B1 and related SINEs in mammalian genomes

Nikita S. Vassetzky, Oleg A. Ten, Dmitri A. Kramerov*

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

Received 18 March 2003; received in revised form 30 May 2003; accepted 2 July 2003

Received by S.M. Mirkin

Abstract

Although B1 and Alu were the first discovered Short Interspersed Elements (SINEs), the studies of these genomic repeats were mostly limited to mice and humans and little data on their presence in other animals were available. Here we report the presence of these SINEs in a wide range of rodents (in all 15 tested families) as well as primates and tree-shrews and their absence in other mammals. Distribution pattern of these SINEs in mammals supports close relationship between rodents and primates as well as tree-shrews. Sequence analysis of these elements, apparently descending from cellular 7SL RNA indicates their rearrangements such as dimerization (Alu), quasi-dimerization (B1), acquiring a tRNA-related unit (B1-dID), extended deletions, etc., preceding their active expansion in the genomes. The revealed common pattern of microenvironment of some rearrangement hot spots in SINEs (internal duplications and deletions) suggests involvement of short direct repeats in the mechanism of such rearrangements. This hypothesis allows us to explain short rearrangements in these and other short retroposons.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Short retroposon; SINE; Evolution; Phylogeny; Rodentia; Scandentia; Primates; Mammalia; Euarchontoglires; Tupaia

1. Introduction

Genomes of many eukaryotes carry repeated sequences 80–400 bp long 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. Sequences of SINEs feature certain variability (usually 5–35%) within a family. Commonly, each SINE family is specific for organisms of one or several families or orders. Most SINE families originate from various tRNAs. However, two SINE families that were discovered first—B1 of mice, rats, and hamsters (Krayev et al., 1980; Haynes et al., 1981) and Alu of humans (Deininger et al., 1981)—originate from 7SL RNA, a component of cytoplasmic ribonucleoprotein called signal recognition particle involved in translation of secreted proteins in all eukaryotes (Ullu and Tschudi, 1984). Both SINE families are composed of sequences corresponding to the terminal regions of 7SL RNA with the central 144–184 nucleotides deleted. Alu (~ 300 bp) is a dimer, apparently formed by fusion of two similar but not identical monomers. Rare free left and right monomers (FLAM and FRAM, respectively) as well as their presumed common precursor, a fossil Alu monomer (FAM), have been found among human sequences (Jurka and Zuckerkandl, 1991; Quentin, 1992b; Quentin, 1994b).

In contrast to Alu, murine B1 (~ 140 bp) is a monomer. However, there is an internal 29 bp duplication which allowed Labuda et al. (1991) to consider B1 as a quasi-dimer. In addition, murine B1 has a 9-bp deletion in the central region of the element. Quentin (1994b) revealed a few proto-B1 (pB1) sequences without the duplication in the mouse genome quite similar to FLAM. Apparently, both pB1 and FLAM descend from 7SL RNA after a yet unknown genetic event in the common ancestor of primates and rodents. In this context, it is important to establish the range of species with 7SL RNA-derived SINEs in their genomes. Until recently, the presence of Alu has been

Abbreviations: SINE, short interspersed elements; FAM, fossil Alu monomer; FRAM, free right Alu monomer; FLAM, free left Alu monomer.

* Corresponding author. Tel.: +7-95-135-9997; fax: +7-95-135-1405.
E-mail address: kramerov@eimb.ru (D.A. Kramerov).

0378-1119/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0378-1119(03)00805-9
demonstrated in all tested primates including various prosi-
mians (Daniels and Deininger, 1983; Zietkiewicz et al.,
1998). Rodents are a very large order totaling over 30
families; however, B1 elements are well documented in
mice and rats (Muridae) and closely related hamsters (Cri-
cetidae) (Krayev et al., 1982; den Dunnen and Schoen-
makers, 1987). In addition, Zietkiewicz and Labuda (1996)
demonstrated B1 in the genomes of chipmunk (Sciuridae)
and Guinea pig (Caviidae) by sequencing PCR products.
Here we report the presence of B1-related SINEs in the
genomes of additional rodent families as well as tree-shrews
and their absence in other mammals (naturally, except
primates). Analysis of the obtained and published B1
sequences provides an important insight into the evolution
of these SINEs as well as their hosts.

2. Materials and methods

2.1. DNA samples, cloning, hybridization, and sequencing

Tissues of mountain beaver (Aplodontia rufa) and scaly-
tailed squirrel (Anomalurus sp.) were kindly provided by Dr.
Catzeflis (Institut des Sciences de l'Evolution, Université
Montpellier II, France). Common tree-shrew (Tupaia glis)
tissue was a generous gift of Dr. Likhnova (Institute of
Ecology and Evolution, Moscow). Acouchi (Myoprocta
acouchy), capybara (Hydrochoerus capybara), lion-tailed
macaque (Macaca silenus), elephant shrew (Macroscelides
proboscideus), walrus (Odobenus rosmarus), and tree
kangaroo (Dendrolagus bennettianus) tissues were kindly
provided by S. Popov (Moscow Zoo). The sources of all
other samples analyzed (Table 1) were described elsewhere
(Borodulina and Kramerov, 2001; Gogolevskaya and Kra-
merov, 2002). DNA was isolated from fresh, frozen, or
ethanol-preserved tissues (liver, kidney, or muscle) by
incubation with proteinase K followed by phenol/chloro-
form extraction. Genomic libraries were constructed in
pGEM7f+ (Promega) digested with HindIII and EcoRI.
The libraries were screened by a labeled DNA fragment of
mouse B1 (pos. 16-135). This fragment was prepared by
PCR of mouse genomic DNA (0.1 ng) with primers
GCAYRCCTTTAATCCCAG and TGAGACAGGTTTTCTCTG.
After agarose gel purification, the fragment was PCR-labeled
with α[32P]dATP (Serdobova and Kram-
were carried out with this probe at 60 °C and the mem-
branes were washed at 42 °C (for details, see
Serdobova and

Table 1

<table>
<thead>
<tr>
<th>Order</th>
<th>Family</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rodentia</td>
<td>Myomorpha</td>
<td>Mus musculus, house mouse</td>
</tr>
<tr>
<td></td>
<td>Muridae</td>
<td>Apodemus peninsulae, Korean field mouse</td>
</tr>
<tr>
<td></td>
<td>Cricetidae</td>
<td>Microtus dauricus, mole</td>
</tr>
<tr>
<td></td>
<td>Spalacidae</td>
<td>Tatera indica, Indian gerbil</td>
</tr>
<tr>
<td></td>
<td>Zapodidae</td>
<td>Spalax microphthalmus, mole rat</td>
</tr>
<tr>
<td></td>
<td>Dipodidae</td>
<td>Allactaga major jerboa</td>
</tr>
<tr>
<td>Sciuromorpha</td>
<td>Geomyidae</td>
<td>Thymomys bottae, Botta's pocket gopher</td>
</tr>
<tr>
<td></td>
<td>Sciuridae</td>
<td>Marmota caudata, long-tailed marmot</td>
</tr>
<tr>
<td></td>
<td>Aplodontidae</td>
<td>Aplodontia rufa, mountain beaver</td>
</tr>
<tr>
<td></td>
<td>Gliridae</td>
<td>Dryomys nitedula, forest dormouse</td>
</tr>
<tr>
<td></td>
<td>Ctenodacylidace</td>
<td>Ctenodactylus gundi, gundi</td>
</tr>
<tr>
<td></td>
<td>Anomaluridae</td>
<td>Anomalurus sp., scaly-tailed squirrel</td>
</tr>
<tr>
<td></td>
<td>Castoridae</td>
<td>Castor fiber, beaver</td>
</tr>
<tr>
<td>Caviomorpha</td>
<td>Dasyproctidae</td>
<td>Myoprocta acouchy, acouchi</td>
</tr>
<tr>
<td></td>
<td>Caviidae</td>
<td>Cavia porcellus, Guinea pig</td>
</tr>
<tr>
<td>Primates</td>
<td>Hominidae</td>
<td>Homo sapiens, human</td>
</tr>
<tr>
<td>Macropodidae</td>
<td>Cercopithecidae</td>
<td>Macaca silenus, lion-tailed macaque</td>
</tr>
<tr>
<td>Insectivora</td>
<td>Talpidae</td>
<td>Mogera robusta, large Japanese mole</td>
</tr>
<tr>
<td></td>
<td>Soricinae</td>
<td>Crocidura suaveolens, scaly shrew</td>
</tr>
<tr>
<td></td>
<td>Erinaceidae</td>
<td>Erinaceus amurensis, Chinese hedgehog</td>
</tr>
<tr>
<td></td>
<td>Macroscelidae</td>
<td>Macroscelides proboscideus, elephant shrew</td>
</tr>
<tr>
<td></td>
<td>Chiroptera</td>
<td>Eonycteris spelaea, cave fruit bat</td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Myotis daubentoni, Daubentoon's bat</td>
</tr>
<tr>
<td>Lagomorpha</td>
<td>Felidae</td>
<td>Oryctolagus cuniculus, rabbit</td>
</tr>
<tr>
<td></td>
<td>Odobenidae</td>
<td>Felis catus, cat</td>
</tr>
<tr>
<td>Carnivora</td>
<td></td>
<td>Odobenus rosmarus, walrus</td>
</tr>
<tr>
<td>Scandentia</td>
<td></td>
<td>Tupaia glis, tree-shrew</td>
</tr>
<tr>
<td>Carnivora</td>
<td></td>
<td>Tursiops truncatus, bottle-nose dolphin</td>
</tr>
<tr>
<td>Artiodactyla</td>
<td></td>
<td>Bos taurus, cow</td>
</tr>
<tr>
<td>Marsupialia</td>
<td></td>
<td>Dendrolagus bennettianus, tree kangaroo</td>
</tr>
</tbody>
</table>
Kramerov, 1998). Double-stranded plasmid DNAs were sequenced using dideoxynucleotides and Sequenase 2.0 (Amersham) according to manufacturer’s instructions with modifications (Redston and Kern, 1994).

2.2. Computer-aided sequence analysis

Sequence similarity search in nucleotide banks was carried out using FASTA algorithm (Pearson and Lipman, 1988) with default parameters. Typically, we searched for the full-length B1 consensus without the tail. The alignments were generated manually using GeneDoc software (Nicholas and Nicholas, 1997). Consensus sequences were also produced by GeneDoc from the alignments of both full-length and partial elements with manual replacement of the hypervariable sites CG/CA/TG with CG. All the alignments as well as consensus sequences mentioned below are available on request.

The number of SINE copies in the genome was roughly estimated from the proportion of B1 elements in bank sequences for a given taxon using equation

\[ N = n \times 3 \times 10^9/L \]

where \( n \) is the number of found B1 in the bank sequences of total length \( L \); \( 3 \times 10^9 \) states for the length of haploid mammalian genome. In the case of actively sequenced mouse and rat genomes, random samples of bank sequences shorter than 20 kb were used to make comparable samples.

Unless stated otherwise, the numeration is based on the murine B1 consensus (Quentin, 1989).

2.3. Sequence accession numbers

The original nucleotide sequences were deposited in GenBank under the following accession numbers: birch mouse (Sicista tianshanica), AY247039 AY247040 AY247041 AY247042; great jerboa (Allactaga major), AY247034 AY247035 AY247036 AY247037 AY247038; long-tailed marmot (Marmota caudata), AY247045 AY247046; Berdmore’s squirrel (Menetes berdmorei), AY247043 AY247044; forest dormouse (Dryomys nitedula), AY247047; beaver (Castor fiber), AY247048 AY247049 AY247050; Guinea pig (Cavia porcellus), AY247051 AY247052 AY247053 AY247054 AY247055; and common tree-shrew (T. glis), AY247056 AY247057 AY247058 AY247059 AY247060 AY247061 AY247062. Accession numbers of other 7SL-related sequences found in the banks and mentioned below are available on request.

3. Results

3.1. B1 distribution in mammalian genomes

In order to establish the range of B1/Alu distribution in mammals, we carried out dot-hybridization of mouse B1 probe to genomic DNA of various species (Table 1). Strong hybridization signal was observed for representatives of all 15 tested rodent families. These families (representing roughly half of all rodent families) cover all three suborders (Myomorpha, Sciuromorpha, and Caviomorpha) and strongly indicate the presence of B1 or related repeats in all rodents. Note a less pronounced hybridization signal for squirrels and related families, which can point to a relatively low number of these SINEs in the genomes or their limited similarity with the probe. As expected, strong signal was also specific for human and lion-tailed macaque DNA due to cross-hybridization between B1 and Alu elements. No signal was observed for representatives of other nine mammalian orders except a weak signal for tree-shrew (Table 1). Dot-hybridization was negative for all tested DNAs from birds, fish, invertebrates, and plants (data not shown). According to the obtained data, abundant 7SL RNA-related retroposons appear specific only for primates and rodents, most likely, for all their numerous species.

Since dot-hybridization has limited sensitivity and can miss minor repeated sequences, we tried Southern hybridization for the species negative in the previous experiment. We used the same murine DNA as a probe as well as human and Guinea pig DNA (50 times less compared to the tested species) as a positive control. In addition to these, the tree-shrew sample showed strong hybridization signal smeared along the track, which is typical for interspersed genomic repeats (Fig. 1). No hybridization was observed for other species. These data demonstrate the presence of B1 or Alu-related sequences only in tree-shrews (Scandentia order) although they are less abundant than in primates and rodents.

3.2. Cloning and sequencing of 7SL RNA-related retroposons

In order to unambiguously confirm distribution of 7SL RNA-related retroposons in various rodent families as well as tree-shrew, we cloned and sequenced them. Genomic libraries of birch mouse (Zapodidae), great jerboa (Dipodidae), long-tailed marmot and Berdmore’s squirrel (Sciuridae), forest dormouse (Gliridae), beaver (Castoridae), Guinea pig (Caviidea), and common tree-shrew (Scandentia order) were generated. Their screening by murine B1 probe has revealed multiple positive clones; their sequencing demonstrates the presence of B1-related sequences in all tested species. We have sequenced five Guinea pig and jerboa elements each, four elements of squirrels and birch mice each, three beaver elements, and single dormouse elements. Sequence analysis of these copies from diverse rodent lineages confirms that they belong to B1 (or a related SINE family; see below). We also managed to obtain seven B1-related sequences from tree-shrew. All full-length elements had A-rich tails typical for B1 and Alu SINEs and many of them had target site duplications.
3.3. Search for 7SL RNA-related retroposons in sequence banks

Computer search for B1 reveals this element in rodent families both ample (mice (Muridae), hamsters (Cricetidae), squirrels (Sciuridae), and cavies (Caviidae)) and underrepresented in sequence banks: tuco-tucos (Ctenomyidae), four sequences; pocket gophers (Geomyidae), one sequence; kangaroo mice (Heteromyidae) and octodont rodents (Octodontidae), three sequences each. Ample 7SL-related sequences (actually, Alu) are found in the primate banks; in addition, one 7SL-related element have been found among the few available bank sequences of tree-shrews (Scandentia). At the same time, thorough search reveals no such elements in rabbits (Lagomorpha) and other mammalian orders as well as non-mammalian vertebrates.

3.4. Number of B1 copies in the genomes

The proportion of positive clones in the genomic libraries allows us to estimate the number of B1-related SINEs in the genomes of birch mouse, long-tailed marmot, Berdmore’s squirrel, forest dormouse, beaver, and Guinea pig within the range from 50,000 (Guinea pig) to 100,000 (dormouse) per haploid genome. An analogous estimation in the tree-shrew genome yields 500–1000 B1-related copies.

Our estimation of B1 copy number in the genomes based on their frequency in the sequence bank yields similar results for the rodents decently represented in sequence banks: 200,000 in hamsters, and 150,000 in cavies, rat, mouse, and squirrels (including the more abundant B1-dID; see below). The samples for other rodents were too small. Although these estimations are very rough, they correspond to the genome sequence data: over 500,000 copies in mouse (Mouse Genome Sequencing Consortium, 2002).

3.5. Structural variants of B1

Thorough analysis allowed Quentin (1994b) to recognize several ancestor B1 families: pB1 that differs from 7SL RNA by a large 184 bp deletion (pos. 74/75); and pB1d7, pB1d9, and pB1d10 that have one more deletion of 7, 9, and 10 bp, respectively (pos. 62–64). The canonical B1 has both these deletions and an additional 29-nt tandem duplication (Ullu and Tschudi, 1984) (Fig. 2). This most abundant variant of mouse B1 can be further divided into six subfamilies by several diagnostic positions (Quentin, 1989; Kass et al., 2000).

Since B1 from mouse and rat are among the best studied SINEs and due to enormous sequence data for these species, we concentrated on the elements from other rodent families. The sequences that we compared were very heterogeneous and we grouped them by large rearrangements such as the region of internal tandem duplication and significant single-nucleotide changes beyond the CpG hypervariable sites (Table 2).

Hamsters (Cricetidae). We divided the available 121 hamster B1 sequences into three main types; type B (45 sequences) was very similar to the canonical murine B1 by the 29 bp repeat; type A (51 sequences) featured additional three nucleotides (GCT) within the second repeat, while type C (25 sequences) had no duplication and corresponded to the pB1d7/d10. A subtype of the A variant can be recognized by two missing nucleotides (pos. 118–119 in type A consensus). The consensus sequences of these types are shown in Fig. 3.

The d7 variant predominated in the type A (only two d9 sequences vs. 41 explicit d7 ones); there were ten d9 sequences and six d10 sequences without considerable rearrangements in this region out of 40 B type elements; while the C type included eight d7 and d10 variants (data not shown). In addition, we observed much higher homogeneity of type A sequences than in types B and C (76%, 52%, and 43% median similarity, respectively) (Table 2).

Zietkiewicz and Labuda (1996) also revealed several subfamilies of B1 in hamster generally corresponding to...
our types; namely, several variants with modified murine internal duplication (h-B1, h-B2, and h-C ≈ Hamster-A) and murine internal duplication (h-A1 ≈ Hamster-B); other ones (h-A2 and h-D) were singular in our sample. Finally, there were considerable conflicts between the obtained and published consensus sequences at their ends (apparently, since Zietkiewicz and Labuda obtained them by PCR).

Birch mice (Zapodidae). All obtained sequences of B1 from birch mouse also had a murine-like internal duplication; however, there were one 4-bp and one 2-bp deletions (pos. 88–91 and 118–119) and one additional nucleotide (pos. 80/81) (Fig. 3). There were four d7 and one d9 variants; the median similarity of the birch mice elements was 80% (Table 2).

Jerboas (Dipodidae). Two out of five sequenced B1 of jerboa (Dipodidae) featured the mouse-like internal duplication; the other ones were truncated in this region. Only the d7 deletion as well as missing T35 could be found. Note the G80/81 insertion and deleted 118/119 dinucleotide that they share with the birch mice B1 (Fig. 3; Table 2).

Cavies (Caviidae). The sequenced and nucleotide bank B1s from Guinea pig could be divided into type I with the murine-like 29 bp duplication and type II without it (32 and 19 sequences, respectively). The type I included 24 d7, four d9, and three d10 variants, while type II included seven d7, eight d10, and even one pB1-like variant (no deletion). As with hamsters, the B1 sequences without the duplication were less homogeneous as compared to the type I (54% and 62% median similarities, respectively) (Table 2). The Guinea pig B1 consensus sequences are shown in Fig. 3.

Dormice (Gliridae) and squirrels (Sciuridae) stand apart from other rodents since their most abundant SINE is a dimer composed of 5′-B1 and 3′-ID units (Kramerov et al., 1999; Kramerov and Vassetzky, 2001). Moreover, the B1 part of the prevalent variant of this SINE (gsB1-dID or Gliridae-Sciuridae B1-dID) also carries a tandem repeat in the region of the murine duplication. However, the repeated unit is different: it is shorter (19 bp) and closer to the 3′-end of the element (Fig. 2). A similar internal duplication (gs-duplication hereafter) can be found in the B1 monomer of another composite element MEN from Berdmore's squirrel (Serdobova and Kramerov, 1998) (Fig. 3). We managed to clone several monomeric squirrel and dormouse B1 elements; three of them had the gs-duplication (“+ DR” hereafter), while other ones had no duplication in this region (“−DR”) similar to the alternative variant of B1-dID (Fig. 3). As in the case of B1-dID, most of the sequences had the d10 deletion; however, rare d9 variants also occurred. Note also that many (but not all) B1-related elements from squirrels and dormice preserved nucleotide 35 (pB1 numeration) deleted in B1s with the murine 29 bp duplication (Table 2).

Beavers (Castoridae). We united the three monomeric B1 sequences from beaver with three B1-dIDs since they shared many features; the consensus sequence is given in...
Fig. 3. The beaver B1 had no internal duplication; all sequences had d10 deletion. The main difference from pB1d10 was deleted T35 in the monomeric B1s (it was present in two beaver B1-dID sequences, while the third one had a large deletion in this region). The median similarity of beaver B1-related sequences was 58% (Table 2).

Other rodents. One B1 sequence of octodonts (Octodontidae) had an internal duplication similar to the murine one except for a large part of the second repeat missing (such rare sequences lacking a portion of the duplicated region can be found in other species, e.g., in Guinea pig (not shown) or pocket gopher); in addition, there was a d10 deletion (Fig. 3; Table 2).

The only full-size out of three obtained B1 sequences of kangaroo rats (Heteromyidae) had a gs-duplication and a d10 deletion; however, all these sequences preserved T35 (Fig. 3). The only available B1 sequence of pocket gopher also had T35 but lacked the internal duplication (Fig. 3; Table 2).

Only two out of four available tuco-tuco (Ctenomyidae) B1 sequences were full-size and one of them had a large deletion in the central region, which allows no general conclusions. There were neither internal duplication nor T35 in the sequences where these regions are available (Fig. 3; Table 2).

Tree-shrews (Scandentia). In addition to the cloned T. glis SINEs and a sequence found in the bank, a recent publication by Nishihara et al. (2002) presents SINEs in another tree-shrew species, T. belangeri. They revealed two types of 7SL-related SINEs: a dimeric Tu type I composed of a 5 V-tRNA-related unit and a 3 V-7SL-related one; and a trimeric Tu type II carrying an additional (shorter) 7SL-related unit at the 3 V-end. Indeed, one of our sequences (Tgl-91) belongs to type I, while another one (Tgl-12) as well as the one found in the sequence bank (Tgl164) belong to the type II. At the same time, all other sequences that we obtained (Tgl-05, Tgl-02, Tgl-13, Tgl-06, and Tgl-04) belong to neither of these types and seem to be monomeric.

We aligned all available 7SL-related sequences of tree-shrews (Fig. 4). One can recognize two major variants: variant A resembling pB1d10 (upper part of the alignment) and variant B with a 21 bp deletion (pos. 61–82; lower part of the alignment). Variant A is specific for Tu type I, the middle unit of Tu type II, and most monomeric SINEs, while variant B is specific for the 3 V-unit of the Tu type II and one monomeric SINE (Tgl-04). All available 7SL-related sequences from tree-shrews have no internal duplication and have only the d10 deletion (Fig. 4). Nucleotide 35 is present similar to the proto-B1 sequences. The median similarities of both variants of tree-shrew 7SL-related

### Table 2

<table>
<thead>
<tr>
<th>Character</th>
<th>Sequence</th>
<th>Internal duplication</th>
<th>Deletion</th>
<th>Pos. 24</th>
<th>Pos. 35</th>
<th>Pos. 80/81</th>
<th>Pos. 118–119</th>
<th>Pos. 124</th>
<th>Median similarity</th>
<th>No. of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>pB1, pB1d10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>G</td>
<td>T</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>nd</td>
<td>cons</td>
</tr>
<tr>
<td>Mouse</td>
<td>murine</td>
<td>d9, minor d7, d10</td>
<td>T</td>
<td>–</td>
<td>–</td>
<td>CA</td>
<td>A</td>
<td>nd</td>
<td>&gt;100</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>murine</td>
<td>d9, less d7, d10</td>
<td>T/G</td>
<td>–</td>
<td>–</td>
<td>CA</td>
<td>A</td>
<td>nd</td>
<td>&gt;100</td>
<td></td>
</tr>
<tr>
<td>Hamster A</td>
<td>murine, modified</td>
<td>d7</td>
<td>T</td>
<td>–</td>
<td>–</td>
<td>CA</td>
<td>A</td>
<td>76%</td>
<td>51(45)</td>
<td></td>
</tr>
<tr>
<td>Hamster B</td>
<td>murine</td>
<td>d9, d10</td>
<td>T</td>
<td>–</td>
<td>–</td>
<td>CA</td>
<td>A/G</td>
<td>52%</td>
<td>45(40)</td>
<td></td>
</tr>
<tr>
<td>Hamster C</td>
<td>–</td>
<td>d7, d10</td>
<td>T/R</td>
<td>–</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>43%</td>
<td>25(19)</td>
<td></td>
</tr>
<tr>
<td>Birch mouse</td>
<td>murine, modified</td>
<td>d7</td>
<td>T</td>
<td>–</td>
<td>G</td>
<td>–</td>
<td>G</td>
<td>80%</td>
<td>5(5)</td>
<td></td>
</tr>
<tr>
<td>Jerboa</td>
<td>murine, modified</td>
<td>d7</td>
<td>T</td>
<td>–</td>
<td>G</td>
<td>–</td>
<td>G</td>
<td>nd</td>
<td>5(2)</td>
<td></td>
</tr>
<tr>
<td>Guinea pig I</td>
<td>murine</td>
<td>d7, less d9, d10</td>
<td>G</td>
<td>–</td>
<td>–</td>
<td>CA</td>
<td>A</td>
<td>62%</td>
<td>32(28)</td>
<td></td>
</tr>
<tr>
<td>Guinea pig II</td>
<td>–</td>
<td>d7, d10</td>
<td>G/A</td>
<td>–</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>54%</td>
<td>19(19)</td>
<td></td>
</tr>
<tr>
<td>Tuco-tucos</td>
<td>–</td>
<td>d7 (?)</td>
<td>G</td>
<td>–</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>nd</td>
<td>4(2)</td>
<td></td>
</tr>
<tr>
<td>Octodont</td>
<td>murine (?)</td>
<td>d10</td>
<td>G/C</td>
<td>–</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>nd</td>
<td>3(1)</td>
<td></td>
</tr>
<tr>
<td>MENf (squirrel)</td>
<td>gs</td>
<td>d10</td>
<td>G</td>
<td>–</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>60%</td>
<td>6(6)</td>
<td></td>
</tr>
<tr>
<td>Squirrels/dormice B1 + DRg</td>
<td>gs</td>
<td>d10</td>
<td>G</td>
<td>–</td>
<td>–</td>
<td>YA</td>
<td>73%</td>
<td>19(19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kangaroo rats</td>
<td>gs, modified</td>
<td>d10</td>
<td>G</td>
<td>T</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>nd</td>
<td>3(1)</td>
<td></td>
</tr>
<tr>
<td>Squirrels/dormice B1–DRg</td>
<td>–</td>
<td>d10</td>
<td>G/A</td>
<td>G/T</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>66%</td>
<td>22(14)</td>
<td></td>
</tr>
<tr>
<td>Large porcupinesd</td>
<td>–</td>
<td>d10</td>
<td>na</td>
<td>T/C</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>59%</td>
<td>6(0)</td>
<td></td>
</tr>
<tr>
<td>Beaversd</td>
<td>–</td>
<td>d10</td>
<td>A/G</td>
<td>–</td>
<td>T</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>58%</td>
<td>6(6)</td>
</tr>
<tr>
<td>Pocket gopher</td>
<td>?</td>
<td>d10</td>
<td>G</td>
<td>T</td>
<td>–</td>
<td>na</td>
<td>?</td>
<td>nd</td>
<td>1(1)</td>
<td></td>
</tr>
<tr>
<td>Tree-shrews A</td>
<td>–</td>
<td>d10</td>
<td>G/A</td>
<td>A/N</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>58%</td>
<td>15(15)</td>
<td></td>
</tr>
<tr>
<td>Tree-shrews B</td>
<td>–</td>
<td>d = 21 nt</td>
<td>G</td>
<td>T/–</td>
<td>–</td>
<td>na</td>
<td>G</td>
<td>58%</td>
<td>17(17)</td>
<td></td>
</tr>
</tbody>
</table>

Numeration by murine B1 consensus; ‘—’ denotes absence (deletion); for other explanations see text.

- a Number of full-length sequences is given in parentheses.
- b Consensus sequences (Quentin, 1994b).
- c Not applicable since pos. 118/119 fall within missing internal duplication region.
- d Not determined if less than five full-size sequences were available. For dimeric elements, only the B1 region was considered.
- e Consensus sequence (Quentin, 1989).
- f Consensus sequence (Serdobova and Kramerov, 1998).
- g B1-dID (Kramerov and Vassetzky, 2001) and monomeric B1 elements.
- h B1-dID elements; (Kramerov and Vassetzky, 2001).
Fig. 3. Alignment of rodent and tree-shrew B1-related SINEs, consensus (with grayed names) and individual (not truncated) sequences. The diagnostic positions are marked by arrows above the alignment. The regions of the murine and gs-duplications are indicated by the above solid arrows (note short terminal redundancies shown on the right); dotted arrows indicate short direct repeats possibly associated with the deletions below (one of them is in the deleted region but is present in sequence(s) without the deletion above); gs indicates SINEs from Gliridae and Sciuridae; for other designations see Fig. 2 and text.
Fig. 4. Alignment of B1-related sequences from tree-shrews. Tgl sequences correspond to the sequenced T. glis elements (except the bank sequence Tgl164, AF339164), while Tbe sequences of T. belangeri were obtained by (Nishihara et al., 2002). The numbers indicate the original clone nos., while ‘R’ and ‘L’ specify the right and left 7SL-derived units of Tu type II. The terminal direct repeats possibly associated with the integration are shaded.
sequences are 58%. The consensus sequences of the two tree-shrew 7SL-related sequences are given in Fig. 3.

4. Discussion

4.1. Evolution of mammalian hosts

Analysis of distribution of SINEs and other retroelements finds increasing use in phylogenetic studies of vertebrates (Martignetti and Brosius, 1993; Shimamura et al., 1997; Stoneking et al., 1997; Serdobova and Kramerov, 1998; Verneau et al., 1998; Kramerov et al., 1999; Kuryshkin et al., 2001; Schmitz et al., 2002; Vassetzky and Kramerov, 2002). Here we demonstrate the presence of numerous B1-related SINEs not only in mice, rats, and hamsters, but in all tested rodents of 15 families covering the main rodent lineages. Most likely, B1 is specific for all rodents. In addition to rodents (B1), 7SL-related sequences were detected in primates (Alu) and tree-shrews but not in other tested mammals or in other animals.

Until recently, neither morphological nor molecular data confirmed that rodents and primates are closer to each other than to other mammals (Carrol, 1988; Reyes et al., 2000; Liu et al., 2001). However, a recent publication (Murphy et al., 2001) based on extensive sequence data of nuclear genes in representatives of all mammalian orders clustered rodents, primates, and tree-shrews (as well as rabbits and flying lemurs (Dermoptera)) into a superorder called Euarchontoglires (Fig. 5). A similar cluster called Supraprimates was revealed by advanced analysis of mitochondrial DNA sequences (Waddell et al., 2001). Fig. 5 presents the tree of mammals by Murphy et al. (2001) complemented with the data on 7SL-derived SINE distribution.

Most likely a molecular event that resulted in emergence of 7SL-derived SINE occurred only once during mammalian evolution; in this case, the observed pattern of its distribution in mammalian genomes points to the common origin of rodents, primates, and tree-shrews. Thus, finding 7SL-related SINEs in rodents, primates, and tree-shrews but not in other tested mammals confirms monophyly of these tree orders. These data confirm the concept of Euarchontoglires with one exception. While the presence of 7SL-related SINEs in flying lemurs could still be expected, their existence in rabbits is challengeable. The hybridization data as well as thorough search in the available bank sequences revealed no such elements in rabbits (for comparison, we found over 50 B1s in a 10 times smaller sample of the available Guinea pig sequences). In this case, the Lagomorpha branch becomes an outgroup to other Euarchontoglires.

Still, independent appearance of 7SL-derived SINEs in more than one mammalian lineage cannot be excluded. However, Nishihara et al. (2002) also considered these possibilities and, based on the distribution of SINEs in the context of their structure, proposed that common origin of the 7SL-derived SINE is more plausible.
Initially, tree-shrews were assigned to insectivores; later they were long considered as a primate family. Then tree-shrews were recognized as a separate Scandentia order on the basis of morphological traits (Butler, 1972). Recent revisions based on the molecular data bring them together with rabbits (Schmitz et al., 2000) or primates and flying lemurs (Murphy et al., 2001). Finding 7SL-related SINEs in *Tupaia belangeri* (Nishihara et al., 2002) and *T. glis* (this paper) substantiates that tree-shrews belong to Euarchontoglires. Moreover, our finding of a monomeric tree-shrew 7SL-derived SINE quite similar to pB1d10 common in rodents but, apparently, missing in primates (at least, in human) further corroborates the inference of Nishihara et al. (2002) on a closer relation of tree-shrews to rodents rather than to primates.

### 4.2. Relations between B1-related SINEs

Certain correlations between the characters presented in Table 2 can be noted. For instance, the murine internal duplication occurs with T24, while purine in this position is specific for the elements with gs-duplication or with no duplication. The A124 is specific for the mouse-related rodents with the duplication (hamster (A and B), rat, and mouse). Two characters, G80 and dinucleotide deletion at 118/119 are found only in birch mice and jerboas (among the variants with the internal duplications, since the deletion falls within the duplicated region), which is not surprising considering close relation between these two rodent families (Carrol, 1988; Serdobova and Kramerov, 1998). Generally, the elements with the internal duplications are more homogeneous than those without it (Table 2) confirming evolutionary recency of this event. T35 is a hallmark of ancestral B1 and the squirrel-related lineage; apparently, it was deleted at the early stages of B1 evolution in the other rodents. At the same time, other characters show mosaic distribution; e.g., d7 and d10 deletions can be found in B1s both with and without the internal duplication (Table 2).

We cannot be positive if the variant with the internal duplication in birch mice is a secondary modification of the murine duplication (with two deletions and one insertion) or another (third) duplication variant (actually, a degenerated 25-bp tandem repeat GAGTGNAGGC-CAGCTGGRCTRMA can be proposed). Draft search for such variants revealed many rat B1 sequences with the same 4-bp deletion but without the two other distinctions (data not shown), while those (but not the 4-bp deletion) could be found in the jerboa sequences. Existence of such “intermediates” favors the first hypothesis.

A similar situation is specific for the octodont B1 with the murine-like duplication. The second repeat unit is shorter than the first one (Fig. 3). However, similar variants were found among the Guinea pig elements (data not shown).

The presence of 7SL-related SINEs (particularly, the monomeric ones) in tree-shrews confirms the common origin of B1 and Alu (from 7SL RNA). Assuming a common ancestor of tree-shrews, rodents, and primates, we can speculate that the proto-B1/monomeric Alu SINEs were already active in this ancestor (although not as successful as their dimeric and quasi-dimeric descendants). This agrees with similar rough estimates of B1-related SINE numbers in the genome of tree-shrew (1000) and B1 without the internal duplication in mouse genome (5000).

### 4.3. Possible mechanism of SINE variability

While analyzing the internal duplication region, we noticed a redundancy of both mouse- and gs-type duplications in pro-B1 sequences: the repeated region is followed by the first five nucleotides of it (GYGAG and ASCCT, respectively; Fig. 2) suggesting their involvement in the duplication mechanism.

Possibly, a similar “natural” tandem duplication was observed after a plasmid transfection and genomic integration in mouse cells (van Rijk et al., 1999). Although the region of duplication was much longer (~1.8 kb) than in our case, this process was apparently triggered by short homologies CCCAT flanking the region. As an alternative to such DNA-mediated rearrangements, this process can be associated with RNA and the retroposition process, thus, introducing additional variability in short retroposons.

In any case, the duplications could be enabled by appearance of direct repeats. For the 29 bp duplication, such repeats (missing in both 7SL RNA and pB1) appeared after the d7 or d10 deletions. Fig. 6 presents a possible scenario of such duplication for the d10 variant. After the deletion of 10 bp in pB1, a perfect repeat GCGAG (missing in the ascending sequences) is formed, thus, enabling the duplication. Later mutations degenerated the repeats. Note that this short repeat also appears in the d7 variant (Fig. 2) and a similar scheme can be proposed. At the same time, no such repeats can be found in the best known murine d9 variant, which confirms that d9 is secondary relative to d7 and d10 as proposed by Quentin (1994a).

Moreover, short direct repeats flank the d7/d10 deletions (at least for pB1d7, pB1d10, B1 Hamster-B, and gsB1 with these deletions; dotted arrows in Fig. 2). It is possible that we do not see such repeats in all cases since the actual sequences where these rearrangements occurred are unavailable; still we see perfect repeats for the deletions in FRAM-A and *Galago* type II SINEs when compared to 7SL RNA (dotted arrows in Fig. 2) as well as in the variant B of 7SL-related region in tree-shrew SINEs (dotted arrows in Fig. 3). A possible series of events giving rise to the d7 and d10 variants is presented in Fig. 7: a mutation in pB1 led to the appearance of the
direct repeat CGC in a hypothetical intermediate, and this repeat mediated the 10 (or 7) bp deletion, giving rise to pB1d10 (or pB1d7).

Existence of such mechanisms for duplication and/or deletion makes possible independent emergence of the same mutations in the SINE copies with the same short repeats. This can explain “mosaic evolution” of short retroposons first noted by Zietkiewicz and Labuda (1996). For instance, there can be no correlation between d7/d9/d10 deletion and the presence or absence of the internal duplication. Bains and Temple-Smith (1989) also noted non species-specific clustering of B1 sequences.

Alternatively, such mosaic can be explained by a pool of variants that preexisted in the genome, so that some of them can be independently realized later in various rodent families. This may be true for some but not all of them; for instance, we failed to find the elements with gs-duplication in the rat (data not shown).

Thus, the 7SL-related SINEs proved to be more diverse than previously considered. Further comparative analysis of this diversity can bring light to essential features of short retroposons as well as to the evolution of their hosts.

Acknowledgements

We are grateful to all researchers who provided animal tissues and DNA and to Dr. Youri Kravatsky for the help with computers. This work was supported by the Russian Foundation for Basic Research (projects 02-04-48644 and 02-04-49772).

References


