Available online at www.sciencedirect.com

ScienceDirect

Genomics 89 (2007) 678–686

GENOMICS

www.elsevier.com/locate/ygeno

B1 SINEs in different rodent families☆

Natalia A. Veniaminova, Nikita S. Vassetzky, Dmitri A. Kramerov *

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov Street, Moscow 119991, Russia

Received 30 October 2006; accepted 28 February 2007
Available online 12 April 2007

Abstract

B1 SINEs were studied in 22 families covering all major rodent lineages. The number of B1 copies considerably varies, from 1×10^4 in Geomyidae to 1×10^6 in Myodonta. B1 sequences can be divided into three main structural variants: B1 with a 20-bp tandem duplication (found in Gliridae, Sciuridae, and Aplodontidae), B1 with a 29-bp duplication (found in other families), and proto-B1 without duplication (pB1). These variants can be further subdivided according to their characters, including specific 7-, 9-, or 10-bp deletions. Different B1 subfamilies predominate in different rodent families. The analysis of B1 variants allowed us to propose possible pathways for the evolution of this SINE in the context of rodent evolution.

Keywords: Retroposon; Rodentia; Evolution; Phylogeny; 7SL RNA

Genomes of many eukaryotes carry repeated 80- to 400-bp-long sequences called short interspersed elements (SINEs) or short retroposons, since they propagate in the genome by retroposition, a process involving reverse transcription of RNA and subsequent integration into the genome [1–3]. This is accomplished by the reverse transcriptase encoded in another class of mobile genetic elements, long interspersed elements (LINEs) [4,5].

Copies of a given short retroposon (a SINE family) display some sequence variation (5–35% depending on their age). Commonly, each SINE family is specific for organisms of one or several families or orders. Cellular RNA polymerase III can transcribe SINEs due to the presence of an internal promoter in their 5′ region, which is composed of A and B boxes spaced 30–40 nucleotides apart. Most SINE families originate from various tRNAs. However, two SINE families that were discovered first, B1 of mice, rats, and hamsters [6–8] and *Alu* of humans [9], originated from 7SL RNA, a component of a cytoplasmic ribonucleoprotein called *signal recognition particle* involved in translation of secreted proteins in all eukaryotes [10]. Later, 7SL RNA-derived SINEs were found in tree shrews [11,12]. All these SINE families include sequences corresponding to the terminal regions of 7SL RNA with the central 144–182 nucleotides deleted. *Alu* (~300 bp) is a dimer, apparently formed by the fusion of two similar but not identical monomers. Rare free left and right *Alu* monomers (FLAM and FRAM, respectively), as well as their precursor, fossil *Alu* monomer, have been found in human sequences [13–15]. The presence of *Alu* has been demonstrated in all tested primates including various prosimians [16–18]. *Alu* amplification was most active in early primate evolution, while the current *Alu* activity decreased 100-fold relative to that of 30–50 million years ago [2]. Most tree shrew 7SL RNA-derived SINEs are dimeric or trimeric and include a tRNA-derived monomer and one or two 7SL RNA-derived monomers [11,12], although monomeric 7SL RNA-derived SINEs also exist [12,19].

In contrast, murine or rat B1 (~140 bp) is a monomer. However, it has an internal 29-bp duplication, which prompted Labuda et al. to consider B1 as a quasi-dimer [20]. In addition, mouse/rat B1 has a 9-bp deletion in the central region of the element. Quentin [21] identified a few mouse and rat proto-B1 (pB1) sequences without the duplication, which were quite similar to FLAM. pB1 variants (pB1d7 and pB1d10) with a 7- or 10-bp deletion were also found in the genomes of these rodents [21]. Apparently, pB1/FLAM emerged from 7SL RNA in the common ancestor of primates and rodents. B1 elements

☆ Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. EF042308–EF042578.

* Corresponding author. Fax: +7 495 135 1405.
E-mail address: kramerov@eimb.ru (D.A. Kramerov).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.02.007

were studied in hamsters (family Cricetidae) [8,22], mice, and rats (Muridae) [23–25], particularly after the genomes of *Mus musculus* and *Rattus norvegicus* were sequenced [26,27]. However, rodents are a very large order, totaling over 30 families, and B1 SINEs remained underexplored in other rodent families. Zietkiewicz and Labuda [28] have found B1 in the genomes of chipmunk (Sciuridae) and guinea pig (Caviidae), but sequencing of PCR products allowed no detailed analysis of B1 structure. Later, we described two dimeric SINE families containing a monomer of ID, another rodent SINE [29], in addition to B1: (i) MEN in squirrels of the genera *Menetes* and *Callosciurus* [30] and (ii) B1-dID in squirrels (Sciuridae) and dormice (Gliridae) [31]. Interestingly, a 20-bp, rather than the 29-bp, internal duplication was identified in B1 monomers of squirrels and dormice. Lee et al. also reported B4, a composite B1-containing SINE, in the mouse genome [32].

Recently, we demonstrated the presence of 7SL RNA-derived SINEs in rodents of all 15 tested families as well as in primates and tree shrews but not in other mammalian orders using hybridization with mouse B1 probe [12]. Pilot cloning and sequencing of genomic DNA fragments confirmed the presence of B1 copies in representatives of various rodent families: jerboas, birch mice, squirrels, beavers, and guinea pigs [12].

In this work, we have sequenced more than 300 B1 copies cloned from 23 rodent species from 22 families. This large-scale B1 study covering the main rodent lineages allowed us to evaluate the diversity of B1 elements, to identify their taxon-specific properties, and to propose possible pathways of this SINE evolution in the context of rodent phylogeny.

Results

Number of B1 genomic copies in various rodent families

The number of B1 copies in the genomes of different rodents was evaluated by dot hybridization of their genomic DNA with a *M. musculus* B1 probe (Fig. 1). One can see hybridization signal in all tested rodents, while the signal intensity varied significantly. The number of B1 copies was determined from the signal intensity using the number of copies in *M. musculus* established by complete genome sequencing [26] as reference (Supplementary Table 1). Most families were represented in this experiment by a single species; however, Muridae (mice, rats, gerbils), Dipodidae (jerboas), and Sciuridae (squirrels) were represented by several genera. Note that the number of B1 copies varied within these families (2.0, 4.2, and 5.6, respectively).

Considerable progress has been made in the recent years in our understanding of the phylogenetic relationships between rodent families, although many issues remain unclear. Fig. 2 shows a plausible evolutionary tree compiled from several molecular phylogenies [33–38] (only the families studied here were included). The highest number of B1 copies (6.3×10^5–1.2×10^6) was found in all families of the Myodonta clade (Muridae, Cricetidae, Spalacidae, Rhizomyidae, Zapodidae,

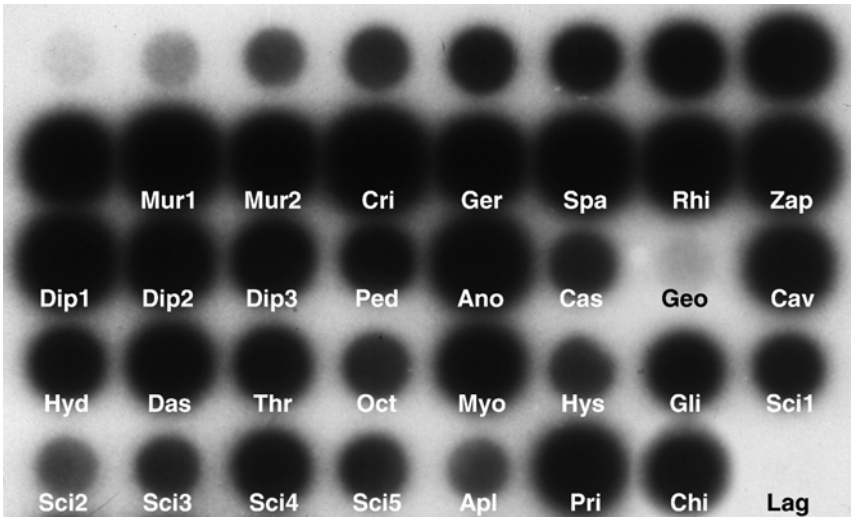
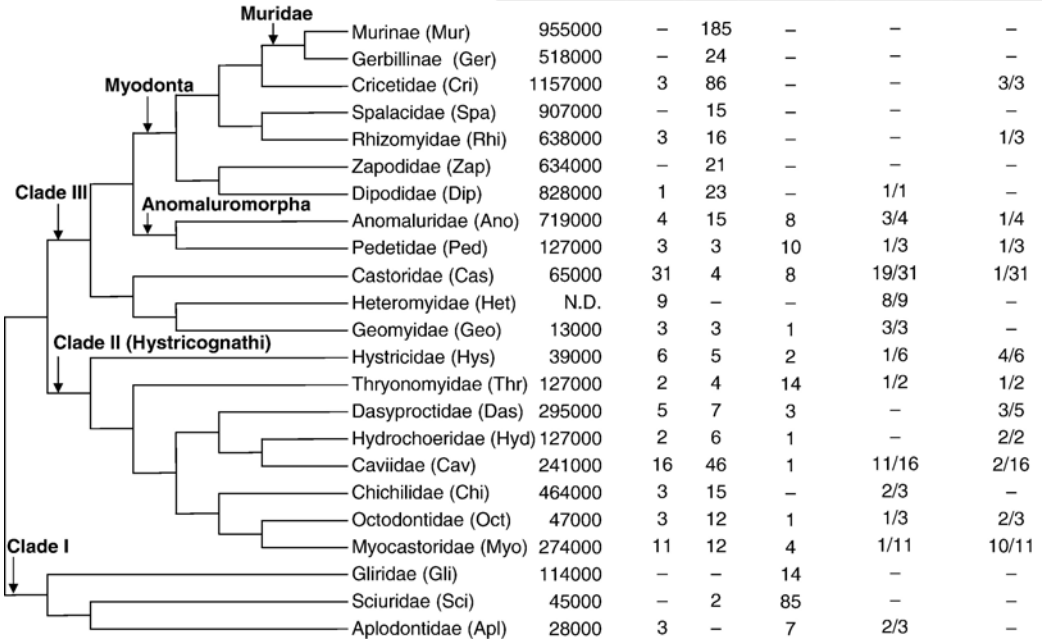


Fig. 1. Dot hybridization of rodent genomic DNA with mouse B1 probe. Murinae: Mur1, *Mus musculus* (house mouse); Mur2, *Rattus norvegicus* (Norway rat). Cricetidae: Cri, *Microtus socialis* (social vole). Gerbilinae: Ger, *Tatera indica* (Indian gerbil). Spalacidae: Spa, *Spalax microphthalmus* (Russian mole rat). Rhizomyidae: Rhi, *Tachyoryctes splendens* (East African mole rat). Zapodidae: Zap, *Sicista tianshanica* (Tien Shan birch mouse). Dipodidae: Dip1, *Eremodipus lichtensteini* (Lichtensten's jerboa); Dip2, *Allactodipus bobrinskii* (Bobrinski's jerboa); Dip3, *Alactagulus pygmaeu*s (lesser fivetoed jerboa). Pedetidae: Ped, *Pedetes capensis* (springhare). Anomaluridae: Ano, *Anomalurus* sp. (scaly-tailed flying squirrel). Castoridae: Cas, *Castor fiber* (Eurasian beaver). Geomyidae: Geo, *Thomomys bottae* (Botta's pocket gopher). Caviidae: Cav, C*avia porcellus* (guinea pig). Hydrochoeridae: Hyd, *Hydrochoerus hydrochaeris* (capybara). Dasyproctidae: Das, *Myoprocta acouchy* (acouchi). Thryonomyidae: Thr, *Thryonomys gregorianus* (lesser cane rat). Octodontidae: Oct, *Octodon degus* (degu). Myocastoridae: Myo, *Myocastor coypus* (nutria). Hystricidae: Hys, *Hystrix indica* (Indian crested porcupine). Gliridae: Gli, *Dryomys nitedula* (forest dormouse). Sciuridae: Sci1, *Marmota caudata* (long-tailed marmot); Sci2, *Spermophilus fulvus* (yellow ground squirrel); Sci3, *Sciurus carolinensis* (gray squirrel); Sci4, *Tamias asiaticus* (Siberian chipmunk); Sci5, *Menetes berdmorei* (Indochinese ground squirrel). Aplodontidae: Apl, *Aplodontia rufa* (mountain beaver). Primates: Pri, *Homo sapiens* (human). Chinchillidae: Chi, *Chinchilla laniger* (chinchilla). Lagomorpha: Lag, *Oryctolagus cuniculus* (rabbit). First 10 spots are twofold dilutions of mouse genomic DNA (from 1 to 500 ng). Human DNA (Pri) hybridization was due to *Alu* SINE.



	Number of copies in genome	Number of copies analyzed			pB1 copies with d10 or d7 / pB1 copies analyzed	
		pB1	B1	composite B1	d10	d7
Murinae (Mur)	955000	–	185	–	–	–
Gerbillinae (Ger)	518000	–	24	–	–	–
Cricetidae (Cri)	1157000	3	86	–	–	3/3
Spalacidae (Spa)	907000	–	15	–	–	–
Rhizomyidae (Rhi)	638000	3	16	–	–	1/3
Zapodidae (Zap)	634000	–	21	–	–	–
Dipodidae (Dip)	828000	1	23	–	1/1	–
Anomaluridae (Ano)	719000	4	15	8	3/4	1/4
Pedetidae (Ped)	127000	3	3	10	1/3	1/3
Castoridae (Cas)	65000	31	4	8	19/31	1/31
Heteromyidae (Het)	N.D.	9	–	–	8/9	–
Geomyidae (Geo)	13000	3	3	1	3/3	–
Hystricidae (Hys)	39000	6	5	2	1/6	4/6
Thryonomyidae (Thr)	127000	2	4	14	1/2	1/2
Dasyproctidae (Das)	295000	5	7	3	–	3/5
Hydrochoeridae (Hyd)	127000	2	6	1	–	2/2
Caviidae (Cav)	241000	16	46	1	11/16	2/16
Chichilidae (Chi)	464000	3	15	–	2/3	–
Octodontidae (Oct)	47000	3	12	1	1/3	2/3
Myocastoridae (Myo)	274000	11	12	4	1/11	10/11
Gliridae (Gli)	114000	–	–	14	–	–
Sciuridae (Sci)	45000	–	2	85	–	–
Aplodontidae (Apl)	28000	3	–	7	2/3	–

Fig. 2. Analysis of B1 sequences in rodent genomes. A tree of studied rodent families compiled from published molecular data is shown on the left. The family Muridae is represented by subfamilies Murinae and Gerbillinae. The number of copies was evaluated by dot hybridization of genomic DNA for rodent species used to make genomic libraries. Note that this procedure can underestimate the number of pB1 and composite SINEs. The following B1 types were recognized: pB1, the proper B1, and composite (dimeric) B1-containing SINEs. All B1 copies in Murinae and Caviidae as well as most copies in Cricetidae and Sciuridae were extracted from nucleotide sequence databases. N.D., not determined due to insufficient DNA quantities.

and Dipodidae) (Fig. 2). The Anomaluromorpha clade demonstrated significant variation in the number of copies: 7.2×10^5 and 1.3×10^5 in Anomaluridae and Pedetidae, respectively. The lowest number of copies (1.3×10^4) was observed in *Thomomys bottae* (Geomyidae). Most analyzed families of the clade Hystricognathi (Thryonomyidae, Dasyproctidae, Hydrochoeridae, Caviidae, Chinchillidae, and Myocastoridae) had a relatively high number of B1 copies (1.3×10^5–4.6×10^5), although it was as low as 3.9×10^4 and 4.7×10^4 in *Hystrix indica* (Hystricidae) and *Octodon degu* (Octodontidae), respectively. The genomes of squirrel-like rodents had a relatively low (Sciuridae and Aplodontidae) or moderate (Gliridae) number of B1 copies (Fig. 2). Note that even closely related families (Myocastoridae and Octodontidae) can differ considerably (sixfold) in the number of B1 copies, while most other related families demonstrated similar B1 abundance.

Structural variants of B1 elements in rodent families

Genomic libraries were constructed for 23 species representing 22 families of the order Rodentia, and the clones showing positive hybridization signals with labeled *M. musculus* B1 probe were sequenced. Overall, 305 B1 sequences were determined. In addition, 269 B1 copies were extracted from the genomic databases of analyzed rodent families. Most B1 nucleotide sequences of Cricetidae, Caviidae, and Sciuridae were obtained by database screening.

All identified nucleotide sequences of B1 could be divided into three types: (i) canonical B1 with an internal tandem duplication, (ii) pB1 lacking such application, and (iii) dimeric SINEs including B1 as one of two monomers. Fig. 2 demonstrates that the canonical B1 dominates over pB1 by the number of copies in most rodent families. However, Pedetidae, Geomyidae, Hystricidae, and Myocastoridae seem to have roughly equal proportions of these two B1 types. Moreover, the beaver (Castoridae) genome has eight times more pB1 than the variant with the duplication. None of nine B1 sequences in pocket mouse (Heteromyidae) represented the canonical B1, which at least points to the prevalence of pB1 over B1 in this rodent genome. Note that the families with minor canonical B1 have a basal position on the phylogenetic tree relative to most other rodent families. Gliridae, Sciuridae, and Aplodontidae seem to have an even more basal position and these rodents have prevalent composite SINE B1-dID: only two copies of monomeric B1 and three copies of monomeric pB1 were found in Sciuridae and Aplodontidae, respectively (Fig. 2).

pB1 structure in rodent families

While individual pB1 copies are highly diverged in all studied species, their consensus sequences are quite similar in different rodents (the median similarity for individual pB1 copies in some decently represented rodent families ranged

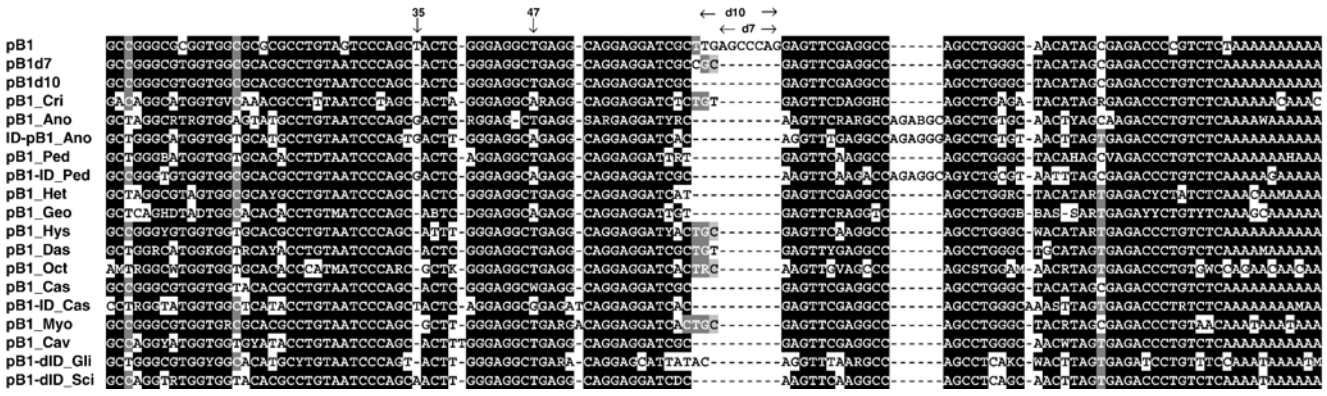


Fig. 3. Alignment of pB1 consensus sequences from different rodent families (designated by first three letters of the family, see Fig. 2). pB1 consensus sequences are shown for species with at least three copies sequenced. Composite SINEs (pB1-ID or ID-pB1) are given for five species. Upper three lines represent pB1 consensus sequences with or without specific deletions described by [21] in mouse and rat. Significant positions are marked above the alignment (pB1 numbering system).

from 55 to 67%, while it was 76% for pB1 consensus sequences; Fig. 3). Most observed differences between the consensus sequences correspond to hypervariable CpG sites (methylation targets). However, there are several other variable positions. 7SL RNA and pB1 have T at position 35, while this T35 is deleted in the dominant pB1 variants (pB1d7 and pB1d10 with a specific 7- or 10-bp deletion) described by Quentin [21] in mice and rats. This was also true for nearly all sequenced pB1 copies from other rodent families. However, some pB1d10 variants (largely from composite SINEs, ID-pB1_Ano, pB1-ID_Ped, and pB1-dID_Sci) had no deletion at position 35 (Fig. 3). Mouse and rat pB1 has T at position 47 [21], which was also observed in many rodent species, while others have T47 replaced with A (with no correlation to their position on the phylogenetic tree). Amazingly, three consensus sequences from anomalures and springhares (pB1_Ano, ID-pB1_Ano, and pB1-ID_Ped) shared a specific hexanucleotide insertion, AGAGGS, as well as three specific single-nucleotide substitutions. This similarity supports the recently proposed relationship between Anomaluridae and Pedetidae families demonstrated by other molecular approaches [33]. (These specific characters were found in none of four pB1_Ped sequences but can be expected as new copies become available).

The 7- and 10-bp deletions in pB1 deserve special consideration. Only 4 of 108 analyzed pB1 sequences from different rodent families had no such deletion. Many species had both deletions (Fig. 2), while others had predominant pB1d7 (nutria, Myocastoridae) or pB1d10 (guinea pig, Caviidae; beaver, Castoridae). It is of interest that pB1d10 predominated in basal families of clade III (see below):

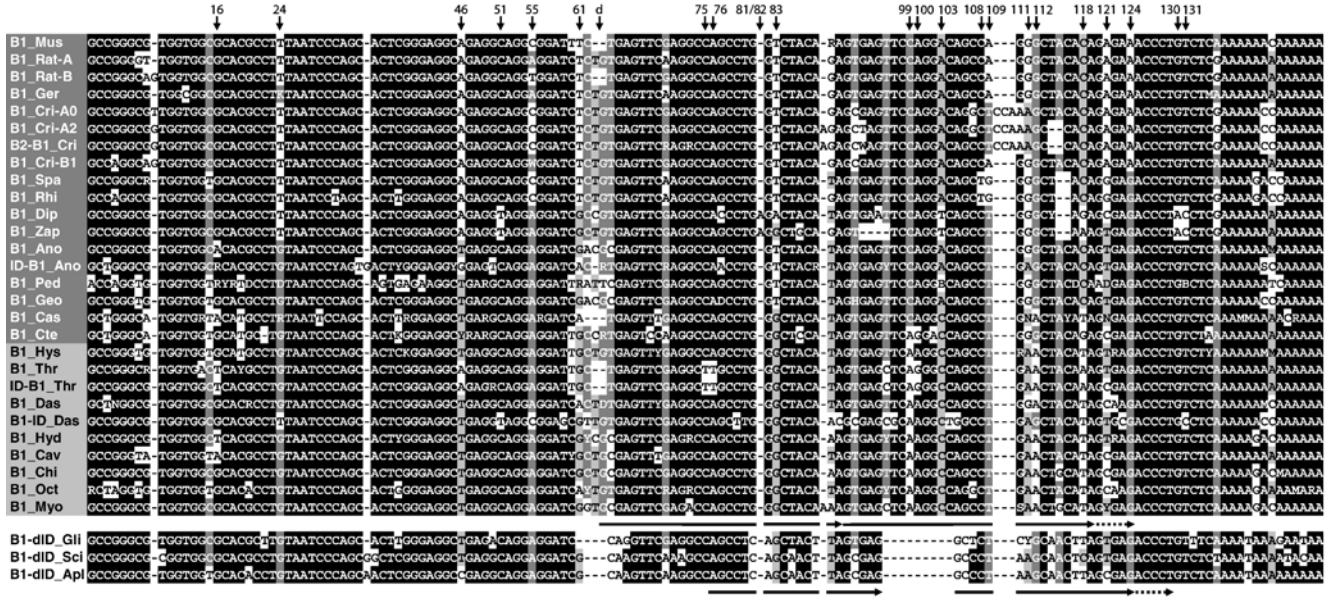


Fig. 4. Alignment of B1 consensus sequences from different rodent families (designated by first three letters of the family, see Fig. 2, except for the upper four sequences generated for the genomes of *M. musculus* (Mus), *R. norvegicus* (Rat-A and Rat-B), and subfamily Gerbillinae (Ger)). Sequence names of clades I, II, and III are shown against white, light gray, and dark gray, respectively. Significant characters are marked above the alignment (mouse B1 numbering system). The arrows below sequences indicate the 29- and 20-bp tandem duplications (dotted arrows show their redundant parts [12]). For other explanations, see text.

Castoridae, Heteromyidae, and Geomyidae (Fig. 2). Thus, replication efficiency of pB1d7 and pB1d10 could vary significantly between rodent families. While deletions and even insertions of lengths other than 7 or 10 nucleotides could be found; these were minor and likely resulted from later mutations (Supplementary Table 3).

B1 structure in different rodent families

In contrast to pB1, canonical B1 (with the internal duplication) often demonstrated specific structural features in different rodent families. Fig. 4 shows the alignment of family-specific consensus sequences of monomeric and composite B1 SINEs. The Muridae family is represented by four consensus sequences: B1_Mus (*M. musculus,* subfamily Murinae), B1_Rat-A and B1_Rat-B (*R. norvegicus,* Murinae), and B1_Ger (subfamily Gerbillinae). Four consensus sequences were also generated for the family Cricetidae: monomeric B1_Cri-A0, B1_Cri-A2, and B1_Cri-B1 and dimeric B2-B1_Cri.

Note that the variation between individual B1 copies was lower compared to pB1: the median similarity for individual B1 copies in some decently represented rodent families ranged from 62 to 86 and 81% for B1 consensus sequences; while it was 55–67 and 76% for pB1, respectively.

Comparison of canonical B1 sequences allowed us to identify certain positions distinguishing at least some of the consensus sequences. The analysis of the 31 significant characters is shown in Supplementary Table 2. Since nucleotides could vary in B1 copies from the same species (family), the proportions of the major variant are given (these values show character conservation within rodent species/families).

The distribution of B1 structural characters was analyzed in the context of the phylogenetic tree of rodent families (Fig. 5). The changes in these characters (duplications, indels, and nucleotide substitutions) were mapped onto the tree. Some of the character states (solid triangles) were congruent with this phylogenetic tree, while others (empty triangles) were not. For instance, the 29-bp internal duplication (direct repeat 29) is typical of B1 (B1DR29) in all rodent families except Gliridae, Sciuridae, and Aplodontidae (clade I), while the clade I B1 has the 20-bp duplication (B1DR20), suggesting that clade I emerged before the split of clades II and III. (Moreover, no B1DR20 has been found in clades II and III, and vice versa, no B1DR29 has been found in clade I.) Overall, 42 of 55 character states were synapomorphic and 13 were homoplastic.

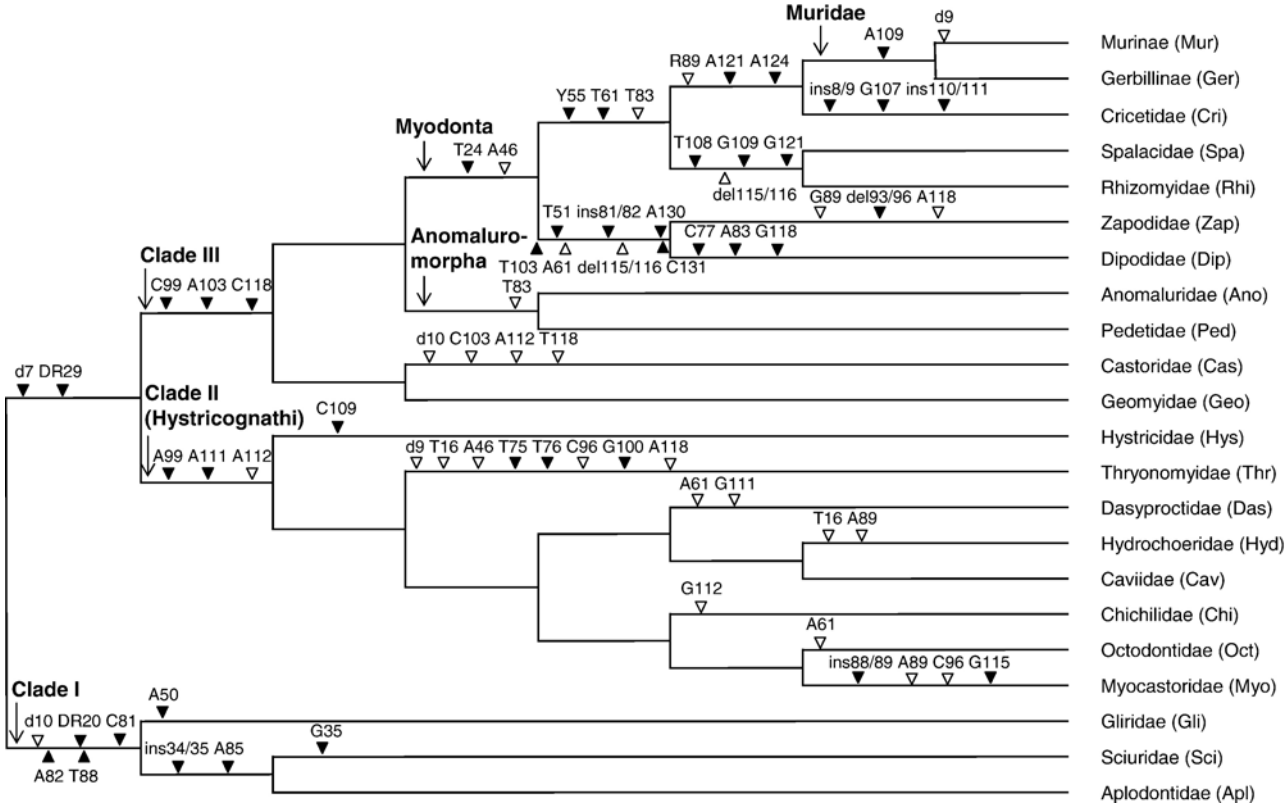


Fig. 5. Phylogenetic relationships among the studied rodent families and significant changes in their B1 sequences. Clades I, II, and III as well as the Myodonta clade are indicated on the tree by arrows, while the moments of significant duplications, substitutions, and indels in B1 are shown as triangles (mouse B1 numbering system). Only pronounced characters observed in >50% conspecific sequences were considered. In the case of Murinae and Cricetidae, in which three B1 subfamilies were recognized each, only characters pronounced in at least two subfamilies were considered. B1 sequences from composite SINEs were not considered except for clade I, in which monomeric SINEs are minor. In total, 38 characters and 55 character states are shown. Solid triangles indicate character states congruent with the tree of rodents (synapomorphic and those occurring in a single family only). Empty triangles indicate homoplastic character states. Character states inherited by B1 from pB1 are not shown. Overall, 42 and 13 character states are represented by solid and empty triangles, respectively.

Note a higher concentration of changes (triangles in Fig. 5) in Myodonta and Thryonomyidae. It is also of interest to trace the history of the d7/d10/d9 deletions. D7 appeared in the ancestor of clades II and III and B1d7 became predominant in these clades (with the only exception of Castoridae). More recently, d9 appeared in Thryonomyidae and in the ancestor of Muridae and Cricetidae (it actively amplified in Murinae but minor B1d9 can be found in Gerbillinae and Cricetidae as well).

Discussion

This work demonstrates the presence of B1 elements in rodents of all 22 tested families by their cloning and sequencing. In most cases, two types of sequences were obtained: (i) the proper B1 with a 29- or 20-bp tandem duplication and (ii) pB1 with no such duplication. pB1 was initially described in the mouse and rat genomes by Quentin as evolutionary precursors of the canonical B1 SINEs [21]. We failed to reveal pB1 in only two families, Spalacidae and Zapodidae, as well as in subfamily Gerbillinae (Fig. 2). This can be attributed to the predominance of the canonical B1 elements as well as to a higher divergence of pB1 sequences providing for a weaker hybridization signal. Overall, there are no grounds to question the presence of pB1 in all rodent genomes.

Similar to the mouse and rat genomes [21], most pB1 elements analyzed in this work had a specific 10- or 7-bp deletion, which points to a higher retropositional activity of pB1d10 and pB1d7 compared to the proper pB1.

Fig. 6 shows a plausible model of B1 evolution based on the original and published data. The emergence of pB1/FLAM from 7SL RNA was a unique event in the common ancestor of rodents, primates, and tree shrews [11,12]. Previously, the absence of 7SL RNA-derived SINEs in the genomes of other

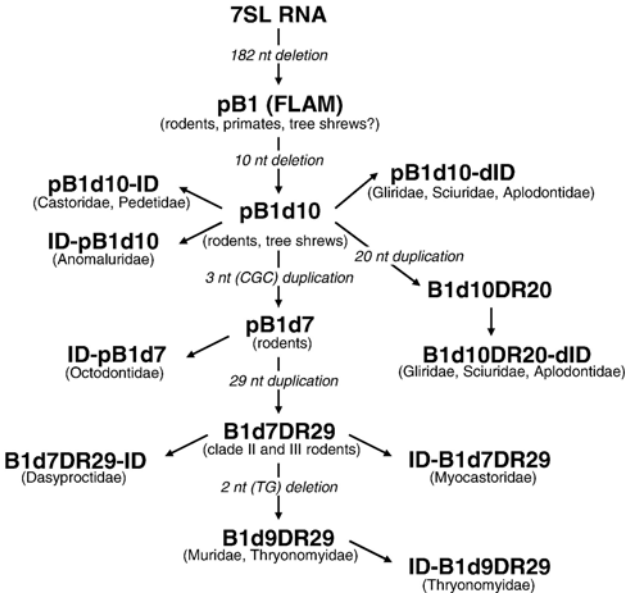


Fig. 6. Schematic evolution of B1 SINEs showing the emergence of specific deletions and duplications as well as composite ID-containing SINEs. The distribution range of B1 variants is shown in parentheses. See text for other explanations.

mammalian orders was demonstrated by hybridization [12]. In this work, we tried to find pB1 in the complete dog genome by computer search. None of 42 identified 7SL RNA pseudogenes had the pB1 182-bp internal deletion (or any internal deletion; these sequences were either full length or terminally truncated). This 182-bp deletion seems to be the primary event in the evolution of B1 and *Alu* SINEs.

The d10 deletion in the ancestors of rodents and tree shrews could be the next step. The precedence of pB1d10 relative to pB1d7 is supported by the finding of pB1d10 but not pB1d7 in the genomes of tree shrews [11,12] and clade I rodents (Gliridae, Sciuridae, and Aplodontidae) ([12]; this work). Moreover, 17 copies of pB1d10 have been recently identified in the human genome, suggesting that this variant appeared in the ancestor of Supraprimates but was not successful in primates, lagomorphs, and flying lemurs [19], so that 7SL RNA-derived SINEs are undetectable in rabbit and flying lemur [12,39]. It seems probable that pB1d7 emerged from pB1d10 as a result of a tandem duplication of trinucleotide CGC (see below).

Note that pB1 also includes sequences with, e.g., 8-, 9-, and 11-bp deletions; however, their number is less by far compared to pB1d7 and/or pB1d10 (Supplementary Table 3). Clearly, these minor variants could originate from pB1d7 or pB1d10 through additional single-nucleotide indels, although other scenarios can be proposed.

The obtained data indicate that most rodent genomes carry both pB1d7 and pB1d10, although one of them could predominate. For instance, pB1d10 clearly predominated in Caviidae, Geomyidae, Heteromyidae, and Castoridae. At the same time, nearly all B1 elements with the 29-bp duplication had the 7-bp deletion (or the 9-bp deletion in *M. musculus, R. norvegicus,* and *Thryonomys gregorianus*) irrespective of the dominant pB1 variant. It looks probable that the 29-bp duplication occurred in the common ancestor of clades II and III in a pB1d7 copy, which gave rise to a highly efficient B1 variant in most rodents of these clades. The derived B1DR29 with the 9-bp deletion has recently become efficient in some species. However, beaver (Castoridae) B1DR29 conflicts with this pattern since all (four) available copies have the 10-bp deletion. In clade I, a different tandem duplication (DR20) occurred in the same region of pB1d10.

Apart from tandem duplications, formation of dimeric SINEs from B1 and ID elements was a significant advancement in their evolution (Fig. 6). We have identified a number of such dimeric SINE families in rodent species (Figs. 3 and 4) differing by the monomer order (B1-ID or ID-B1) and specific additional sequences or deletions in the ID unit. (The structure and evolution of ID-derived SINEs will be described elsewhere.) Note that these dimeric SINEs could contain both B1 and pB1. The structure of the pB1/B1 unit in the dimers largely corresponded to the family-specific monomeric consensus sequence. In rare cases, some specific features appeared in the dimers; for instance, ID-B1 in Anomaluridae has the 8-bp deletion in contrast to monomeric B1d7.

Previously, we proposed [12] that the 29-bp duplication was mediated by a short direct repeat, GCGAG, which provided for a reverse transcriptase jump from the second repeat backward

to the first one [3]. This model was proposed for pB1d10 but can be extended to pB1d7, considering that it is a more likely precursor of B1DR29 (Fig. 6). As mentioned above, pB1d7 could arise from pB1d10 as a result of a tandem duplication of trinucleotide CGC (Fig. 7), which maintained the GCGAG repeat and allowed the 29-bp duplication mediated by reverse transcriptase slippage. Further single-nucleotide substitutions gave rise to different subfamilies of current B1 elements (Fig. 7). Note that the clade III variant changed more than the clade II B1.

While the 29-bp duplication was reported long ago [10,20], the emergence of a second B box of the RNA polymerase III remained unnoticed (Fig. 7). It is not improbable that an additional B box could contribute to the retropositional activity of B1DR29 considering that pB1 is not numerous in all rodents analyzed. At the same time, the low numbers of B1DR29 in Geomyidae and Castoridae suggest that the duplication alone is not sufficient to increase amplification efficiency of this SINE. Note that at least one more SINE (bat VES) with a second B box is known [40].

B1 sequences in some rodent families have more specific features (single-nucleotide substitutions and small indels). Apparently, the rate of B1 evolution was higher in such rodent lineages (Myodonta and Thryonomyidae) compared to others (other families of clade II) (Fig. 5). The differences observed between related families (e.g., Zapodidae and Dipodidae or Spalacidae and Rhizomyidae) indicate that most B1 copies were amplified after their divergence dated 45 and 23 Mya, respectively [41,42]. Apparently, all Cricetidae have B1 variants with CCAAA110–111 (although it is minor in Sigmodontinae). Clearly, these variants appeared in the Cricetidae ancestor before their split into subfamilies. On the other hand, Cricetidae has the subfamily B1_Cri-B1 highly similar to gerbil B1_Ger and rat B1_Rat-A (Fig. 4), which suggests that some B1 subfamilies can be maintained without much modification over long evolutionary periods—at least 25 My.

B1 with the 9-bp deletion exemplifies active amplification of a specific B1 subfamily, which took place at a relatively recent stage of rodent evolution. Although minor B1d9 copies can be found in different rodent families, considerable quantities are found only in *M. musculus, R. norvegicus,* and *Thr. gregorianus,* representing different rodent lineages. It is of interest that B1d9 copies are minor in Gerbillinae and Cricetidae and amount to one-third of all B1 in rat, while the bulk of mouse B1 have the 9-bp deletion.

Nevertheless, the mouse genome contains minor and highly diverged B1d7 known as subfamily B1F [21]. Analysis of its significant characters compared to those shown in Fig. 4 for other rodents allows us to date the emergence of B1F to the early divergence of clades II and III (data not shown). Clearly, this subfamily was not very efficient and became inactive long ago in the mouse lineage. In this context, it is of interest that the number of B1 copies in the basal rodent families in clade II (Hystricidae and Thryonomyidae) and clade III (Pedetidae, Castoridae, Heteromyidae, and Geomyidae) is relatively low, from 1.3×10^4 to 1.3×10^5 (Fig. 2). Apparently, the genomes of their ancestors contained a B1F-like subfamily, which neither became very successful nor gave rise to other B1 subfamilies with a higher retropositional activity.

While the number of B1/pB1 copies is not high in the genomes of pocket gophers (Geomyidae) and kangaroo mice

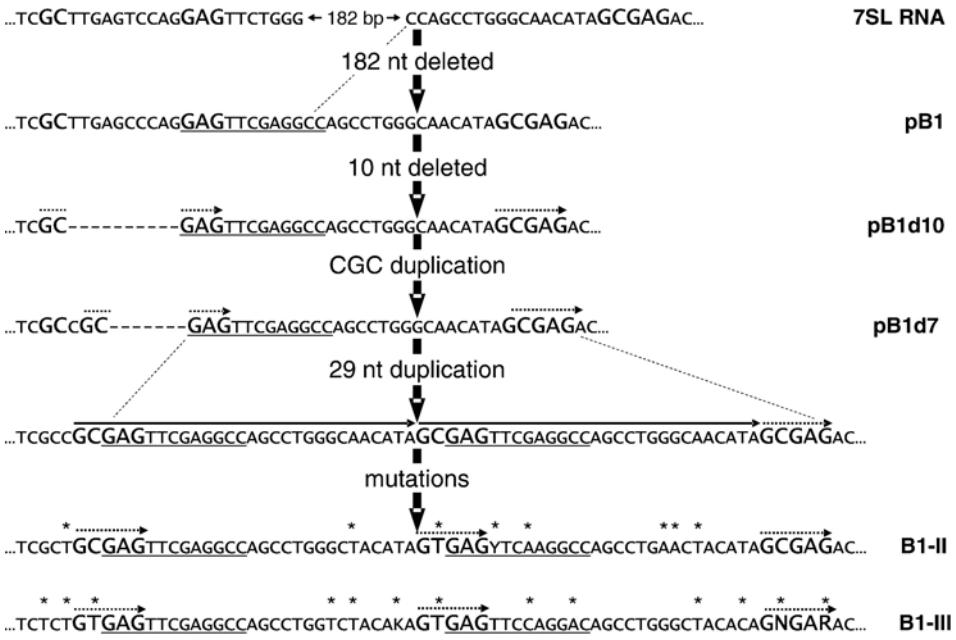


Fig. 7. Possible mechanism of evolution of B1 with the 29-bp tandem duplication and the 7-bp deletion. Pentanucleotide direct repeats (marked by dotted arrows) were formed from boldfaced nucleotides after a 10-bp deletion and a 3-bp duplication. A reverse transcriptase backward jump from one direct repeat to the other resulted in the 29-bp duplication (marked by solid arrows), later affected by mutations (indicated by asterisks) in the clade II (B1-II) and III (B1-III) sequences. The original and duplicated B boxes of the RNA polymerase III promoter are underlined.

(Heteromyidae), they have another highly abundant young SINE, IDL-Geo [43]. We assumed that other rodents with a relatively low number of B1/pB1 copies can also contain other highly abundant SINEs. However, large-scale screening of the beaver genomic library (over 4×10^4 clones) identified no interspersed repeats apart from L1, B1, and ID (also not numerous).

It looks like pB1 and old B1 variants were moderately active during early rodent evolution but gradually became inactive. In most lineages, they were replaced with more efficient SINEs, e.g., B1d9DR29 in mice, B1-dID in squirrels, or IDL-Geo in pocket gophers. A similar B1 inactivation was also reported in Sigmodontinae rodents, in which it correlated with the inactivation of the partner LINE L1 in nearly all cases [44]. In contrast, L1 remained functional in pocket gophers and drove active IDL-Geo amplification. Thus, low B1 activity was not due to L1 inactivation in this case. Finally, beavers are the only known rodent family that lost active SINEs relatively long ago. This prompts us to speculate that low activity of SINEs, a powerful factor of genome variation, can decrease species fitness under changing conditions, particularly considering that only two species of the once diverse Castoridae family survived.

A considerable progress has been recently made in resolving the relationships between rodent families, particularly involving the analysis of nuclear gene sequences [33–37]. The conclusions reached allowed us to present the phylogenetic relationships of the studied rodent families as a tree in Fig. 5. The topology of certain regions of the tree is not sufficiently supported. For instance, the position of the clade including Gliridae, Sciuridae, and Aplodontidae remains unclear. The presence of the 20-bp duplication in these families confirms their basal position relative to other families with the 29-bp duplication in B1 SINEs. Generally speaking, many of the B1 characters are synapomorphic and support the current phylogeny of rodents (Fig. 5). Still, there are homoplastic ones (although not too many—13 of 55 recognized character states), which complicate direct inference of host phylogeny based on the characters of SINE subfamilies. While this approach can become useful, more data are required to be conclusive.

Materials and methods

DNA samples

Animal, tissue, and DNA sources are shown in Supplementary Table 1. DNA was isolated from fresh, frozen, or ethanol-preserved tissues (liver, kidney, or muscle) by incubation with proteinase K followed by phenol/chloroform extraction. DNA was quantified by fluorometry using Hoechst dye 33258.

Library construction and screening

Rodent genomic DNA (1.5–5.0 μg) was digested with *Eco*RI and *Hin*dIII and separated by electrophoresis in 1% agarose gel. (*Eremodipus lichtensteini* DNA was digested with *Kpn*I and *Hin*dIII because the bulk of jerboa B1 sequences have an internal *Eco*RI site). DNA fragments of 0.5–1.2 kb were collected by reverse electrophoresis on a DEAE membrane inserted in the gel. DNA was eluted from the membrane in 400 μl of 1 M NaCl, 1 mM EDTA, and 10 mM Tris–HCl, pH 8.0, for 30 min at 60 °C. DNA was precipitated by ethanol using 10 μg glycogen as a carrier. The isolated genomic fragments (0.1–0.5 μg) were ligated into 0.1–0.3 μg pGEM3Z, digested with *Eco*RI and *Hin*dIII (or *Kpn*I and *Hin*dIII in the case of jerboa), and used to transform XL-1 Blue *Escherichia coli* cells. Colony hybridization was carried out at 60 °C in 4× SSC, 0.5% SDS, 5× Denhardt's solution, 0.1 mg/ml boiled herring sperm DNA, and ^{32}P-labeled murine B1 probe [12]. Nitrocellulose filters were washed in 0.1× SSC and 0.1% SDS at 42 °C, and positive colonies were identified by autoradiography. Colonies with both strong and weak hybridization signals were selected. B1-contaning *E. coli* clones were purified by two additional rounds of colony hybridization.

Interspersed repeats were screened in the *Castor fiber* genome as described above using ^{32}P-labeled total genomic DNA as a probe. In this case, only clones with high-copy-number sequences were selected. All positive clones were further purified by colony hybridization and used for sequencing.

In the case of *Ctenodactylus gundi* (Ctenodactylidae), only degraded DNA was available; so we isolated 150 to 250-bp fragments from 3% agarose gel and blunt-ended them with Klenow fragment. Then the 3′ ends were adenylated using *Taq* polymerase and dATP and ligated into pGEM-T (Promega). The resulting library was screened by hybridization with the B1 probe as described above. (Since only two of five sequences included full-length sequences, these data were considered preliminary and included only in Fig. 4.)

Dot-blot hybridization

Rodent genomic DNAs (500 ng) were incubated in 10 μl of 0.5 M NaOH for 1 h at 37 °C. After the incubation 20 volumes of 6× SSC, 6% formaldehyde, and 0.025 M NaH_2PO_4 were added, and the DNA was transferred to a Hybond N membrane. Hybridization and washing conditions as well as radioactive probe preparation were the same as in colony hybridization screening of the genomic libraries. The number of genomic B1 copies was evaluated from the signal intensity measured with a Cyclone phosphorimager (Packard). Mouse genomic DNA with the number of monomeric and composite B1 copies (9.55×10^5) known from the complete genome sequence [26] was used as a standard.

DNA sequencing and computer analysis

Cloned B1-containing DNA fragments were sequenced using standard M13 primers, the BigDye Terminator sequencing kit, and an ABI Prism 3100-Avant sequencer (Applied Biosystems).

The nucleotide sequences of cloned DNA fragments were deposited with GenBank under the following accession numbers: EF042308–EF042319 (*Myocastor coypus*), EF042320–EF042330 (*O. degus*), EF042331–EF042338 (*Chaetodipus californicus*), EF042339–EF042345 (*Tho. bottae*), EF042346–EF042366 (*Anomalurus* sp.), EF042367–EF042382 (*Pedetes capensis*), EF042383–EF042389 (*Microtus socialis*), EF042390–EF042399 (*Tatera indica*), EF042400–EF042409 (*Aplodontia rufa*), EF042410–EF042416 (*Sciurus carolinensis*), EF042417–EF042421 (*Dryomys nitedula*), EF042422–EF042434 (*Chinchilla lanigera*), EF042435–EF042443 (*Hydrochoerus hydrochaeris*), EF042444–EF042455 (*Myoprocta acouchy*), EF042456–EF042492 (*Ca. fiber*), EF042493–EF042497 (*Allactaga major*), EF042498–EF042509 (*Er. lichtensteini*), EF042510–EF042519 (*Sicista tianshanica*), EF042520–EF042530 (*H. indica*), EF042531–EF042549 (*Thr. gregorianus*), EF042550–EF042553 (*Ct. gundi*), EF042554–EF042568 (*Tachyoryctes splendens*), and EF042569–EF042578 (*Spalax microphthalmus*).

SINE sequences were extracted from nucleotide sequence databases using FASTA search (their accession numbers can be found in the alignments in supplementary figures). Multiple alignments were generated manually or using ClustalW and manual correction in GeneDoc [45]. Consensus sequences were also generated by GeneDoc from the alignments with manual replacement of the hypervariable sites CG/CA/TG with CG. All the alignments and consensus sequences are available in the supplementary materials or on request.

Acknowledgments

Thanks are due to François Catzeflis (Institut des Sciences de l'Evolution, Université Montpellier 2, France); Ron DeBry (University of Cincinnati, Cincinnati, OH, USA); Terry

Robinson (University of Stellenbosch, Matieland, South Africa); Georgy Shenbrot, Elena Ivanitskaya, and Leonid Lavrenchenko (Institute of Ecology and Evolution, Moscow); Sergey Popov (Moscow Zoo); Elena Lyapunova (Institute of Developmental Biology, Moscow); and Andrei Puzachenko (Institute of Geography, Moscow) for providing tissues or DNA. We also thank Marina Grachtchouk (University of Michigan, Ann Arbor, MI, USA) for help in some experiments. This study was supported by the Russian Foundation for Basic Research (05-04-49553 and 07-04-00462).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.02.007.

References

[1] N. Okada, SINEs, Curr. Opin. Genet. Dev. 1 (1991) 498–504.
[2] P.L. Deininger, M.A. Batzer, Mammalian retroelements, Genome Res. 12 (2002) 1455–1465.
[3] D. Kramerov, N. Vassetzky, Short retroposons in eukaryotic genomes, Int. Rev. Cytol. 247 (2005) 165–221.
[4] M. Kajikawa, N. Okada, LINEs mobilize SINEs in the eel through a shared 3′ sequence, Cell 111 (2002) 433–444.
[5] M. Dewannieux, C. Esnault, T. Heidmann, LINE-mediated retrotransposition of marked Alu sequences, Nat. Genet. 35 (2003) 41–48.
[6] D.A. Kramerov, A.A. Grigoryan, A.P. Ryskov, G.P. Georgiev, Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: evidence from DNA cloning experiments, Nucleic Acids Res. 6 (1979) 697–713.
[7] A.S. Krayev, et al., 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 (1980) 1201–1215.
[8] S.R. Haynes, T.P. Toomey, L. Leinwand, W.R. Jelinek, The Chinese hamster Alu-equivalent sequence: a conserved highly repetitious, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element, Mol. Cell. Biol. 1 (1981) 573–583.
[9] P.L. Deininger, D.J. Jolly, C.M. Rubin, T. Friedmann, C.W. Schmid, Base sequence studies of 300 nucleotide renatured repeated human DNA clones, J. Mol. Biol. 151 (1981) 17–33.
[10] E. Ullu, C. Tschudi, Alu sequences are processed 7SL RNA genes, Nature 312 (1984) 171–172.
[11] H. Nishihara, Y. Terai, N. Okada, Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins, Mol. Biol. Evol. 19 (2002) 1964–1972.
[12] N.S. Vassetzky, O.A. Ten, D.A. Kramerov, B1 and related SINEs in mammalian genomes, Gene 319 (2003) 149–160.
[13] J. Jurka, E. Zuckerkandl, Free left arms as precursor molecules in the evolution of Alu sequences, J. Mol. Evol. 33 (1991) 49–56.
[14] Y. Quentin, 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 (1992) 487–493.
[15] Y. Quentin, Emergence of master sequences in families of retroposons derived from 7sl RNA, Genetica 93 (1994) 203–215.
[16] G.R. Daniels, P.L. Deininger, A second major class of Alu family repeated DNA sequences in a primate genome, Nucleic Acids Res. 11 (1983) 7595–7610.
[17] E. Zietkiewicz, C. Richer, D. Sinnett, D. Labuda, Monophyletic origin of Alu elements in primates, J. Mol. Evol. 47 (1998) 172–182.
[18] C. Roos, J. Schmitz, H. Zischler, Primate jumping genes elucidate strepsirrhine phylogeny, Proc. Natl. Acad. Sci. USA 101 (2004) 10650–10654.
[19] J.O. Kriegs, G. Churakov, J. Jurka, J. Brosius, J. Schmitz, Evolutionary history of 7SL RNA-derived SINEs in Supraprimates, Trends Genet. 23 (2007), doi:10.1016/j.tig.2007.02.002.
[20] D. Labuda, D. Sinnett, C. Richer, J.M. Deragon, G. Striker, Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved, J. Mol. Evol. 32 (1991) 405–414.
[21] Y. Quentin, 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 (1994) 2222–2227.
[22] W. Bains, K. Temple-Smith, Similarity and divergence among rodent repetitive DNA sequences, J. Mol. Evol. 28 (1989) 191–199.
[23] A.S. Krayev, et al., Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing, Nucleic Acids Res. 10 (1982) 7461–7475.
[24] J.T. den Dunnen, J.G. Schoenmakers, Consensus sequences of the Rattus norvegicus B1- and B2 repeats, Nucleic Acids Res. 15 (1987) 2772.
[25] D.H. Kass, M.E. Raynor, T.M. Williams, Evolutionary history of B1 retroposons in the genus Mus, J. Mol. Evol. 51 (2000) 256–264.
[26] R.H. Waterston, et al., Initial sequencing and comparative analysis of the mouse genome, Nature 420 (2002) 520–562.
[27] R.A. Gibbs, et al., Genome sequence of the Brown Norway rat yields insights into mammalian evolution, Nature 428 (2004) 493–521.
[28] E. Zietkiewicz, D. Labuda, Mosaic evolution of rodent B1 elements, J. Mol. Evol. 42 (1996) 66–72.
[29] J. Kim, J.A. Martignetti, M.R. Shen, J. Brosius, P. Deininger, Rodent BC1 RNA gene as a master gene for ID element amplification, Proc. Natl. Acad. Sci. USA 91 (1994) 3607–3611.
[30] I.M. Serdobova, D.A. Kramerov, Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny, J. Mol. Evol. 46 (1998) 202–214.
[31] D.A. Kramerov, N.S. Vassetzky, Structure and origin of a novel dimeric retroposon B1-diD, J. Mol. Evol. 52 (2001) 137–143.
[32] I.Y. Lee, et al., Complete genomic sequence and analysis of the prion protein gene region from three mammalian species, Genome Res. 8 (1998) 1022–1037.
[33] D. Huchon, et al., Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes, Mol. Biol. Evol. 19 (2002) 1053–1065.
[34] S. Steppan, R. Adkins, J. Anderson, Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes, Syst. Biol. 53 (2004) 533–553.
[35] R.M. Adkins, A.H. Walton, R.L. Honeycutt, Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes, Mol. Phylogenet. Evol. 26 (2003) 409–420.
[36] D. Huchon, E.J. Douzery, From the Old World to the New World: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents, Mol. Phylogenet. Evol. 20 (2001) 238–251.
[37] J.C. Opazo, A molecular timescale for caviomorph rodents (Mammalia, Hystricognathi), Mol. Phylogenet. Evol. 37 (2005) 932–937.
[38] A. Farwick, et al., Automated scanning for phylogenetically informative transposed elements in rodents, Syst. Biol. 55 (2006) 936–948.
[39] J. Schmitz, M. Ohme, B. Suryobroto, H. Zischler, The colugo (Cynocephalus variegatus, Dermoptera): the primates' gliding sister? Mol. Biol. Evol. 19 (2002) 2308–2312.
[40] O.R. Borodulina, D.A. Kramerov, Wide distribution of short interspersed elements among eukaryotic genomes, FEBS Lett. 457 (1999) 409–413.
[41] M. Hugueney, P. Mein, A comment on the earliest Spalacinae (Rodentia, Muroidae), J. Mamm. Evol. 1 (1993) 215–223.
[42] R.L. Carrol, Vertebrate Paleontology and Evolution, Freeman, New York, 1988.
[43] K.P. Gogolevsky, D.A. Kramerov, Short interspersed elements (SINEs) of the Geomyoidea superfamily rodents, Gene 373 (2006) 67–74.
[44] T.A. Rinehart, R.A. Grahn, H.A. Wichman, SINE extinction preceded LINE extinction in sigmodontine rodents: implications for retrotranspositional dynamics and mechanisms, Cytogenet. Genome Res. 110 (2005) 416–425.
[45] K.B. Nicholas, H.B.J. Nicholas, GeneDoc: analysis and visualization of genetic variation, http://www.cris.com/~Ketchup/genedoc.shtml (1997).