Genome Analysis

# Evolutionary history of 7SL RNA-derived SINEs in Supraprimates

**Jan Ole Kriegs[1*], Gennady Churakov[1*], Jerzy Jurka[2], Jürgen Brosius[1] and Jürgen Schmitz[1]**

[1] Institute of Experimental Pathology, ZMBE, University of Münster, Von-Esmarch-Str. 56, D-48149 Münster, Germany
[2] Genetic Information Research Institute, 1925 Landings Drive, Mountain View, CA 94043, USA

**The evolutionary relationships of 7SL RNA-derived SINEs such as the primate *Alu* or the rodent B1 elements have hitherto been obscure. We established an unambiguous phylogenetic tree for Supraprimates, and derived intraordinal relationships of the 7SL RNA-derived SINEs. As well as new elements in *Tupaia* and primates, we also found that the purported ancestral fossil *Alu* monomer was restricted to Primates, and provide here the first description of a potential chimeric promoter box region in SINEs.**

## Introduction

Short interspersed elements (SINEs) are known to occur throughout the genomes of eukaryotes. They originate from retroposition of small RNAs such as 7SL RNA, tRNA or their derivatives featuring internal RNA polymerase III promoters [1]. Active SINEs can cause hereditary diseases [2], enable genomic rearrangements [3] and be exapted into functional proteins [4–6]. SINEs originally derived from 7SL RNA [7,8], a component of the cytoplasmic signal recognition particle [9], have been found in primates [10], scandentians (tree shrews) [11] and rodents [12], all members of the placental mammalian clade Supraprimates [13,14] (Euarchontoglires [15]), which also includes dermopterans (flying lemurs) and lagomorphs (rabbits). Assuming that 7SL RNA-derived SINEs arose in a common ancestor of these orders [16], they should also be present in the other two Supraprimate orders, but this has yet to be demonstrated [17].

The fossil *Alu* monomer (FAM) [16] was thought to be the oldest common ancestor of all 7SL RNA-derived SINEs [18]. Compared with human 7SL RNA, FAMs have a large central deletion (Figure 1) [16] and are believed to have given rise to subsequent monomeric sequences, namely the primate-specific free right *Alu* monomer (FRAM) [16], and to the free left *Alu* monomer (FLAM) of the A-subtype (FLAM-A, or FLA) [19]. In primates FLAM-C is a descendant of FLAM-A, and its fusion to FRAM gave rise to the predominating dimeric *Alu* family of 7SL RNA-derived SINEs [20].

The rodent proto-B1 element (PB1) is nearly identical to FLAM-A [18]. Rodent B1 elements are monomeric SINEs, but contain a diagnostic internal 29-bp duplication

(Figure 1: B1 R). They descended from FLAM-A (or PB1) through PB1D with a characteristic deletion [17] that is lacking in the primate 7SL RNA-derived SINEs of the *Alu* family.

In Scandentia, the tree shrew-specific chimeric Tu type II SINEs (composed of tRNA and 7SL RNA parts) share with the rodents this diagnostic deletion in the left 7SL RNA-derived part [11]. Nishihara *et al.* used this shared deletion to suggest a possible evolutionary scenario with a more recent common ancestry of Tu type II and rodent B1 SINEs and, as a potential consequence, a sister group relationship of the orders Rodentia and Scandentia [11].

Because 7SL RNA-derived SINEs are specific for Supraprimates, their evolutionary histories are also of particular interest. However, because the current evidence for ordinal relationships within Supraprimates is partially contradictory, the distribution patterns, and the evolutionary history, of 7SL RNA-derived SINEs might differ from proposed scenarios. Some large-scale sequence data support the Euarchonta theory combining Primates, Dermoptera and Scandentia in one group [13,15], whereas other studies indicate the placement of tree shrews with lagomorphs [21–23]. Furthermore, similarities in lineage-specific 7SL RNA-derived SINEs [11] support a clustering of Rodentia and Scandentia, further confusing the issue. The phylogenetic position of lagomorphs within the placental tree is also equivocal, with some data supporting the monophyletic group Glires, comprising rodents and lagomorphs [15,24,25], and others placing lagomorphs as a sister group to a clade Cetferungulata+Xenarthra+Afrotheria [22]. Thus, to analyze the evolutionary history of Supraprimate-specific 7SL RNA-derived SINE retroposons, we chose first to resolve current controversies concerning the interordinal relationships within the Supraprimate clade and establish a robust evolutionary tree of this superorder.

## Resolving the evolutionary tree of Supraprimates

We searched genomic sequences of mouse, rat, guinea pig, rabbit, tree shrew, chimpanzee, human and rhesus monkey with the local version of RepeatMasker (http://www.repeatmaster.org/) for phylogenetically informative retroposon presence–absence data conclusively to accept or reject current hypotheses of Supraprimate evolution (see Online Supplementary Material). We found 14 informative genomic loci that unambiguously define two main branches within Supraprimates (Figure 2, and Online Supplementary Material), one MLT1A1 and four MSTD
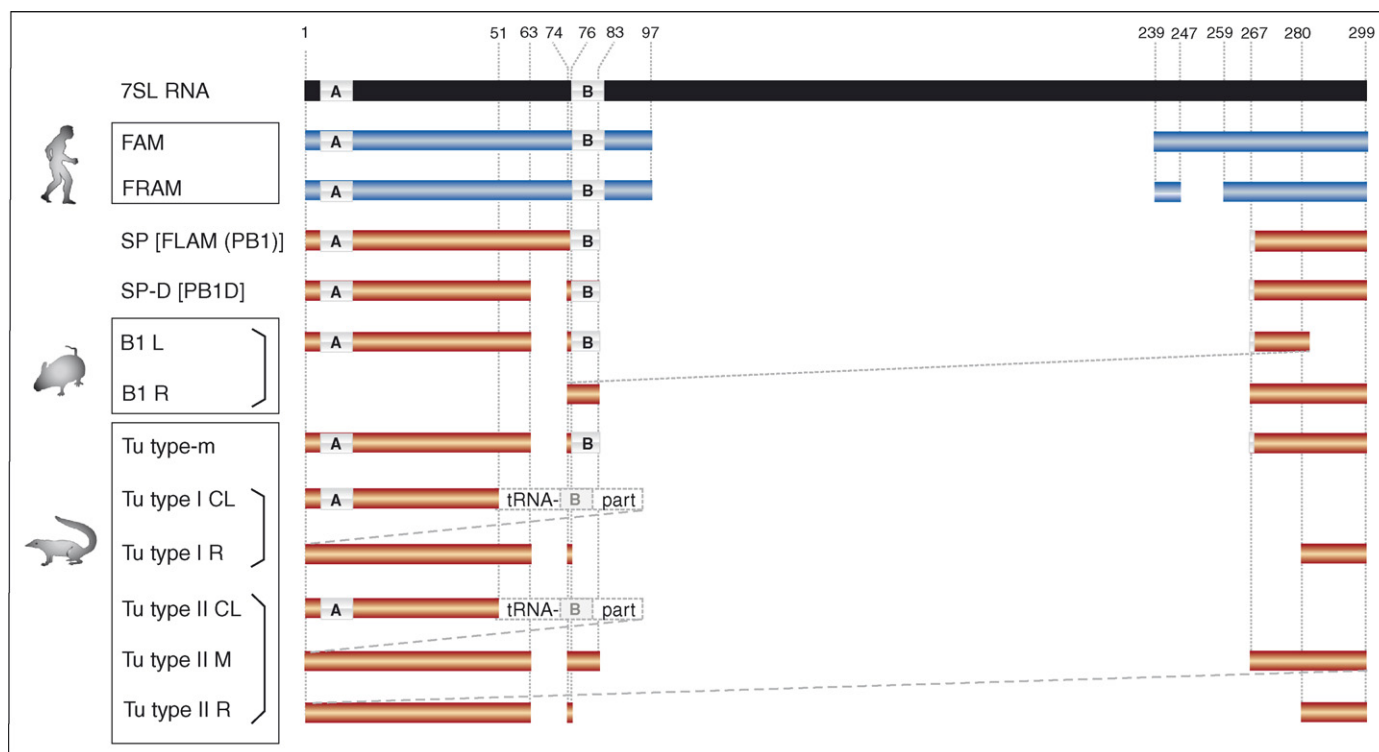
**Figure 1.** Illustration of 7SL RNA and the derived SINE sequence regions. All sequence positions (numbers on top) correspond to those of human 7SL RNA (top, solid black bar). All presented 7SL RNA-derived sequences probably arose by two independent events. A FAM retroposon evolved from a 7SL RNA sequence and gave rise to the FRAM part of *Alu* SINEs (blue bars). Independently an SP – FLAM (PB1) – sequence evolved from a 7SL RNA sequence and eventually gave rise, through the SP-D (PB1D) element to all other 7SL RNA-derived SINE elements (red bars). A and B denote the two RNA polymerase III promoter-box elements. L, M and R refer to the left, middle and right parts, CL to the chimeric left part, of dimeric 7SL RNA-derived SINE elements. The tRNA-derived part including the promoter B box is not related to 7SL RNA and is shown in a dashed-enclosed box. SP refers to the suggested new terminology for Supraprimates. By realigning the right 7SL RNA-derived parts of Tu types I and II, we identified the same diagnostic deletion as described for the left 7SL RNA-derived part of Tu type II [11].

long terminal repeats (LTRs) retroposons inserted in independent genomic loci of human and tree shrew. One of these loci, containing an LTR element, was amplifiable in the flying lemur as well. The elements in these loci are clearly absent in the mouse, rat and dog genomes, and although two of the loci could be detected in rabbit trace sequences, no retroposons were present in either case, thereby establishing the monophyly of Euarchonta. Nine additional independent retroposed elements (2 × MLT1A0, MLT1A1, 4 × L1MA5, L1MA6 and ORR1E) were present in orthologous genomic loci in mouse, rat and rabbit, whereas the orthologous target sites were empty in human, chimpanzee, rhesus monkey and dog. These data, along with two other retroposon markers [25], clearly group the Lagomorpha and Rodentia in a common clade, Glires.

### Elucidating the evolutionary history of 7SL RNA-derived SINEs

Placing the tree shrew in the Euarchonta clade, however, apparently contradicts one current hypothesis of 7SL RNA-derived SINE evolution [11] that makes use of a common diagnostic deletion (positions 64–73) in Rodent B1 and scandentian Tu type II SINEs to suggest a sister group relationship of Rodentia and Scandentia [11,17]. By contrast, our support for the separate Euarchonta and Glires clades, coupled with the common deletion in Rodentia and Scandentia, suggest that there is a common ancestry of all 7SL RNA-derived SINEs. Thus, one would also expect to find some SINE sequences with this deletion in

Lagomorpha, Dermoptera and Primates. To test this, we performed multiple computational searches of available genomic and trace sequences of various Supraprimate species.

In human, we identified 17 complete copies (see Online Supplementary Material) of a sequence related to a PB1D-like monomer, containing the diagnostic deletion common to Rodentia and Scandentia. Sixteen of these copies are also conserved at orthologous positions in chimpanzee and rhesus monkey. This is the first time such an element has been found in Primates. None of these elements were flanked by perfect direct repeats, suggesting that they are relatively old, no longer active SINEs, which might explain why they were not previously identified.

PB1D elements were thought to be rodent-specific; however, it might be appropriate to designate them as Supra-primate-specific elements. We propose 'SP-D', 'SP' for Supraprimates and 'D' for the diagnostic deletion at positions 64–73. Accordingly, FLAM-A (PB1)-like sequences should be named 'SP-SINEs'.

In available genomic trace sequences of *Tupaia belangeri* we found a previously unknown 7SL RNA-derived monomeric SINE that was present in high copy number (∼300 copies) and is similar to the rodent PB1D10, which we designate as Tu type-m (for monomer) (Figure 1, and Online Supplementary Material). Flanking short direct repeats were detected in 57 examples (see Online Supplementary Material) and indicate recent activity of the corresponding master gene.
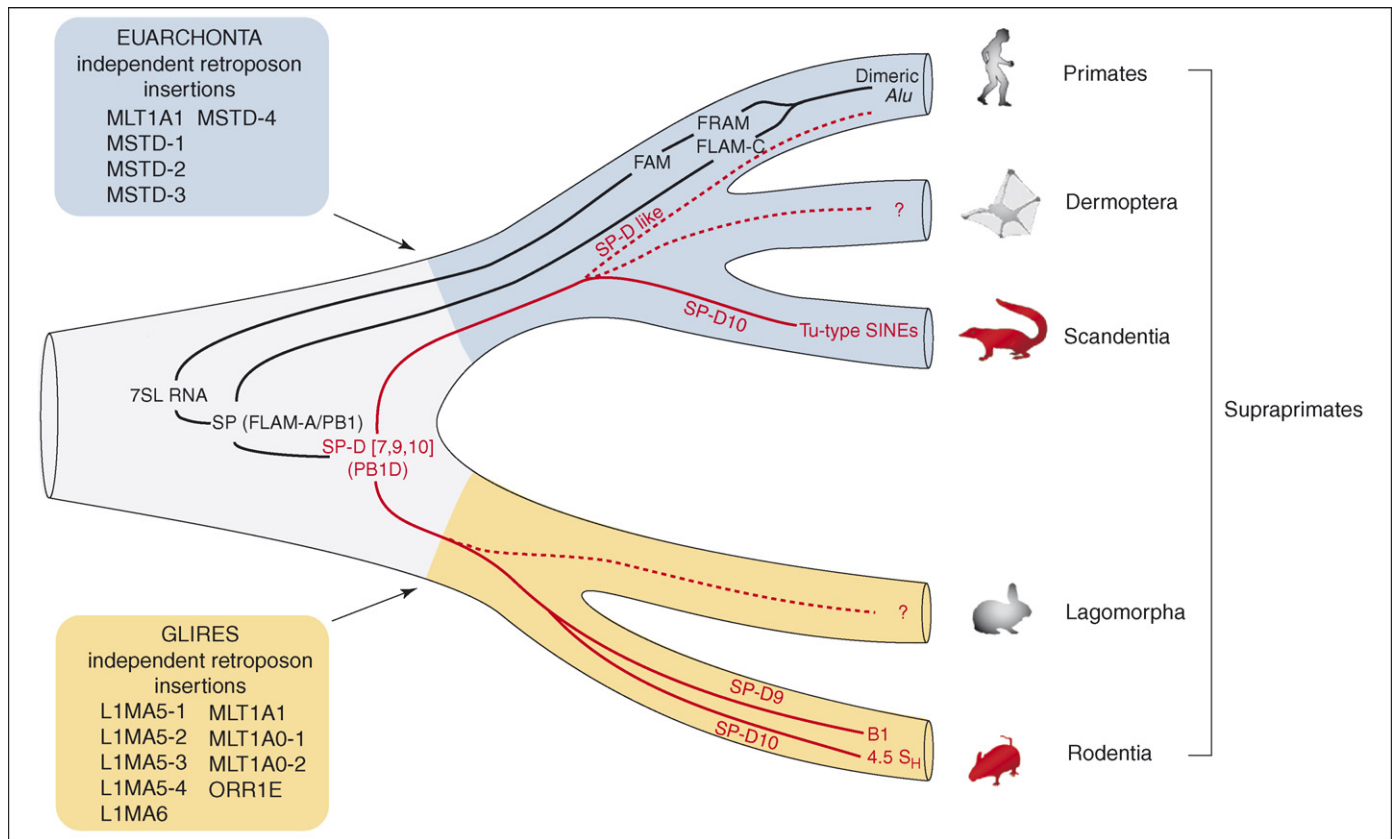
**Figure 2**. Phylogenetic tree of Supraprimates showing distribution and evolution of 7SL RNA-derived SINEs. The retroposed elements in the blue panel represent single retroposed insertions that group the species of the Euarchonta clade, whereas those in the beige panel depict single insertions that group the species of the Glires clade. Bold lines on the branches of the tree depict SINEs generated from recently active master genes, whereas the dotted lines symbolize SINEs from inactive master genes. The SINEs depicted before the main branching point of the tree were active before this divergence and hence are found as molecular fossils in both branches of Supraprimates. The lack of genomic information concerning SP-D elements in the flying lemur and the rabbit is indicated by question marks. Animals in red silhouettes host probably recently active 7SL RNA-derived SINEs with SP-D as a common ancestral form. The fates of SP-D-related sequences are shown as red lines. $4.5S_H$ denotes the SINE offspring of a 7SL RNA-derived $4.5S_H$ master gene.

We did not identify any complete PB1D-like elements in the rabbit. The sequence divergence is probably too great to recognize clear PB1D-like sequences in the small amount of rabbit genomic data that is available as trace sequences. Although the phylogenetic position of dermopterans implies the presence of 7SL RNA-derived SINEs, searchable genomic data are not yet available to test this prediction. Although no FLAM (PB1)-like sequences have yet been discovered in Dermoptera and Lagomorpha species, our new human data support the conclusion that they are distributed throughout all Supraprimate orders.

Quentin proposed that FAM was the progenitor of all other 7SL RNA-derived SINEs [16]. This implies that not only FLAM (PB1)-like elements should be present in all Supraprimate genomes but also their presumed common progenitor, FAM. Interestingly, in spite of extensive BLAST searches, not a single FAM sequence was detected in tree shrew, rabbit or guinea pig genomic trace sequences or in mouse and rat full genomes. Fragments of FAM-like sequences that were occasionally identified by RepeatMasker did not include sequences overlapping the FAM-specific diagnostic deletion (positions 98–238). Assuming they are found to be absent in flying lemurs as well, FAM sequences seem to be restricted to Primates (Figure 2).

Inasmuch as the central deletion of FLAM overlaps the central deletion of FAM (Figure 1), two possible evolutionary scenarios, one with successive deletions (7SL RNA-dele-

tion→FAM-deletion→FLAM [16]) or one with independent events acting on 7SL RNA (7SL RNA-deletion→FAM; 7SL RNA-deletion→FLAM) are equally parsimonious. Our finding that FAM elements are probably restricted to Primates supports the second of these scenarios and contradicts the notion of a FAM progenitor for all 7SL RNA-derived SINE forms. Furthermore, the presence of FLAM-A (PB1)- and PB1D-like sequences in species of both major branches of Supraprimates indicates that the common ancestor of all 7SL RNA-derived SINEs was probably a FLAM-A (PB1) or subsequent PB1D-like form. We propose that FAM, rather than being a progenitor for all 7SL RNA-derived SINEs, is a primate-specific monomer that gave rise to FRAM monomers, whereas SP-SINEs contributed the left part of the *Alu* dimer. This would then make SP-D the progenitor of both B1 elements in Rodentia and the 7SL RNA-derived Tu-type SINE parts of Scandentia (Figure 2).

Our discovery of a new SP-D-like monomeric SINE form in Scandentia suggests that the Tu type I and II SINEs evolved from dimers of Tu type-m (Figure S1 in Online Supplementary Material). These elements were originally thought to be derived from tRNA and 7SL RNA parts [11]; however, with the benefit of updated trace sequences we find that the 5′ end of the 'tRNA' has a greater sequence similarity to Tu type-m (70%, 51 nt) and the 3′ end is 75% (41 nt) identical to tRNA$^{Gly}$, suggesting that the first 92 nt is a chimeric sequence rather than a completely tRNA-

derived sequence (see Online Supplementary Material). This would imply that the box A component of the promoter is derived from a 7SL RNA-like polymerase III promoter sequence and the box B component from tRNA$^{Gly}$. To our knowledge this is the first description of a promoter box fusion in SINEs.

## Concluding remarks

Using a combination of genomic and trace sequence information we have established a clear phylogenetic framework of the Supraprimates that unambiguously confirms the clades Euarchonta and Glires. This enables us to elucidate and redefine the evolutionary history of 7SL RNA-derived SINE elements, the birth of which was traced to the common ancestor of Supraprimates. Instead of the previously proposed FAM element ancestor, which probably gave rise to only primate-specific SINEs, we propose a new designation of SP elements – FLAM-A (PB1) – as the ancestral, Supraprimate-specific 7SL RNA-derived SINE. These elements are recognizable as fossil remnants in human, but were recently still active in rodents and scandentians. In *Tupaia* a complex rearrangement of one such SP-D-like element and an active tRNA-derived SINE component presumably gave rise to a 7SL–tRNA mixed promoter box region in these elements.

## Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.02.002.

## References

1 Kramerov, D.A. and Vassetzky, N.S. (2005) Short retroposons in eukaryotic genomes international review of cytology. In *A Survey of Cell Biology* (Jeon, K.W., ed.), pp. 165–221, Academic Press

2 Callinan, P. and Batzer, M.A. (2006) Retrotransposable elements and human disease. In *Genome and Disease. Genome Dynamics* (Vol. 1) (Volff, J., ed.), pp. 104–115, Karger

3 Sen, S.K. *et al.* (2006) Human genomic deletions mediated by recombination between Alu elements. *Am. J. Hum. Genet.* 79, 41–53

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

5 Sorek, R. *et al.* (2002) Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060–1067

6 Krull, M. *et al.* (2005) Alu-SINE exonization: En route to protein-coding function. *Mol. Biol. Evol.* 22, 1702–1711

7 Zieve, G. *et al.* (1977) Synthesis of two classes of small RNA species *in vivo* and *in vitro*. *Biochemistry* 4, 4520–4525

8 Ullu, E. and Tschudi, C. (1984) Alu sequences are processed 7SL RNA genes. *Nature* 312, 171–172

9 Walter, P. and Blobel, G. (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* 299, 691–698

10 Deininger, P. *et al.* (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol.* 151, 17–33

11 Nishihara, H. *et al.* (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

12 Krayev, A. *et al.* (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

13 Waddell, P.J. *et al.* (2001) A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.* 12, 141–154

14 Kriegs, J.O. *et al.* (2006) Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* 4, e91

15 Murphy, W.J. *et al.* (2001) Resolution of the early placental mammal radiation using bayesian phylogenetics. *Science* 294, 2348–2351

16 Quentin, Y. (1992) Origin of the Alu family: A family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res.* 20, 3397–3401

17 Vassetzky, N.S. *et al.* (2003) B1 and related SINEs in mammalian genomes. *Gene* 319, 149–160

18 Quentin, Y. (1994) A master sequence related to a Free Left Alu Monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic Acids Res.* 22, 2222–2227

19 Jurka, J. (1995) Origin and evolution of Alu repetitive elements. In *The Impact of Short INterspersed Repeats (SINEs) on the Host Genome* (Maraia, R., ed.), pp. 29–41, Springer

20 Quentin, Y. (1992) Fusion of a Free Left Alu Monomer and a Free Right Alu Monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res.* 20, 487–493

21 Schmitz, J. *et al.* (2000) The complete mitochondrial genome of *Tupaia belangeri* and the phylogenetic affiliation of Scandentia to other eutherian orders. *Mol. Biol. Evol.* 17, 1334–1343

22 Arnason, U. *et al.* (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8151–8156

23 Madsen, O. *et al.* (2001) Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409, 610–614

24 Lin, Y-H. *et al.* (2002) Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Mol. Biol. Evol.* 19, 2060–2070

25 Nishihara, H. *et al.* (2006) Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9929–9934