may have been exchanged by newly acquired 3′ tails derived from another LINE. This situation can lead to a structure in which only the middle part of the SINE exhibits a similarity with the LINE. *ERE4B* is an example of such a chimeric SINE. Manual inspection of the Censor results noted above revealed several candidates for this type of chimeric SINE (fig. 1 and table 2). The alignments between SINEs and LINEs appear in supplementary figure S3, Supplementary Material online. Among them, *ERE4B* is described earlier. The 3′ end of *MARE3* corresponds to the middle of 5′-UTR of *RTE-14_Lch*, suggesting the current consensus sequence of *MARE3* is 3′-truncated.

### *EbuSINE2*

*EbuSINE2* has been reported to be a family of Deu-SINEs with a tRNA-derived head (Nishihara et al. 2006). The sequence downstream of the Deu-domain (278–321) exhibits similarity with the 5′-UTR (129–176) of *RTE-3_MD* (table 2 and supplementary fig. S1, Supplementary Material online). Although the 3′ terminus of *EbuSINE2* exhibits no sequence similarity with any TEs in Repbase, this region may be derived from the 3′-UTR of an unknown *RTE*.

### *MonoRep87A* and *MonoRep87B*

*MonoRep87A* and *MonoRep87B* are two SINE families from the platypus *Ornithorhynchus anatinus*. Their consensus sequences start with a tRNA-like sequence and end with (CAT)$_n$ microsatellites, indicating that they are full-length sequences of SINE2. Although there is no sequence similarity to known LINEs or SINEs in their 3′ termini, the upstream sequences exhibit similarity with the 5′-UTR of *RTE* (table 2 and supplementary fig. S1, Supplementary Material online). The middle regions of these two SINE2 families are similar, but no close relatives have been found.

**Table 2**
Internal Fragments of LINE 5′-UTRs Seen in the Middle of SINEs

| SINE | Region | LINE | Region | Identity |
|---|---|---|---|---|
| *ERE4B* (185) | 82–116 | *RTE1-N1b_LA* (470) | 2–37 | 92% |
| *EbuSINE2* (370) | 270–321 | *RTE-3_MD* (3228) | 129–176 | 86% |
| *MARE3* (180) | 101–178 | *RTE-14_Lch* (3944) | 220–298 | 69% |
| *MonoRep87A* (523) | 391–455 | *RTE-3_PM* (3975) | 294–359 | 76% |
| *MonoRep87B* (537) | 403–467 | *RTE-14_Lch* (3944) | 286–348 | 78% |
| *SINE2-1_AMi* (161) | 61–127 | *Vingi-2_Gav* (3128) | 2–74 | 84% |
| *IdioSINE2* (423) | 130–367 | *RTE-2_Croc* (4296) | 259–486 | 75% |
| *OegopSINE* (370) | 130–220 | *RTE-3_BF* (4202) | 325–414 | 79% |
| | 225–281 | *RTE-12_AMi* (3904) | 182–237 | 74% |
| *SepiaSINE* (278) | 127–213 | *RTE-3_BF* (4202) | 325–414 | 76% |
| *Sepioth-SINE1* (292) | 134–239 | *RTE-3_BF* (4202) | 325–423 | 79% |
| *Sepioth-SINE2A* (294) | 133–238 | *RTE-3_BF* (4202) | 325–423 | 77% |

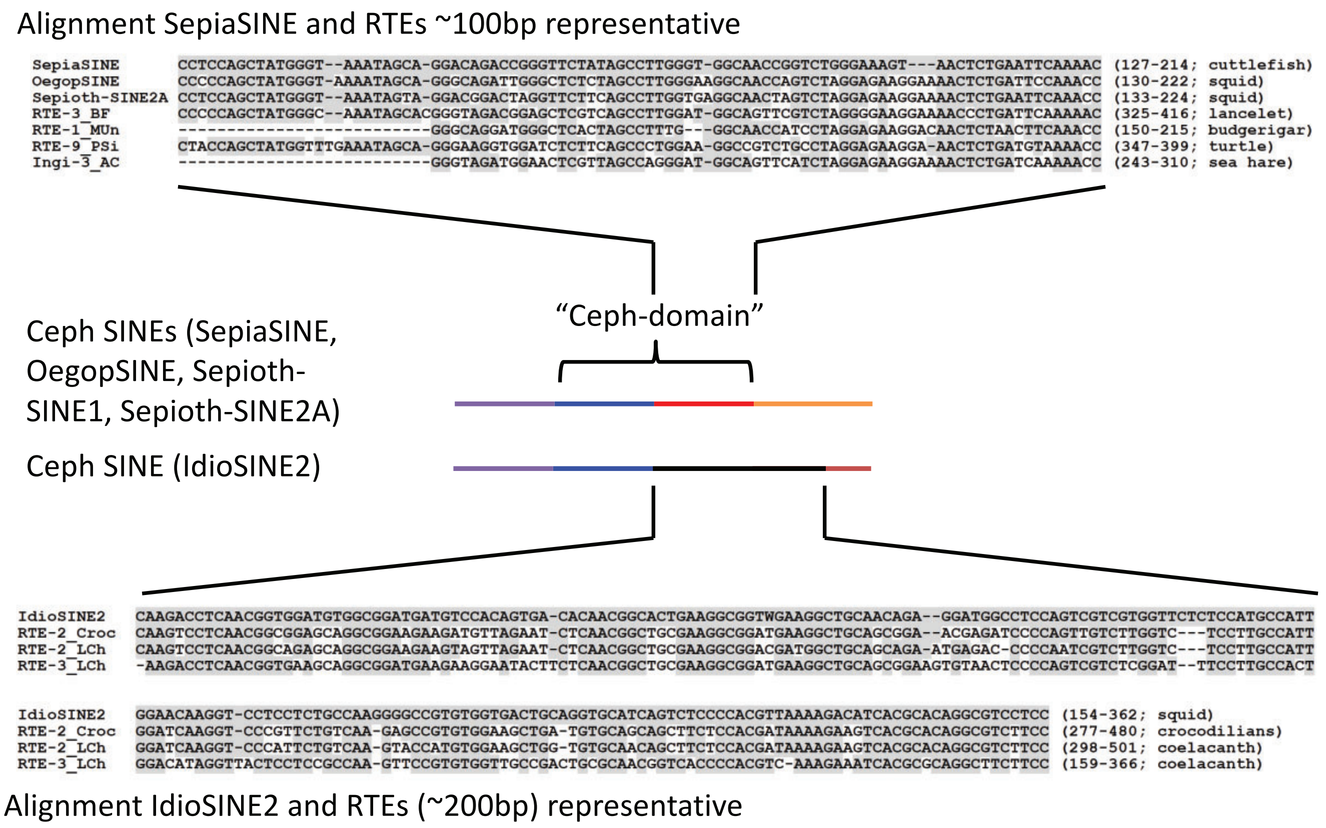
Note.—If the same region of SINE hits several different LINEs, only the LINE with the highest CENSOR score is shown. The length of LINE/SINE is shown in parenthesis.

### *SINE2-1_AMi*

*SINE2-1_AMi* is a SINE2 family found from *Alligator mississippiensis*. Just downstream of the 5′ tRNA-derived head, there is a sequence similar to the 5′ end of *Vingi-2_Gav* from the gharial *Gavialis gangeticus* (table 2 and supplementary fig. S1, Supplementary Material online). This example is the only newly identified SINE containing a fragment of a LINE other than *RTE*.

### The 3′ Half of the Ceph-Domain Originates from RTE 5′-UTR

The 3′ halves of the "Ceph-domain" of *SepiaSINE*, *OegopSINE*, *Sepioth-SINE1*, and *Sepioth-SINE2A* are similar to several LINEs belonging to the *RTE* clade (table 2 and supplementary fig. S1, Supplementary Material online). A repeated Censor search with these 3′ halves of Ceph-domains and *RTE* sequences in Repbase revealed a well-conserved domain between *RTE*, *RTE*-derived SINEs as well as *Ingi-3_AC* and *R4-1_ADi* (fig. 2 and supplementary fig. S4, Supplementary Material online). *RTEs* from diverse animals—including vertebrates, echinoderms, annelids, arthropods, and cnidarians— contain this conserved domain. The hits included recently characterized LINEs and SINEs from


Alignment SepiaSINE and RTEs ~100bp representative
"Ceph-domain"
Ceph SINEs (SepiaSINE, OegopSINE, Sepioth-SINE1, Sepioth-SINE2A)
Ceph SINE (IdioSINE2)
Alignment IdioSINE2 and RTEs (~200bp) representative

```
SepiaSINE       CCTCCAGCTATGGGT--AAATAGCA-GGACAGACCGGGTTCTATAGCCTTGGGT-GGCAACCGGTCTGGGAAAGT---AACTCTGAATTCAAAAC (127-214; cuttlefish)
OegopSINE       CCCCCAGCTATGGGT-AAAATAGCA-GGGCAGATTGGGCTCTCTAGCCTTGGGAAGGCAACCAGTCTAGGAGAAGGAAAACTCTGATTCCAAACC (130-222; squid)
Sepioth-SINE2A  CCTCCAGCTATGGGT--AAATAGTA-GGACGGACTAGGTTCTTCAGCCTTGGTGAGGCAACTAGTCTAGGAGAAGGAAAACTCTGAATTCAAACC (133-224; squid)
RTE-3_BF        CCCCCAGCTATGGGC--AAATAGCACGGGTAGACGGAGCTCGTCAGCCTTGGAT-GGCAGTTCGTCTAGGGGAAGGAAAACCCTGATTCAAAAC (325-416; lancelet)
RTE-1_MUn       -------------------------GGGCAGGATGGGCTCACTAGCCTTTG---GGCAACCATCCTAGGAGAAGGACAACTCTAACTTCAAACC (150-215; budgerigar)
RTE-9_PSi       CTACCAGCTATGGTTTGAAATAGCA-GGGAAGGTGGATCTCTTCAGCCCTGGAA-GGCCGTCTGCCTAGGAGAAGGA-AACTCTGATGTAAAACC (347-399; turtle)
Ingi-3_AC       -------------------------GGGTAGATGGAACTCGTTAGCCAGGGAT-GGCAGTTCATCTAGGAGAAGGAAAACTCTGATCAAAAACC (243-310; sea hare)
```

```
IdioSINE2   CAAGACCTCAACGGTGGATGTGGCGGATGATGTCCACAGTGA-CACAACGGCACTGAAGGCGGTWGAAGGCTGCAACAGA--GGATGGCCTCCAGTCGTCGTGGTTCTCTCCATGCCATT
RTE-2_Croc  CAAGTCCTCAACGGCGGAGCAGGCGGAAGAAGATGTTAGAAT-CTCAACGGCTGCGAAGGCGGATGAAGGCTGCAGCGGA--ACGAGATCCCCAGTTGTCTTGGTC---TCCTTGCCATT
RTE-2_LCh   CAAGTCCTCAACGGCAGAGCAGGCGGAAGAAGTAGTTAGAAT-CTCAACGGCTGCGAAGGCGGACGATGGCTGCAGCAGA-ATGAGAC-CCCCAATCGTCTTGGTC---TCCTTGCCATT
RTE-3_LCh   -AAGACCTCAACGGTGAAGCAGGCGGATGAAGAAGGAATACTTCTCAACGGCTGCGAAGGCGGATGAAGGCTGCAGCGGAAGTGTAACTCCCCAGTCGTCTCGGAT--TTCCTTGCCACT

IdioSINE2   GGAACAAGGT-CCTCCTCTGCCAAGGGGCCGTGTGGTGACTGCAGGTGCATCAGTCTCCCCACGTTAAAAGACATCACGCACAGGCGTCCTCC (154-362; squid)
RTE-2_Croc  GGATCAAGGT-CCCGTTCTGTCAA-GAGCCGTGTGGAAGCTGA-TGTGCAGCAGCTTCTCCACGATAAAAGAAGTCACGCACAGGCGTCTTCC (277-480; crocodilians)
RTE-2_LCh   GGATCAAGGT-CCCATTCTGTCAA-GTACCATGTGGAAGCTGG-TGTGCAACAGCTTCTCCACGATAAAAGAAGTCACGCACAGGCGTCTTCC (298-501; coelacanth)
RTE-3_LCh   GGACATAGGTTACTCCTCCGCCAA-GTTCCGTGTGGTTGCCGACTGCGCAACGGTCACCCCACGTC-AAAGAAATCACGCGCAGGCTTCTTCC (159-366; coelacanth)
```

**Fig. 2.**—Sequence similarity of Ceph-domains with some *RTE* LINEs. Nucleotides identical to those in representative Ceph-domains (*SepiaSINE* and *IdioSINE2*) are shaded. The positions of consensus sequences are shown in parentheses with their origins.

birds (Suh et al. 2016). *BuceSINE*, *MeloSINE*, *ManaSINE1*, and *ManaSINE2* are assumed to have originated independently, but they all contain sequences showing similarity with the 3′ half of the Ceph-domain. Some *RTEs*, such as *AviRTE*, contain two regions corresponding to this conserved domain in their 5′-UTRs. It is noteworthy that this conserved domain is not located at the 5′ end but rather in the middle of the 5′-UTRs.

An unexpected finding was that *Ingi-3_AC* and *R4-1_ADi*, very distant LINEs from *RTE*, contained a sequence similar to Ceph-SINE and *RTE*. The *RTE*-like sequence in *Ingi-3_AC* (243–310) is in the latter half of 5′-UTR (1–450). Because *Ingi-3_AC* is a LINE from the California sea hare *Aplysia californica*, one species of mollusks, it is possible that the recombination between the *Ingi* LINE and the Ceph-SINE contributed to this sequence similarity. *R4-1_ADi* is from coral *Acropora digitifera*. The sequence similar to *RTE* is located at 135–179, in the former half of 5′-UTR (1–652).

### Another Ceph-SINE IdioSINE2 Has a Different 5′-UTR Fragment of RTE

The 3′ half of the Ceph-domain of *IdioSINE2* is similar to vertebrate *RTE* families such as *RTE-2_Croc* from crocodilians and *RTE-2_LCh* from coelacanths (table 1 and fig. 2). This *RTE*-like sequence is not similar to *RTE*-like sequences from other Ceph-SINEs. However, upstream of this region, *IdioSINE2* contains a short, 23-bp *RTE*-like sequence (CCTCCAGCTATGGGTTAAATAGT) similar to that of other Ceph-SINEs. It corresponds to the 5′ terminal sequence of the *RTE*-like sequence from other Ceph-SINEs. Considering the occasional replacement of LINE-like 3′ terminal sequences in SINE evolution, *IdioSINE2* was likely generated via tail replacement by another *RTE* LINE with a short 23-bp fragment of original RTE-like sequence remaining.

### Similarity between the CORE-Domain and the 5′ Half of the Ceph-Domain

It is now clear that the 3′ half of the Ceph-domain derives from the 5′-UTR of *RTE*. What about the 5′ half of the Ceph-domain? The originally reported Ceph-domain was ∼150 bp long. Excluding the *RTE*-derived region, the 5′ ∼50-bp sequence is here redetermined as Ceph-domain. A Censor search in Repbase revealed that this 5′ half exhibits weak similarity with the CORE-domain (fig. 3). The CORE-domain exhibits a high sequence diversity, and the conserved region among all reported CORE-domains is only ∼25-bp long. The Ceph-domain shares 15 bp with the conserved CORE-

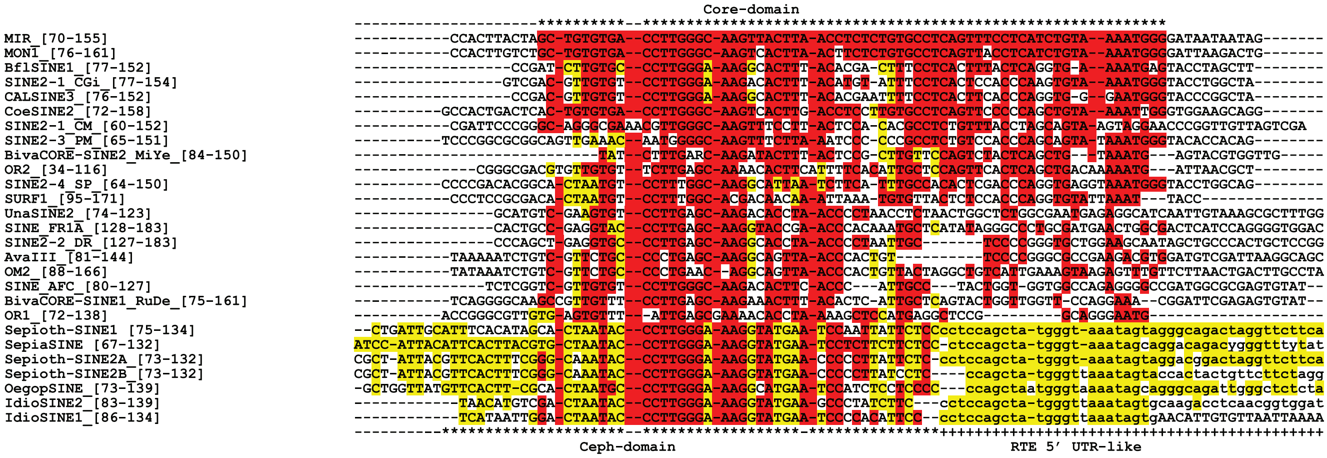


**Fig. 3.**—Alignment of CORE-domains and Ceph-domains. Nucleotides identical to *MIR* CORE-domain is colored in red, whereas nucleotides identical to *Sepioth-SINE1* but not to *MIR* CORE-domain are in yellow. *RTE*-like sequences in Ceph-SINEs are in lower cases.

domain. The conserved sequence CCTTGGG in the Ceph-domain is also present in the CORE-domain. Two CORE-SINEs from mollusks, *SINE2-1_CGi* from the Pacific oyster *Crassostrea gigas* and *CALSINE3* from the California sea hare *Aplysia californica*, share a longer identical sequence with the Ceph-domain CCTTGGGAAAG. It is reasonable to consider that the Ceph-domain is a cephalopod- (or mollusk)-specific derivative of the CORE-domain that has experienced the loss of the 3′ half of the CORE-domain due to tail replacement by *RTE*.

## Discussion

### Bipartite Nonautonomous LINEs and the Birth of New SINEs

In this study, several new SINE families that have the 5′ and the 3′ parts of LINEs were found. Not a few nonautonomous LINEs with solely the 5′ and the 3′ parts of autonomous LINE counterparts have been created by internal deletion (Bao et al. 2015). They can be subclassified into two types: ORF1-absent and ORF1-present. *HeT-A*, *HAL1*, and *Ag-Sponge* can be members of bipartite nonautonomous LINEs even though they encode one protein corresponding to ORF1p (Pardue et al. 1996; Smit 1999; Biedler and Tu 2003; Bao and Jurka 2010). *HeT-A* and related elements were derived from the *Jockey* clade of LINEs, *HAL1* were from the *L1* clade, and *Ag-Sponge* were from the *CR1* clade. The proteins encoded by these nonautonomous elements likely function to multimerize with the proteins encoded by autonomous counterparts and to enhance transposition (Rashkova et al. 2002). The necessity of generating ORF1p excludes the possibility that these protein-coding nonautonomous LINEs function as a source of SINEs; SINEs cannot encode a protein.

The distributions of bipartite ORF-absent nonautonomous LINEs in the classification of LINEs are very biased. Only four clades of LINEs—*RTE*, *Ingi*, *Vingi*, and *R2*—have been reported to produce bipartite nonautonomous LINEs (Bringaud et al. 2003, 2009; Kojima et al. 2011; Eickbush and Eickbush 2012; Bao et al. 2015). *Ingi* and *Vingi* are closely related clades of LINEs (Kojima et al. 2011). Here, only two clades of LINEs, *RTE* and *Vingi*, were revealed to contribute to the middle parts of SINEs. It is obvious that some SINEs are descendants of bipartite nonautonomous LINEs as proposed previously for *Bov-tA* (Okada and Hamada 1997). A striking example is *WALLSI2*. *WALLSI2* is the recombinant between *WALLSI4* and either *WALLSI1*, *WALLSI1A*, or *WALLSI3*. *WALLSI3* is a bipartite nonautonomous LINE, and *WALLSI1* and *WALLSI1A* are likely descendants of *WALLSI3* or *RTESINE2*, the latter of which has an identical structure as *WALLSI3* but is older. *WALLSI2*, as well as *SINE2-1_ACar*, has a CORE-domain upstream of the 5′ part of *RTE*. It is very likely that *SINE2-1_ACar* is also a recombinant of a bipartite nonautonomous LINE and an unknown SINE having a tRNA-derived head and a CORE-domain.

The 5′ sequences of *RTE* observed in SINEs are not always the 5′ ends. In contrast, bipartite nonautonomous LINEs usually possess the 5′ end of their original LINEs. The presence of a self-cleaving ribozyme at the 5′ terminus of some LINEs may be a cause of this distinction (Ruminski et al. 2011). Several *RTE* families are predicted to possess a self-cleaving ribozyme (Ruminski et al. 2011). Considering the structure of SINEs, which possess an RNA-derived head upstream of their LINE-derived parts, the presence of a self-cleaving ribozyme causes 5′-truncation. Six clades of LINEs, *R1*, *R2*, *R4*, *RTE*, *Ingi*, and *LOA*, were revealed to possess a self-cleaving ribozyme at their 5′ ends (Eickbush and Eickbush 2010; Ruminski et al. 2011; Sánchez-Luque et al. 2011). Among them, *R1*, *R2*, and *R4* are target-specific LINEs (Kojima and Fujiwara 2003, 2004) and likely depend on the transcription of target ribosomal RNA genes. They accordingly need to cleave their 5′ ends to generate full-length transcripts (Eickbush and Eickbush 2003, 2010). Three clades, *RTE*, *Ingi*, and *R2*, generate bipartite

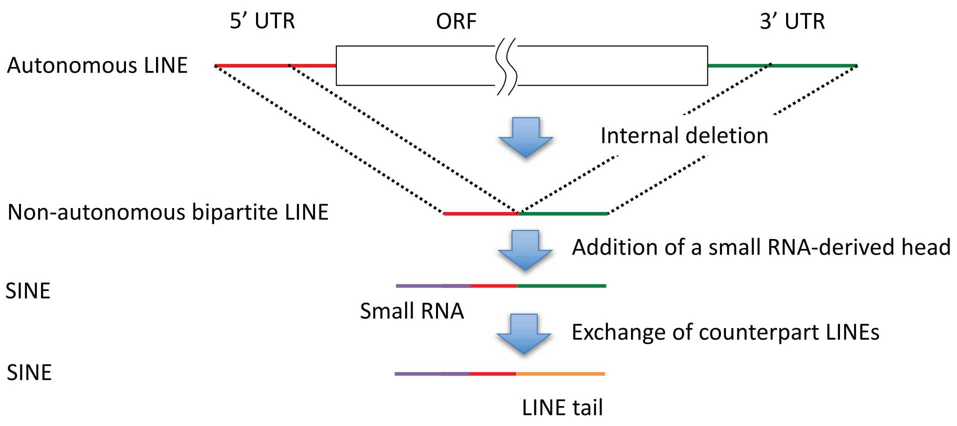

5′ UTR
ORF
3′ UTR
Autonomous LINE
Internal deletion
Non-autonomous bipartite LINE
Addition of a small RNA-derived head
SINE
Small RNA
Exchange of counterpart LINEs
SINE
LINE tail

**Fig. 4.**—Hypothetical origin of SINE body.

nonautonomous LINEs. It is not yet known whether this tendency is caused by sampling bias or by specific requirements of transcription.

The generation of bipartite nonautonomous LINEs may also be related to different requirement of transcription initiation. Bipartite LINEs are very likely transcribed by RNA polymerase II, as is true for their counterpart autonomous LINEs. SINEs, on the other hand, are transcribed by RNA polymerase III. It is known that the 5′ extreme regions of LINEs are responsible for transcription (Takahashi and Fujiwara 1999). The cis-regulatory sequences for transcription by RNA polymerase II may contradict efficient transcription by RNA polymerase III.

## Origins of Conserved SINE Bodies

Currently, the V-domain, CORE-domain, Deu-domain, Nin-domain, Ceph-domain, Inv-domain, Pln-domain, Snail-domain, and Meta-domain have been proposed as conserved SINE bodies (Gilbert and Labuda 1999; Ogiwara et al. 2002; Nishihara et al. 2006, 2016; Akasaki et al. 2010; Piskurek and Jackson 2011; Luchetti and Mantovani 2013; Matetovici et al. 2016). However, Nin-domain and Inv-domain have been reported to be variants or parts of Deu-domain. The Snail-domain and the Nin-domain show similarity at their 5′ ends. In this article, the originally proposed Ceph-domain (Akasaki et al. 2010) is revealed to be composed of two regions of independent origins: the CORE-domain and the 5′-UTR of *RTE*. Although the sequence similarity between the CORE-domain and Ceph-domain is marginal (fig. 3), the sequence diversity among CORE-SINEs can rationalize the classification of the Ceph-domain as a member of the CORE-domain (Gilbert and Labuda 1999).

Recent analysis has revealed that some SINE "superfamilies" share 5′ regions of their bodies but not 3′ regions. Nishihara et al. (2016) reported that two different types of 3′ regions of the CORE-domain are present, and they designated them CORE (original) and CORE2. The Inv-domain is similar to the Nin-domain and is combined with the 3′ flanking Pln-domain in Polyneopteran insects (Luchetti and Mantovani 2013). The Nin-domain and Snail-domain exhibit sequence similarity only in their 5′ regions (Matetovici et al. 2016). The fusion of two bodies, such as the Meta-domain and the Deu-domain, is also observed (Nishihara et al. 2016). These facts suggest that these proposed domains are not minimal functional units. The replacement of parts of the body appears common.

Here, a hypothesis that nonautonomous LINEs that have only 5′ and 3′ regions of original LINEs can be a source of enigmatic middle body of SINEs is proposed (fig. 4). This can be considered as an extension of the hypothesis by Okada and Hamada (1997), in which some SINEs originated from the addition of 5′ heads onto an internally deleted derivative of autonomous LINEs. Very limited groups of LINEs can generate internally deleted derivatives for unknown reasons. Such non-autonomous bipartite LINEs can be transcribed by RNA polymerase II and transpose dependently on the original autonomous LINEs. A template switch can add a 5′ small RNA-derived sequence onto a bipartite LINE, resulting in the birth of a SINE that is transcribed by RNA polymerase III. Due to the occasional exchange of parts of SINEs, the 5′ and 3′ regions of LINEs cannot always be present in combination in SINEs, which is demonstrated by the structure of *ERE4B*. Once the 3′ LINE-derived sequence is exchanged, characterizing the origin of the middle bodies of SINEs is a challenge due to their short lengths and relatively low sequence conservation compared with the rapid sequence evolution of mobile elements. The LINE-originated sequence in *ERE4B* is only 35 nucleotides in length. It would be nearly impossible to characterize the origin of this kind of short fragmented sequence if the counterpart LINE went extinct. This situation is perhaps why no sequence similar to conserved body sequences of SINEs has been found.

SINEs which contain similar *RTE* 5′ regions, such as avian *BuceSINE*, *ManaSINE*, *MeloSINE*, and Ceph-SINEs, have independently evolved. A high sequence similarity of *RTE* 5′ regions between SINEs from diverse animals has been observed. For example, the *RTE* 5′ sequence from *CoeSINE4* from coelacanths is ∼87% identical to that of *SINE2-1_PPo* from butterflies. This high sequence similarity resembles conserved SINE bodies. Conserved SINE bodies are often observed in conserved noncoding elements (Nishihara et al. 2006; Xie et al. 2006). They have been exapted to have a certain biological function, such as enhancer, promoter, or insulator (Bejerano et al. 2006; Sasaki et al. 2008). The ability to bind to a transcriptional regulator can also be useful for SINEs and LINEs, and it can accordingly be speculated that the conservation of the 5′-UTR sequences among diverse *RTE* LINEs as well as SINEs is due to their functional importance in the lifecycle of these mobile elements. Such functional elements can be maintained in evolution and are poised to become integrated into host biological systems.

## Supplementary Material

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

## Acknowledgments

The author thanks Weidong Bao for a critical reading of the manuscript. This work was supported by the Ministry of Science and Technology of the Republic of China (106-2311-B-006-001-MY2, 106-2628-B-006-002-MY3).

## 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(1–2):8–19.

Bao W, Jurka J. 2010. Origin and evolution of LINE-1 derived "half-L1" retrotransposons (HAL1). Gene 465(1–2):9–16.

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

Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res 12(8):1269–1276.

Bejerano G, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441(7089):87–90.

Biedler J, Tu Z. 2003. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. Mol Biol Evol. 20(11):1811–1825.

Bringaud F, Berriman M, Hertz-Fowler C. 2009. Trypanosomatid genomes contain several subfamilies of ingi-related retroposons. Eukaryot Cell 8(10):1532–1542.

Bringaud F, et al. 2003. The ingi and RIME non-LTR retrotransposons are not randomly distributed in the genome of *Trypanosoma brucei*. Mol Biol Evol. 21(3):520–528.

Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet. 35(1):41–48.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Eickbush DG, Eickbush TH. 2003. Transcription of endogenous and exogenous R2 elements in the rRNA gene locus of *Drosophila melanogaster*. Mol Cell Biol. 23(11):3825–3836.

Eickbush DG, Eickbush TH. 2010. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. Mol Cell Biol. 30(13):3142–3150.

Eickbush DG, Eickbush TH. 2012. R2 and R2/R1 hybrid non-autonomous retrotransposons derived by internal deletions of full-length elements. Mob DNA 3(1):10.

Gilbert N, Labuda D. 1999. CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc Natl Acad Sci U S A. 96(6):2869–2874.

Gilbert N, Labuda D. 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. J Mol Biol. 298(3):365–377.

Gogolevsky KP, Vassetzky NS, Kramerov DA. 2008. Bov-B-mobilized SINEs in vertebrate genomes. Gene 407(1–2):75–85.

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

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

Kapitonov VV, Tempel S, Jurka J. 2009. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. Gene 448(2):207–213.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33(2):511–518.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7:474.

Kojima KK. 2015. A new class of SINEs with snRNA gene-derived heads. Genome Biol Evol. 7(6):1702–1712.

Kojima KK, Fujiwara H. 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol. 20(3):351–361.

Kojima KK, Fujiwara H. 2004. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol. 21(2):207–217.

Kojima KK, Kapitonov VV, Jurka J. 2011. Recent expansion of a new Ingi-related clade of Vingi non-LTR retrotransposons in hedgehogs. Mol Biol Evol. 28(1):17–20.

Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J. 2007. Evolutionary history of 7SL RNA-derived SINEs in supraprimates. Trends Genet. 23(4):158–161.

Longo MS, Brown JD, Zhang C, O'Neill MJ, O'Neill RJ. 2015. Identification of a recently active mammalian SINE derived from ribosomal RNA. Genome Biol Evol. 7(3):775–788.

Luchetti A, Mantovani B. 2013. Conserved domains and SINE diversity during animal evolution. Genomics 102(4):296–300.

Matetovici I, et al. 2016. Mobile element evolution playing Jigsaw – SINEs in gastropod and bivalve mollusks. Genome Biol Evol. 8(1):253–270.

Nikaido M, Nishihara H, Hukumoto Y, Okada N. 2003. Ancient SINEs from African endemic mammals. Mol Biol Evol. 20(4):522–527.

Nilsson MA, et al. 2010. Tracking marsupial evolution using archaic genomic retroposon insertions. PLoS Biol. 8(7):e1000436.

Nilsson MA, Janke A, Murchison EP, Ning Z, Hallström BM. 2012. Expansion of CORE-SINEs in the genome of the Tasmanian devil. BMC Genomics 13:172.

Nishihara H, Plazzi F, Passamonti M, Okada N. 2016. MetaSINEs: broad distribution of a Novel SINE superfamily in animals. Genome Biol Evol. 8(3):528–539.

Nishihara H, Smit AF, Okada N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. 16(7):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(2):316–324.

Ohshima K. 2012. Parallel relaxation of stringent RNA recognition in plant and mammalian L1 retrotransposons. Mol Biol Evol. 29(11):3255–3259.

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(7):3756–3764.

Okada N, Hamada M. 1997. The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. J Mol Evol. 44(S1):S52–S56.

Okada N, Hamada M, Ogiwara I, Ohshima K. 1997. SINEs and LINEs share common 3′ sequences: a review. Gene 205(1–2):229–243.

Okonechnikov K, Golosova O, Fursov M, team U. 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28(8):1166–1167.

Pardue ML, Danilevskaya ON, Lowenhaupt K, Wong J, Erby K. 1996. The gag coding region of the Drosophila telomeric retrotransposon, HeT-A, has an internal frame shift and a length polymorphic region. J Mol Evol. 43(6):572–583.

Piskurek O, Jackson DJ. 2011. Tracking the ancestry of a deeply conserved eumetazoan SINE domain. Mol Biol Evol. 28(10):2727–2730.

Rashkova S, Karam SE, Kellum R, Pardue ML. 2002. Gag proteins of the two Drosophila telomeric retrotransposons are targeted to chromosome ends. J Cell Biol. 159(3):397–402.

Ruminski DJ, Webb CH, Riccitelli NJ, Luptak A. 2011. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. J Biol Chem. 286(48):41286–41295.

Sánchez-Luque FJ, López MC, Macias F, Alonso C, Thomas MC. 2011. Identification of an hepatitis delta virus-like ribozyme at the mRNA 5′-end of the L1Tc retrotransposon from *Trypanosoma cruzi*. Nucleic Acids Res. 39(18):8065–8077.

Sasaki T, et al. 2008. Possible involvement of SINEs in mammalian-specific brain formation. Proc Natl Acad Sci U S A. 105(11):4220–4225.

Smit AF. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev. 9(6):657–663.

Suh A, et al. 2016. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. Nat Commun. 7:11396.

Takahashi H, Fujiwara H. 1999. Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm *Bombyx mori*. Nucleic Acids Res. 27(9):2015–2021.

Wang J, et al. 2012. Characterization of three novel SINE families with unusual features in *Helicoverpa armigera*. PLoS One 7(2):e31355.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191.

Xie X, Kamal M, Lander ES. 2006. A family of conserved noncoding elements derived from an ancient transposable element. Proc Natl Acad Sci U S A. 103(31):11659–11664.

**Associate editor**: Esther Betran