

## REVIEW

# Origin and evolution of SINEs in eukaryotic genomes

DA Kramerov and NS Vassetzky

Laboratory of Eukaryotic Genome Evolution, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russian Federation

Short interspersed elements (SINEs) are one of the two most prolific mobile genomic elements in most of the higher eukaryotes. Although their biology is still not thoroughly understood, unusual life cycle of these simple elements amplified as genomic parasites makes their evolution unique in many ways. In contrast to most genetic elements including other transposons, SINEs emerged *de novo* many times in evolution from available molecules (for example, tRNA). The involvement of reverse transcription

in their amplification cycle, huge number of genomic copies and modular structure allow variation mechanisms in SINEs uncommon or rare in other genetic elements (module exchange between SINE families, dimerization, and so on.). Overall, SINE evolution includes their emergence, progressive optimization and counteraction to the cell's defense against mobile genetic elements.

*Heredity* (2011) **107**, 487–495; doi:10.1038/hdy.2011.43; published online 15 June 2011

**Keywords:** repetitive elements; mobile elements; transposons; retrotransposons; SINEs; evolution

## Introduction

The profusion of eukaryotic genomes continues to amaze geneticists: as low as a few percents of eukaryotic genome length correspond to protein-coding sequences. Eukaryotic genes are commonly separated by long regions, and their coding sequences (exons) are intervened by non-coding ones (introns), which run to tens of kilobases. Extensive chromosomal regions free from genes, intergenic regions and introns contain great numbers of repetitive DNA sequences, most of which are mobile genetic elements or transposable elements (TEs). TEs are divided into two major classes: DNA transposons and retrotransposons. DNA transposons encode a transposase enzyme catalyzing the transposon DNA excision and its integration into a new genomic location ('cut and paste' mechanism). Similar to all other TEs, DNA transposons are transmitted vertically from parent to offspring; however, their horizontal transmission between species (sometimes phylogenetically distant) is not uncommon. Unlike other TEs, DNA transposons are found in both eukaryotes and prokaryotes (for review see Feschotte and Pritham, 2007).

Retrotransposons is the most abundant class of TEs. The transposition of all such elements involves the 'copy and paste' mechanism including transcription of the TE gene, reverse transcription of the RNA, and integration of the resulting DNA into a new genomic location. Long terminal repeat (LTR) elements represent the best-studied subclass of retrotransposons. They have a very wide distribution among eukaryotes, from yeast to human. Structurally, LTR elements resemble retroviral genomic copies. Both contain LTRs and open-reading frames encoding

the reverse transcriptase (RT) and the RNA-binding protein (Gag). Some LTR elements also have an open-reading frame encoding the envelope protein (Env). Essentially, such elements are endogenous retroviruses, which result from viral infections of germ cells. Apparently, LTR elements with the *env* gene can sometimes give rise to functional retroviruses. The amplification mechanism of LTR elements and retroviral copies in the host genome is the same and involves a tRNA molecule to prime the reverse transcription (for review see Havecker *et al.*, 2004).

Long INterspersed Elements (LINEs) is another subclass of retrotransposons. They have no LTRs but also encode the activity of RT and, commonly, RNase H and endonuclease as well as a *gag*-like protein. The mechanism of LINE amplification substantially differs from that of LTR elements. After the transcription and translation, the RT binds the LINE mRNA (most likely, the one that has been translated), and the complex is imported back to the nucleus to cleave one of genomic DNA strands using its endonuclease activity. The resulting 3' end of the genomic DNA serves as a primer for the reverse transcription of LINE RNA. During or after the synthesis, RT cleaves the other genomic DNA strand (usually, 8–16 nucleotides away from the first break), jumps to the resulting 3' end of the genomic DNA, and uses it as a primer for the synthesis of the second strand of LINE DNA and an extra fragment of the genomic DNA (target site duplication; Bibillo and Eickbush, 2004; Babushok *et al.*, 2006). In some (but not all) LINEs, the RNase H activity of RT is used to displace RNA from the duplex. Lastly, the gaps in DNA are filled by the cellular DNA repair system.

LINEs are widespread among eukaryotes, but are less common among unicellular ones. To date, dozens of LINE families falling into 17 clades have been described (Lovsin *et al.*, 2001; Eickbush and Malik, 2002; Bailey *et al.*, 2003). The horizontal transmission of LINEs is by far less common compared with DNA transposons and LTR elements; possibly, some LINE families are not horizontally transmitted at all.

Correspondence: Professor DA Kramerov, Laboratory of Eukaryotic Genome Evolution, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov St, Moscow 119991, Russian Federation. E-mail: kramerov@eimb.ru

Received 31 March 2011; revised 13 April 2011; accepted 15 April 2011; published online 15 June 2011

The last subclass of retrotransposons is Short INter-spersed Elements (SINES), whose length ranges from 100 to 600 bp (Kramerov and Vassetzky, 2005; Ohshima and Okada, 2005; Deragon and Zhang, 2006). The genomes can contain tens or hundreds of thousands of SINE copies. These copies are not identical and their sequence can vary by 5–35%. Altogether, these sequences constitute a SINE family. The genomes of a given species can contain several SINE families (usually, 2–4). In contrast to all other TEs transcribed by RNA polymerase II, SINES are transcribed by RNA polymerase III (pol III) and contain a pol III promoter in their sequence. SINES encode no proteins and have to use LINE RT for their retrotransposition (Jurka, 1997; Kajikawa and Okada, 2002; Dewannieux *et al.*, 2003). The transcribed SINE RNA binds to the LINE RT, which is followed by the reverse transcription and integration of a SINE copy into a new genomic location in a way described above for LINES. SINES are widespread among eukaryotes but not as wide as other TEs. Apparently, they can be found in all mammals, reptiles and fishes. SINES have been found in the genomes of some invertebrates including sea squirts, sea urchins, cephalopods and certain insects. SINES are also common in many flowering plants. At the same time, *Drosophila* species lack SINES, and SINES are missing in most unicellular eukaryotes. (Note that some genomes can contain short non-autonomous retrotransposons, largely fragments of LINES, that resemble SINES; however, they are not transcribed by pol III, and hence, cannot be classified as SINES.)

Essentially, SINES are genomic parasites and can cause damage to the host genome through insertional mutagenesis or unequal crossover. At the same time, SINE copies can be beneficial for the host as sources of promoters, enhancers, silencers, insulators, and even genes encoding RNAs and proteins; they can underlie alternative splicing and polyadenylation; finally, SINE RNAs can act as *trans* factors of transcription, translation and mRNA stability (Makalowski, 2000; Ponicsan *et al.*, 2010; Gong and Maquat, 2011).

This review addresses the origin of SINES and pathways of their evolution. After the introductory section, the problem is considered in two planes: the events in SINE evolution (sections Origin of SINE Families and Further Evolution of SINES) and the genetic mechanisms that make possible these events (Mechanisms of SINE Evolution). Finally, the problem is considered in a more general context to outline the peculiarities of SINE evolution and their coevolution with LINES and cells (Overview of SINE Evolution).

## SINE structure and classification

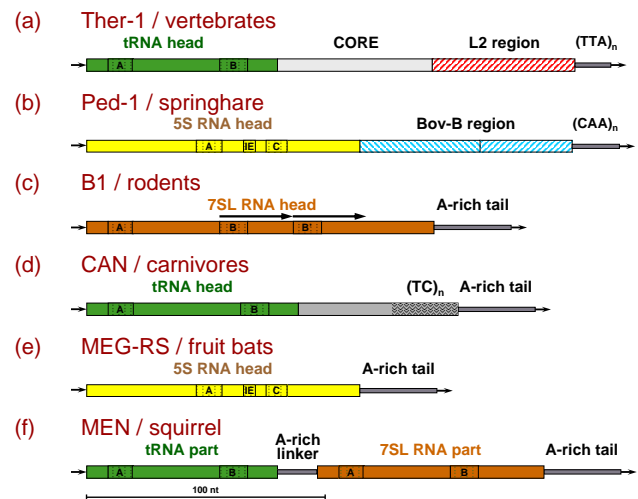
Most SINES consist of two or more modules: 5'-terminal 'head,' 'body' and 3'-terminal 'tail.' The head of all SINE families known to date demonstrate a clear similarity with one of the three types of RNA synthesized by pol III: tRNA, 7SL RNA, or 5S rRNA. The origin of SINE bodies is not easy to trace, although it has a region descending from one of the LINES in a large fraction of SINE families. The tail is a sequence of variable length consisting of simple (often degenerate) repeats.

The SINE head similarity with one of the cellular RNAs suggests its origin from this RNA. SINES originating from tRNAs are particularly abundant (Table 1;

**Table 1** Structural patterns of SINES

Monomeric SINES		(109)
w/o LINE-related region (83)		with LINE-related region (26)
tRNA- [??]- ~~~	(62)	tRNA-CORE-LINE ~~~
tRNA-CORE- [??]- ~~~	(10)	tRNA-LINE ~~~
tRNA-CORE- [??]- ~~~	(2)	tRNA-CORE- [??]- LINE ~~~
5S RNA- [??]- ~~~	(1)	tRNA- [??]- LINE ~~~
Simple SINES		(8)
tRNA ~~~	(6)	5S RNA-LINE ~~~
7SL RNA ~~~	(1)	
5S RNA ~~~	(1)	
Dimeric SINES		(15)
homomeric (8)		heteromeric (9)
tRNA-tRNA ~~~	(4)	tRNA-7SL RNA ~~~
tRNA- [??]- tRNA-CORE- [??]- ~~~	(1)	5S RNA-tRNA-CORE-LINE ~~~
tRNA- [??]- tRNA ~~~	(1)	7SL RNA-tRNA ~~~
7SL RNA-7SL RNA ~~~	(1)	tRNA- [??]- 7SL RNA ~~~
		tRNA-5S RNA ~~~
Trimeric SINES		(2)
tRNA-tRNA-tRNA ~~~	(1)	tRNA-7SL RNA-7SL RNA ~~~

tRNA, 7SL RNA, and 5S RNA are the heads derived from the corresponding RNAs; [??] corresponds to body parts of unknown origin; CORE denotes CORE and similar domains; LINE is the LINE-derived body region; and '~~~' denotes the tail. Numbers in parentheses indicate the number of SINE families with a given structure. (Note that the number of SINES with the LINE-derived region can be underestimated since the LINE partners have not been described yet.)



**Figure 1** SINE structure examples. (a) Ther-1 is a tRNA-derived CORE SINE of stringent recognition group (Gilbert and Labuda, 1999); (b) Ped-1, 5S rRNA-derived SINE of stringent recognition group (with bipartite LINE region; Gogolevsky *et al.*, 2008); (c) B1, 7SL RNA-derived quasi-dimeric SINE of relaxed recognition group (Labuda *et al.*, 1991); (d) CAN, tRNA-derived SINE of relaxed recognition group with a variable polypyrimidine region (Vassetzky and Kramerov, 2002); (e) MEG-RS, simple 5S rRNA-derived SINE of relaxed recognition group (Gogolevsky *et al.*, 2009); (f) MEN, dimeric tRNA/7SL RNA (heterodimeric) SINE of relaxed recognition group (Serdobova and Kramerov, 1998).

Figures 1a, d and f). A particular tRNA species of origin can be confidently identified for many SINE families, although nucleotide substitutions in SINE evolution make it impossible in other ones. To date, 7SL RNA-derived SINES (Figures 1c and f) have been

identified only in rodents (Krayev *et al.*, 1980; Veniaminova *et al.*, 2007), primates (Deininger *et al.*, 1981; Zietkiewicz *et al.*, 1998) and tree shrews (Nishihara *et al.*, 2002; Vassetzky *et al.*, 2003). The 7SL RNA (~300 nt) is found in all eukaryotes as the RNA component of the signal recognition particle (SRP), the ribonucleoprotein that targets secreted proteins to the endoplasmic reticulum. The number of SINE families originating from 5S rRNA is also not high (Table 1; Figures 1b and e); they have been found in some fishes (Kapitonov and Jurka, 2003; Nishihara *et al.*, 2006) and in a few mammals: fruit bats (Gogolevsky *et al.*, 2009) and springhare (Gogolevsky *et al.*, 2008).

The genes of all these RNAs (as well as the corresponding SINEs) have an internal pol III promoter. The promoter in tRNA and 7SL RNA genes consists of two boxes (A and B) of about 11 nt spaced by 30–35 nt, while the 5S rRNA genes have three such boxes: A, IE and C (Schramm and Hernandez, 2002). The presence of the promoter within the transcribed sequence is critical for SINE amplification, as the promoter is preserved in new SINE copies. By the head structure, SINEs are divided into three types according to the RNA of origin (tRNA-, 7SL- and 5S rRNA-derived; Figure 1).

The body of most SINE families (67%; Table 1) consists of a central sequence of unknown origin. The central sequence is specific for each SINE family; however, it can contain domains common for distant families (Table 1; Figure 1a). Currently, four such domains are known: CORE domain in vertebrates (Gilbert and Labuda, 1999), V-domain in fishes (Ogiwara *et al.*, 2002), Deu-domain in deuterostomes (Nishihara *et al.*, 2006) and Ceph-domain in cephalopods (Akasaki *et al.*, 2010). Some researchers recognize SINE superfamilies sharing CORE or similar domains.

A substantial fraction of SINEs (20%; Table 1; Figures 1a and b) has a 30–100 bp region of similarity with the 3'-terminal sequence of LINE, whose RT is involved in SINE amplification (Ohshima and Okada, 2005). Such regions are not only found in most of the SINEs in fishes (Matveev and Okada, 2009), but also occur in other groups including mammals. The LINE-derived regions of SINEs are required for the recognition of their RNA by the RT of some LINES, while RTs of other LINES require no specific recognition sequence. Accordingly, SINEs are divided into the stringent and relaxed recognition groups.

All SINEs have the 3'-terminal tail composed of repeated mono-, di-, tri-, tetra- or pentanucleotides. The tail of many SINEs is a poly(A) or irregular A-rich sequence (A-tail; Figures 1c–f), the amplification of all such SINEs in mammals depends on the RT of LINE1 (L1). In some SINEs, the end of A-rich tails can contain the signals of transcription termination and polyadenylation responsible for the synthesis of poly(A) at the 3' end of SINE RNA (Borodulina and Kramerov, 2001, 2008). By the presence of these signals, SINEs are divided into T<sup>+</sup> and T<sup>-</sup> classes. The tail synthesis in other SINEs is thought to be mediated by the template translocation mechanism similar to that in telomerase (Kajikawa and Okada, 2002; Roy-Engel *et al.*, 2005).

At the same time, not all SINEs have body (in particular, all known 7SL RNA-derived SINEs): 6% of SINE families consist of the head and tail only (Table 1). Such elements resembling pseudogenes of cellular RNAs are called simple SINEs (Figure 1e; Borodulina and

Kramerov, 2005). Simple SINEs can be distinguished from pseudogenes by specific nucleotide substitutions, which indicate their immediate origin from a SINE copy with such substitutions rather than from an RNA gene (Gogolevsky *et al.*, 2009).

On the other hand, the structure of SINEs can be more complex. Two or more SINEs can combine into a dimeric (or a more complex) structure, which is further amplified as a dimer (Table 1). Representatives of the same or different SINE families can combine. One of the first discovered SINEs, Alu in primates, consists of two similar parts derived from 7SL RNA (Deininger *et al.*, 1981; Ullu and Tschudi, 1984). There are dimeric and trimeric SINEs derived from tRNAs (Schmitz and Zischler, 2003; Churakov *et al.*, 2005). On the other part, complex elements composed of different SINE families or even types have been described. There are many such SINEs combining simple 7SL RNA- and tRNA-derived elements (Figure 1f); most of them were described in rodents (Serdobova and Kramerov, 1998; Veniaminova *et al.*, 2007; Churakov *et al.*, 2010), but they also exist in primates (Daniels and Deininger, 1983) and tree shrews (Nishihara *et al.*, 2002; Vassetzky *et al.*, 2003). Hybrid 5S rRNA/tRNA SINEs have been described (Nishihara *et al.*, 2006; Gogolevsky *et al.*, 2009), while no SINEs combining 7SL RNA- and 5S rRNA-derived elements are known yet. Accordingly, complex SINEs are divided into homodimers, heterodimers, trimers, and so on.

## Origin of SINE families

*The origin of a new SINE family is a multistage process. SINE amplification relies on at least two processes, transcription and reverse transcription/integration, and a SINE genomic copy should be efficiently transcribed, while its RNA should be efficiently reverse transcribed. SINEs originate from pseudogenes of tRNAs, 7SL RNA or 5S rRNA. The genomes of higher eukaryotes harbor numerous retropseudogenes of various small cellular RNAs. In mammals, most such pseudogenes have an A-rich tail, which indicates the involvement of L1 RT in their emergence, while similar retropseudogenes commonly have no A-rich tail in the genomes of non-mammalian higher eukaryotes.*

## Transcriptional competence

SINEs should be efficiently transcribed; moreover, their transcription should coincide with the period when active RT is available (LINE proteins are normally synthesized in the early embryogenesis). The majority of 7SL RNA pseudogenes are not transcribed, as the transcription of 7SL RNA genes depend on the regulatory elements upstream of the gene in addition to the internal promoter (Ullu and Weiner, 1985). Accordingly, a 7SL RNA pseudogene transformation into a SINE requires modifications that allow its transcription irrespective of the flanking sequences. It is possible that the deletion of the central region and/or smaller mutations in the 7SL RNA pseudogene in the genome of the common ancestor of primates and rodents have eventually led to the emergence Alu and B1.

Apparently, most tRNA pseudogenes with intact internal promoter can be transcribed, and their conversion into SINEs requires no such radical modifications (thus, SINEs emerged from tRNAs many times but, probably, only once from 7SL RNA). Nevertheless, the

transcriptional control had to be modified in this case as well—the transcriptional patterns of SINEs and tRNAs that gave rise to them substantially differ. As the *in vivo* transcription proceeds from a minor fraction of SINE copies (for example, Maraia, 1991), the flanking genomic sequences are nevertheless of importance: there seem to be additional regulatory signals modulating the transcriptional patterns of individual genomic copies of SINEs (Chesnokov and Schmid, 1996; Deininger *et al.*, 1996; Arnaud *et al.*, 2001).

#### Reverse transcriptional competence

Reverse transcription of foreign molecules (including cellular tRNAs) by LINE RT is an extremely rare event compared with the reverse transcription of LINE RNA. Currently, we know two systems protecting LINE RTs from processing foreign templates: sequence recognition of the RNA encoding the enzyme and *cis*-preference, when the RNA molecule used for RT translation is used by the translated enzyme as the template for reverse transcription (Esnault *et al.*, 2000; Wei *et al.*, 2001; Kajikawa and Okada, 2002). Overcoming this protection is an essential step in SINE formation. In the first case, it is realized by the acquisition of the fragment(s) recognized by the RT. The mechanism of *cis*-preference violation remains unclear; the SINE RNA interaction with the factors of the RT complex can be proposed. For instance, B1 and Alu (as well as 7SL) RNAs form a complex with SRP proteins SRP9/14 (Weichenrieder *et al.*, 2000), which can bind to polyribosomes. This way B1 and Alu transcripts can be presented to the synthesized L1 RT as the template for reverse transcription. A similar mechanism can be proposed for SINEs derived from tRNAs or 5S rRNA, components of the ribosomal complex. The *cis*-preference violation can be mediated by poly(A)-binding protein, which can bind proteins of the translational machinery (Roy-Engel *et al.*, 2002b); in this case, the acquisition of an A-tail should be an essential step in the evolution of SINEs mobilized by an RT with *cis*-preference. In some SINEs (for example, rodent B2), a polyadenylation signal at the 3' end provides for the A-tail synthesis (Borodulina and Kramerov, 2008).

#### Other functions

SINE RNA should not be involved in the processes with a cellular RNA, from which it originates (for example, RNA processing). This assumes the accumulation of changes from the original structure. For instance, transcripts of simple tRNA-derived SINEs cannot form the clover leaf structure, and their nucleotides are not modified as in tRNAs (Rozhdestvensky *et al.*, 2007; Sun *et al.*, 2007). As a result of such changes, SINE transcripts lose the capacity to bind to at least some protein factors of tRNA processing or transport. This excludes SINE transcripts from tRNA biochemical pathways and opens up a way for efficient retroposition. A similar pattern can be expected for the conversion of 7SL RNA and 5S rRNA pseudogenes into SINEs. For instance, B1 and Alu transcripts largely lose the similarity with the 7SL RNA secondary structure, although the structure of two domains is preserved (Labuda and Zietkiewicz, 1994).

In addition to transcription and reverse transcription, SINE replication involves other yet poorly known processes such as SINE RNA degradation or nuclear

export (Kramerov and Vassetzky, 2005). There is evidence that polyadenylation radically increases the lifetime of SINE RNA (Borodulina and Kramerov, 2008). The transport of SINE RNA is likely mediated by the interaction of its domains with cellular factors. For instance, Alu RNA transport is likely mediated by SRP9 and SRP14 (He *et al.*, 1994). It is not improbable that CORE and similar domains found in quite different (sometimes otherwise unrelated) SINE families participate in SINE RNA transport or some other function. Anyway, the absence of universal SINE structure responsible for its transport suggests different pathways of their RNA transport and, accordingly, different pathways of this function acquisition.

#### Further evolution of SINEs

*After the emergence, SINE families can further change. Minor changes in their structure (point mutations and indels) give rise to SINE subfamilies. More substantial changes (module exchange and duplication of modules or whole SINEs) give rise to new SINE families. SINE families and subfamilies can coexist or replace each other. Some of them (or even all) can lose their activity with time and extinct, while their gradually degrading copies remain in the genome.*

#### Emergence of SINE subfamilies

In all likelihood, only a minor fraction of SINE genomic copies is capable for retroposition (Roy-Engel *et al.*, 2002b). Active copies with beneficial (or neutral) modifications can give rise to new SINE subfamilies. One can propose that these changes correspond to the fine-tuning of SINEs to the critical factors of their amplification. For instance, the changes in Alu sequence modulating the Alu RNA capacity to bind the SRP9/14 complex gave rise to subfamilies with different amplification rate (Sarrowa *et al.*, 1997). LINE RT is another factor of SINE amplification. Considering that LINE subfamilies also replace each other in time, the structure of SINEs mobilized by them can also change accordingly (Human Genome Sequencing Consortium, 2001).

#### SINE dimerization

Although the majority of SINEs are monomeric, numerous dimeric (and even trimeric) SINE families exist. According to the number of their genomic copies, dimerization is usually a progressive evolutionary event; however, dimeric SINEs are not necessarily more successful than the monomeric counterparts. For instance, the dimers of B1 and ID are much more ample in the genomes of squirrels and dormice, whereas the opposite pattern is observed in the guinea pig genome (Kramerov and Vassetzky, 2001).

In addition to true dimers, there are SINEs with internal duplications (20–30 nt) called quasi-dimers. The best known (but not the only) example of this kind is rodent B1 (Figure 1c), which is much more successful than its predecessor pB1 without the internal duplication (Veniaminova *et al.*, 2007).

#### Module exchange

Long ago, an unusual property of SINEs was noted: their individual copies can have shuffled characters of different SINE subfamilies. This phenomenon was called

'mosaic evolution' (Labuda and Zietkiewicz, 1994; Zietkiewicz and Labuda, 1996) or 'gene conversion' (Maeda *et al.*, 1988; Roy-Engel *et al.*, 2002a). Such shuffling also occurs with SINE modules. For instance, the genome of wallaby harbors six SINEs, which amplified in different time periods with the help of different LINES (L1, L2, L3 and Bov-B). All of them share a similar tRNA-derived head and a CORE domain but differ in the 3'-terminal module and tail (Figure 2). Similar processes can go in SINE dimerization. Likewise, all combinations of major B1 and ID variants can be found among rodent dimeric SINEs (pB1-ID, B1-ID,

ID-pB1 and ID-B1; Veniaminova *et al.*, 2007; Churakov *et al.*, 2010).

Period of SINE activity

SINE families can lose their activity with time. For instance, Ther-1 and Ther-2 amplified in the genomes of vertebrate ancestors but are no more active at least in mammalian genomes (Human Genome Sequencing Consortium, 2001). B1 and ID have become inactive in the genomes of rat and mouse, respectively (Rat Genome Sequencing Consortium, 2004). A similar pattern is observed in SINE subfamilies remaining active over different evolutionary periods (Ohshima *et al.*, 2003; Liu *et al.*, 2009). Little is known about the factors that determine their duration, but it can substantially vary. Clearly, a decline in LINE activity makes the further amplification of the dependent SINE impossible. Thus, activity correlation is observed for many SINE/LINE partners, for example, Ther-1 and L2 in human and mouse (Human Genome Sequencing Consortium, 2001; Mouse Genome Sequencing Consortium, 2002); MEG and L1 in fruit bats (Cantrell *et al.*, 2008; Gogolevsky *et al.*, 2009) or Alu and L1 subfamilies in human (Ohshima *et al.*, 2003).

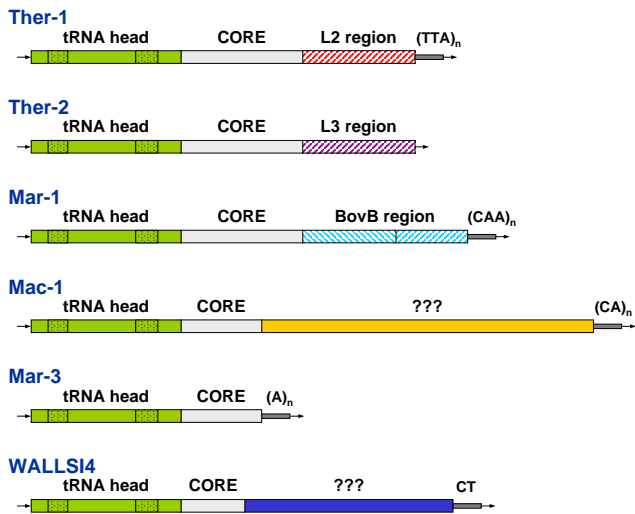


Figure 2 SINEs in the genome of wallaby mobilized by different LINES: Ther-1 (MIR), L2; Ther-2 (MIR3), L3; Mar-1, Bov-B (Gilbert and Labuda, 1999); Mar-3 (WSINE1), L1 (Munemasa *et al.*, 2008). The LINE partners of Mac-1 (WALLSI2; Munemasa *et al.*, 2008) and WALLSI4 (Jurka *et al.*, 2005) remain to be identified. Alternative SINE names are given in parentheses.

Mechanisms of SINE evolution

The life cycle of SINEs includes the DNA and RNA stages; accordingly, they can change in the form of DNA and RNA at different stages of their amplification. Although the 'common' mechanisms of nucleic acid variation can be important for SINE evolution, we will focus on the mechanisms with particular significance for this type of mobile genetic elements (Figure 3).

In the DNA replication cycle (DNA → DNA), two mechanisms of particular significance for SINE variation can be recognized. A huge number of SINE copies in the genome inevitably leads to homologous recombination between their non-allelic copies (for example, Bailey

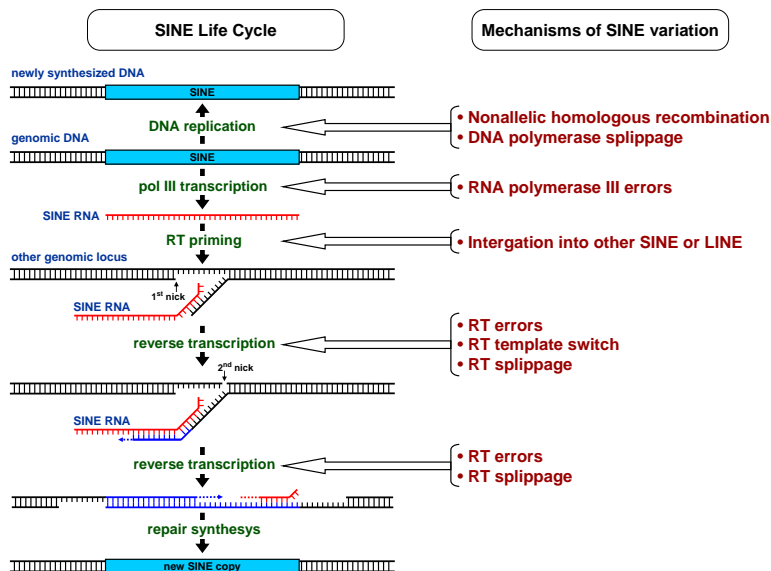


Figure 3 Mechanisms of SINE variation during their life cycle.

*et al.*, 2003). Recombination between copies falling into different SINE subfamilies or even families gives rise to hybrid SINEs (unless the genomic deletion or insertion is lethal). Such events can underlie both minor modifications in SINE structure ('mosaic evolution') and large-scale rearrangements (module acquisition/exchange, A-tail elongation, and dimerization).

Certain SINEs contain stretches of simple repeats (for example, (TC)<sub>n</sub> in CAN and C elements; Figure 1d). The length of such structures may vary significantly (Vassetzky and Kramerov, 2002), which can be attributed to DNA polymerase slippage during DNA replication in a way similar to microsatellites. The same mechanism can be applicable to the length variation in SINE tails.

Reverse transcription of SINEs is linked to the integration of their RNA into the genome. The RT endonuclease activity makes a break in the genomic DNA. The genomic sequence around the break has certain (usually not very high) specificity (for instance, the first break is usually made in 3'-AATTTT in the case of L1, which mobilizes most of currently active mammalian SINEs (Jurka, 1997)). In addition, the integration occurs into chromatin regions that are available when the SINE and LINE are transcribed. Altogether, this increases the probability of SINE integration into a site of previous SINE or LINE integration. This mechanism can be recruited for SINE dimerization as well as in the formation of RNA pseudogene/LINE 3' end hybrids during early SINE evolution.

In all likelihood, all RTs can switch between templates during reverse transcription. For instance, template switch takes place during the replication of retroviruses (Coffin *et al.*, 1997) and the proper LINEs (Bibillo and Eickbush, 2004; Babushok *et al.*, 2006). A switch between LINE and RNA pseudogene templates can underlie the emergence of SINEs (Gilbert and Labuda, 1999; Weiner, 2002), and indeed chimeric structures of this kind can be found in mammalian genomes (Gogvadze and Buzdin, 2005). A similar switch between templates of different SINEs can give rise to different modifications in their structure (module acquisition/exchange, A-tail elongation, and dimerization).

The ability of RTs to slip on the same template underlies the activity of telomerase, which reuses the same sequence as the template (Greider and Blackburn, 1989). A similar pattern has been demonstrated for a LINE RT reusing the same sequence to synthesize SINE tail (Kajikawa and Okada, 2002). It is not improbable that this mechanism underlies the A-tail elongation in SINEs mobilized by L1.

Likewise, certain RTs can jump on a template with direct repeats, for example, in retroviruses (Pathak and Temin, 1990). RT jumping between direct repeats leads to a duplication or deletion depending on the jump direction. Apparently, this mechanism underlies the emergence of many internal duplications and deletions in SINEs (Vassetzky *et al.*, 2003).

Finally, LINE RTs are capable of non-templated synthesis after the template has been read (Bibillo and Eickbush, 2004; Babushok *et al.*, 2006). This capacity can also contribute to SINE evolution by elongating their tail.

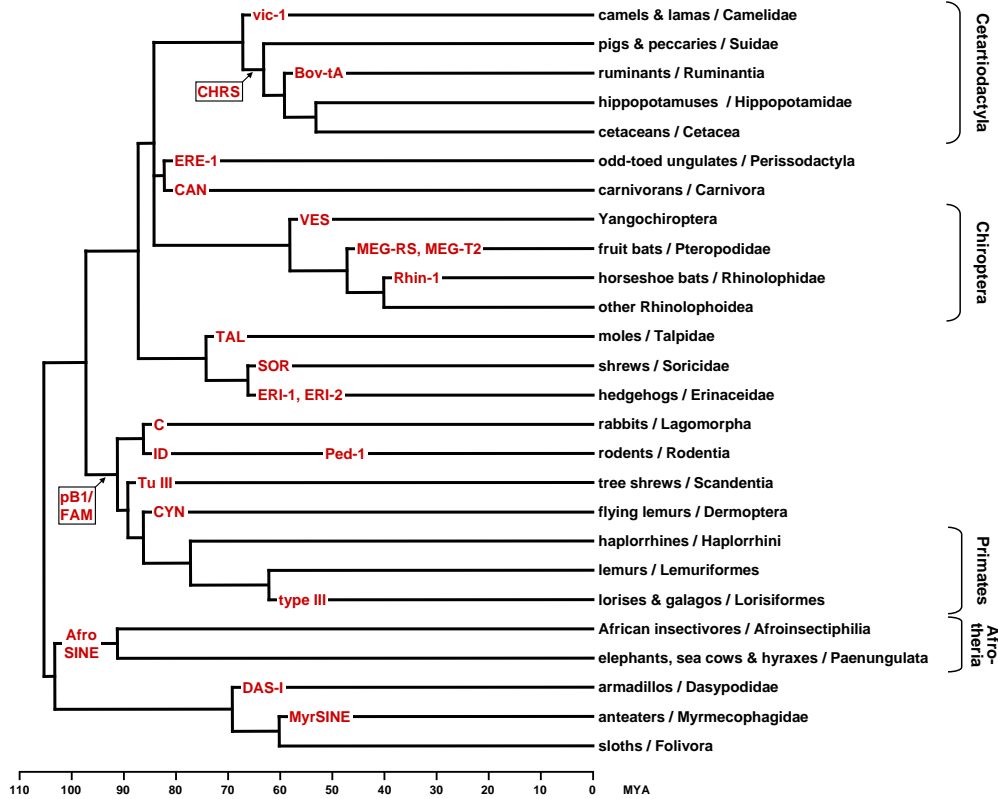
## Overview of SINE evolution

The organism's interaction with SINEs (as well as with other mobile genetic elements) largely resembles the

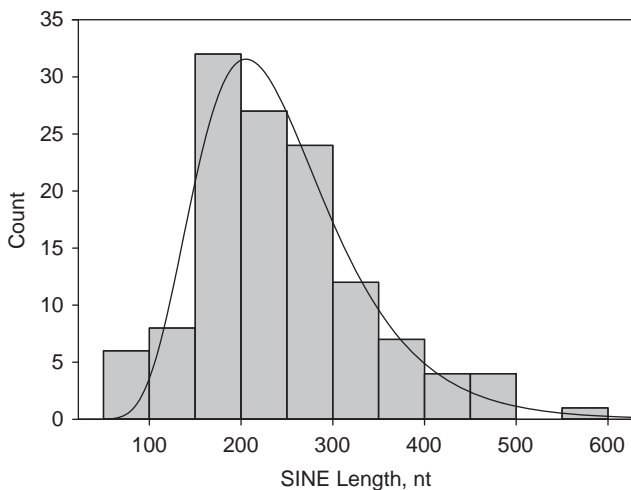
host-parasite coevolution. The integration of new SINE copies often disturbs gene expression; on the other hand, they can serve as a source of genomic innovations and a factor of genome plasticity (Makalowski, 2000). Nevertheless, the organism tries to suppress SINE amplification using, for example, APOBEC3-mediated system (Chiu *et al.*, 2006; Hulme *et al.*, 2007) or SINE DNA methylation (Rubin *et al.*, 1994). As LINE RT is required for SINE amplification, LINE repression also protects the genome from SINE expansion. LINE can be repressed through RNA interference or the APOBEC3 system, and the repression can be fixed by DNA methylation. The evolutionary dynamics of interactions between the organism and SINEs (as well as LINEs) resembles an arms race. At the extremes, too aggressive SINEs (or LINEs) can destroy their host organism and are eliminated by selection; on the other hand, there are many examples of SINE family death (cessation of amplification). More commonly, ups and downs in the activity of particular SINEs or LINEs are observed. This can be exemplified by the evolutionary waves of genome expansion by B1 or Alu subfamilies (Quentin, 1989; Ohshima *et al.*, 2003) or by the 100-times decline in the Alu retroposition frequency in current humans relative to primates 40–50 MYA (Batzer and Deininger, 2002). Amazingly, some dead SINEs can be 'reincarnated.' For instance, after inactivation of a LINE partner, the replacement of the 3'-terminal region with that of another (active) LINE gives rise to a new active SINE family. A demonstrative example of this kind can be found in wallaby genome, where a tRNA-CORE cassette consecutively replaced the 3'-terminal region and LINE partners (L2, L3, Bov-B, and L1; Figure 2). To a large extent, this and many other events in the evolution of SINEs are made possible by the huge number of their genomic copies, a fraction of which is transcribed even if their reverse transcription is impossible.

In contrast to other mobile genetic elements, SINEs emerged in evolution many times. For instance, at least 23 primary SINE families independently appeared in the evolution of placental mammals (currently, 51 mammalian SINE families have been described; Figure 4). This amazing property results, on the one hand, from their simple modular structure and the availability of the source modules (for example, tRNA or 3' end of LINE) in the cell. Moreover, high variation in SINE structures suggests that there are no stringent requirements for their nucleotide sequences excluding several short conserved regions. On the other hand, the emergence and replication of SINEs depend on LINE RT, which is not very secure from processing foreign sequences. Interestingly, some modules and RTs are particularly favorable for SINE emergence. For instance, alanine tRNA<sup>CGC</sup> independently gave rise to three simple SINEs (ID in rodents, vic-1 in camels and DAS-I in armadillos; Borodulina and Kramerov, 2005). Likewise, SINE families mobilized by mammalian L1 are particularly abundant. At present, we have no clue what properties of alanine tRNA and L1 RT proved beneficial for SINE emergence and amplification.

Further SINE evolution involves the complication of their structure by internal duplications, acquisition of new modules (such as CORE) and dimerization. Although simple SINEs can be highly prolific, the majority of successful SINEs are longer than 150 bp and



**Figure 4** The *de novo* emergence of SINEs in placental mammals. The mammalian tree corresponds to the TimeTree Knowledge Base (Hedges *et al.*, 2006).



**Figure 5** Length distribution of SINE families (without tail; plotted for 125 elements).

have a more complex structure (Figure 5). It is worth mentioning one more property of SINE evolution, module exchange. Although such recombination occurs in other genetic elements, it is unusually frequent in SINEs, which provides extra flexibility to their evolution. In a sense, SINE dimerization can also be considered as a special case of module exchange.

Owing to *de novo* emergence of SINEs and module exchange/dimerization, large-scale evolution of SINEs cannot be presented as a common phylogenetic tree (although short periods of SINE evolution can), which

distinguishes it from the evolution of genes and other mobile genetic elements presentable as a common bifurcating tree.

Mammals (placentals, marsupials and monotremes), reptiles, fishes and cephalopods have a large number of different active SINE families. Amazingly, they are absent from *Drosophila* species and chicken (although the chicken genome contains copies of inactive Ther-1, which amplified in the genomes of vertebrate ancestors), at the same time, their genomes have active LINES. One can speculate that these LINES lack some properties essential for SINE mobilization; it is also possible that *de novo* emergence of a SINE is a very rare event, and the odds are that it never occurred in certain genomes. Finally, SINEs could emerge but failed to survive because of some properties of host genomes (for instance, the *Drosophila* genome is relatively small, which can point to the mechanisms counteracting mobile element expansion). The rapid progress in comparative genomics of eukaryotes shows promise that this and other mysteries of SINE origin and evolution will be solved.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

This study was supported by the Molecular and Cellular Biology Program of the Russian Academy of Sciences and the Russian Foundation for Basic Research (project nos. 10-04-01259-a and 11-04-00439-a).

## References

- Akasaki T, Nikaido M, Nishihara H, Tsuchiya K, Segawa S, Okada N (2010). Characterization of a novel SINE superfamily from invertebrates: 'Ceph-SINES' from the genomes of squids and cuttlefish. *Gene* **454**: 8–19.
- Arnaud P, Yukawa Y, Lavie L, Pelissier T, Sugiura M, Deragon JM (2001). Analysis of the SINE S1 Pol III promoter from Brassica; impact of methylation and influence of external sequences. *Plant J* **26**: 295–305.
- Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian Jr HH (2006). L1 integration in a transgenic mouse model. *Genome Res* **16**: 240–250.
- Bailey JA, Liu G, Eichler EE (2003). An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* **73**: 823–834.
- Batzler MA, Deininger PL (2002). Alu repeats and human genomic diversity. *Nat Rev Genet* **3**: 370–379.
- Bibillo A, Eickbush TH (2004). End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J Biol Chem* **279**: 14945–14953.
- Borodulina OR, Kramerov DA (2001). Short interspersed elements (SINES) from insectivores. Two classes of mammalian SINES distinguished by A-rich tail structure. *Mamm Genome* **12**: 779–786.
- Borodulina OR, Kramerov DA (2005). PCR-based approach to SINE isolation: simple and complex SINES. *Gene* **349**: 197–205.
- Borodulina OR, Kramerov DA (2008). Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. *RNA* **14**: 1865–1873.
- Cantrell MA, Scott L, Brown CJ, Martinez AR, Wichman HA (2008). Loss of LINE-1 activity in the megabats. *Genetics* **178**: 393–404.
- Chesnokov I, Schmid CW (1996). Flanking sequences of an Alu source stimulate transcription *in vitro* by interacting with sequence-specific transcription factors. *J Mol Evol* **42**: 30–36.
- Chiu YL, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C *et al.* (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc Natl Acad Sci USA* **103**: 15588–15593.
- Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J (2010). Rodent evolution: back to the root. *Mol Biol Evol* **27**: 1315–1326.
- Churakov G, Smit AF, Brosius J, Schmitz J (2005). A novel abundant family of retroposed elements (DAS-SINES) in the nine-banded armadillo (*Dasypus novemcinctus*). *Mol Biol Evol* **22**: 886–893.
- Coffin JM, Hughes SH, Varmus H (1997). *Retroviruses*. Cold Spring Harbor Laboratory Press: Plainview, NY.
- Daniels GR, Deininger PL (1983). A second major class of Alu family repeated DNA sequences in a primate genome. *Nucleic Acids Res* **11**: 7595–7610.
- Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J Mol Biol* **151**: 17–33.
- Deininger PL, Tiedge H, Kim J, Brosius J (1996). Evolution, expression, and possible function of a master gene for amplification of an interspersed repeated DNA family in rodents. *Prog Nucleic Acid Res Mol Biol* **52**: 67–88.
- Deragon JM, Zhang X (2006). Short interspersed elements (SINES) in plants: origin, classification, and use as phylogenetic markers. *Syst Biol* **55**: 949–956.
- Dewannieux M, Esnault C, Heidmann T (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**: 41–48.
- Eickbush T, Malik H (2002). Origins and evolution of retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AL (eds). *Mobile DNA II*. ASM Press: Washington, DC, pp 1111–1144.
- Esnault C, Maestre J, Heidmann T (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**: 363–367.
- Feschotte C, Pritham EJ (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**: 331–368.
- Gilbert N, Labuda D (1999). CORE-SINES: eukaryotic short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci USA* **96**: 2869–2874.
- Gogolevsky KP, Vassetzky NS, Kramerov DA (2008). Bov-B-mobilized SINES in vertebrate genomes. *Gene* **407**: 75–85.
- Gogolevsky KP, Vassetzky NS, Kramerov DA (2009). 5S rRNA-derived and tRNA-derived SINES in fruit bats. *Genomics* **93**: 494–500.
- Gogvadze EV, Buzdin AA (2005). New mechanism of retrogene formation in mammalian genomes: *in vivo* recombination during RNA reverse transcription. *Mol Biol* **39**: 364–373.
- Gong C, Maquat LE (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**: 284–288.
- Greider CW, Blackburn EH (1989). A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* **337**: 331–337.
- Havecker ER, Gao X, Voytas DF (2004). The diversity of LTR retrotransposons. *Genome Biol* **5**: 225.
- He XP, Bataille N, Fried HM (1994). Nuclear export of signal recognition particle RNA is a facilitated process that involves the Alu sequence domain. *J Cell Sci* **107**(Part 4): 903–912.
- Hedges SB, Dudley J, Kumar S (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**: 2971–2972.
- Hulme AE, Bogerd HP, Cullen BR, Moran JV (2007). Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* **390**: 199–205.
- Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Jurka J (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc Natl Acad Sci USA* **94**: 1872–1877.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**: 462–467.
- 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.
- Kramerov DA, Vassetzky NS (2001). Structure and origin of a novel dimeric retroposon B1-dID. *J Mol Evol* **52**: 137–143.
- Kramerov DA, Vassetzky NS (2005). Short retrotransposons in eukaryotic genomes. *Int Rev Cytol* **247**: 165–221.
- Krayev AS, Kramerov DA, Skryabin KG, Ryskov AP, Bayev AA, Georgiev GP (1980). The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res* **8**: 1201–1215.
- Labuda D, Sinnett D, Richer C, Deragon JM, Striker G (1991). Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. *J Mol Evol* **32**: 405–414.
- Labuda D, Zietkiewicz E (1994). Evolution of secondary structure in the family of 7SL-like RNAs. *J Mol Evol* **39**: 506–518.
- Liu GE, Alkan C, Jiang L, Zhao S, Eichler EE (2009). Comparative analysis of Alu repeats in primate genomes. *Genome Res* **19**: 876–885.
- Lovsin N, Gubensek F, Kordis D (2001). Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. *Mol Biol Evol* **18**: 2213–2224.
- Maeda N, Wu CI, Bliska J, Reneke J (1988). Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. *Mol Biol Evol* **5**: 1–20.



- Makalowski W (2000). Genomic scrap yard: how genomes utilize all that junk. *Gene* **259**: 61–67.
- Maraia RJ (1991). The subset of mouse B1 (Alu-equivalent) sequences expressed as small processed cytoplasmic transcripts. *Nucleic Acids Res* **19**: 5695–5702.
- Matveev V, Okada N (2009). Retroposons of salmonoid fishes (Actinopterygii: Salmonoidei) and their evolution. *Gene* **434**: 16–28.
- Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Munemasa M, Nikaido M, Nishihara H, Donnellan S, Austin CC, Okada N (2008). Newly discovered young CORE-SINEs in marsupial genomes. *Gene* **407**: 176–185.
- Nishihara H, Smit AF, Okada N (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res* **16**: 864–874.
- Nishihara H, Terai Y, Okada N (2002). Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. *Mol Biol Evol* **19**: 1964–1972.
- 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, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N (2003). Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. *Genome Biol* **4**: R74.
- Ohshima K, Okada N (2005). SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenet Genome Res* **110**: 475–490.
- Pathak VK, Temin HM (1990). Broad spectrum of in vivo forward mutations, hypermutations, and mutational hot-spots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc Natl Acad Sci USA* **87**: 6019–6023.
- Ponicsan SL, Kugel JF, Goodrich JA (2010). Genomic gems: SINE RNAs regulate mRNA production. *Curr Opin Genet Dev* **20**: 149–155.
- Quentin Y (1989). Successive waves of fixation of B1 variants in rodent lineage history. *J Mol Evol* **28**: 299–305.
- Rat Genome Sequencing Consortium (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493–521.
- Roy-Engel AM, Carroll ML, El-Sawy M, Salem AH, Garber RK, Nguyen SV *et al.* (2002a). Non-traditional Alu evolution and primate genomic diversity. *J Mol Biol* **316**: 1033–1040.
- Roy-Engel AM, El-Sawy M, Farooq L, Odom GL, Perepelitsa-Belancio V, Bruch H *et al.* (2005). Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet Genome Res* **110**: 365–371.
- Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE *et al.* (2002b). Active Alu element ‘A-tails’: size does matter. *Genome Res* **12**: 1333–1344.
- Rozhdestvensky TS, Crain PF, Brosius J (2007). Isolation and posttranscriptional modification analysis of native BC1 RNA from mouse brain. *RNA Biol* **4**: 11–15.
- Rubin CM, VandeVoort CA, Teplitz RL, Schmid CW (1994). Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic Acids Res* **22**: 5121–5127.
- Sarrowa J, Chang DY, Maraia RJ (1997). The decline in human Alu retroposition was accompanied by an asymmetric decrease in SRP9/14 binding to dimeric Alu RNA and increased expression of small cytoplasmic Alu RNA. *Mol Cell Biol* **17**: 1144–1151.
- Schmitz J, Zischler H (2003). A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. *Mol Phylogenet Evol* **28**: 341–349.
- Schramm L, Hernandez N (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**: 2593–2620.
- Serdobova IM, Kramerov DA (1998). Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. *J Mol Evol* **46**: 202–214.
- Sun FJ, Fleurdepine S, Bousquet-Antonelli C, Caetano-Anolles G, Deragon JM (2007). Common evolutionary trends for SINE RNA structures. *Trends Genet* **23**: 26–33.
- Ullu E, Tschudi C (1984). Alu sequences are processed 7SL RNA genes. *Nature* **312**: 171–172.
- Ullu E, Weiner AM (1985). Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature* **318**: 371–374.
- Vassetzky NS, Kramerov DA (2002). CAN—a pan-carnivore SINE family. *Mamm Genome* **13**: 50–57.
- Vassetzky NS, Ten OA, Kramerov DA (2003). B1 and related SINEs in mammalian genomes. *Gene* **319**: 149–160.
- Veniaminova NA, Vassetzky NS, Kramerov DA (2007). B1 SINEs in different rodent families. *Genomics* **89**: 678–686.
- Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH *et al.* (2001). Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol Cell Biol* **21**: 1429–1439.
- Weichenrieder O, Wild K, Strub K, Cusack S (2000). Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* **408**: 167–173.
- Weiner AM (2002). SINEs and LINEs: the art of biting the hand that feeds you. *Curr Opin Cell Biol* **14**: 343–350.
- Zietkiewicz E, Labuda D (1996). Mosaic evolution of rodent B1 elements. *J Mol Evol* **42**: 66–72.
- Zietkiewicz E, Richer C, Sinnett D, Labuda D (1998). Monophyletic origin of Alu elements in primates. *J Mol Evol* **47**: 172–182.