

Phylogeny of the Order Rodentia Inferred from Structural Analysis of Short Retroposon B1

N. A. Veniaminova^a, N. S. Vassetzky^a, L. A. Lavrenchenko^b,
S. V. Popov^c, and D. A. Kramerov^a

^a Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia;
e-mail: kramerov@eimb.ru

^b Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, 119991 Russia

^c Moscow Zoo, Moscow 123242 Russia

Received February 28, 2006

Abstract—A large-scale study of short retroposon (SINE) B1 has been conducted in the genome of rodents from most of the known families of this mammalian order. The B1 nucleotide sequences of rodents from different families exhibited a number of characteristic features including substitutions, deletions, and tandem duplications. Comparing the distribution of these features among the rodent families, the currently discussed phylogenetic relationships were tested. The results of analysis indicated (1) an early divergence of Sciuridae and related families (Aplodontidae and Gliridae) from the other rodents; (2) a possible subsequent divergence of beavers (Castoridae); (3) a monophyletic origin of the group Hystricognathi, which includes several families, such as porcupines (Hystricidae) and guinea pigs (Caviidae); (4) a possible monophyletic origin of the group formed by the remaining families, including six families of mouse-like rodents (Myodonta). Various approaches to the use of short retroposons for phylogenetic studies are discussed.

DOI: 10.1134/S1022795407070071

INTRODUCTION

Short retroposons, or SINEs (Short Interspersed Elements) are transposable genetic elements 80–400 bp in size that reproduce in the genome via reverse transcription of their RNA [1–3]. As the reverse transcriptase, short retroposons use polypeptides encoded in a long retroposon, or LINE (Long Interspersed Element).

The genomes of mammals of one species have two to four short retroposon families, each of which contain 10^4 – 10^5 copies, whose sequences typically show 65–90% similarity. Among-copy differences are explained by the existence of short retroposon subfamilies and by mutations appearing after integration of each copy. The number of such nucleotide substitutions is proportional to the age of the SINE copy.

Most short retroposons have originated from tRNA molecules. However, there are two classes of SINEs whose origin is related to other RNAs, synthesized by RNA polymerase III. One of them includes several recently discovered fish SINE families, which were derived from 5S rRNA [4, 5]. Short retroposon of the other class (Alu of primates and B1 of rodents) were among the first described SINEs [6–9]. They were derived from 7SL RNA—cytoplasmic RNA of 300 nucleotides in length, which is contained in ribonucleoprotein particles (SRP, signal recognition particles), which recognize the signal peptide of secreted and membrane proteins [10]. In the process of the Alu and B1 appearance,

the 144–182-nucleotide central part of 7SL RNA was lost (Fig. 1). Alu (300 bp) consists of two such similar, but not identical sequences: left (L) and right (R) monomers. Alu was found in the genomes of all primates examined in this respect, including Prosimiae, which testifies to the appearance of Alu in the common ancestor of all primates [11–13].

In contrast to Alu, the B1 element of the genomes of house mouse *Mus musculus* and Norwegian rat *Rattus norvegicus* is a monomer of about 140 bp in length. However, this element has an internal tandem duplication with a 29-bp repeated unit, which may be regarded as a sort of dimerization [14]. Another difference between Alu and B1 is the presence of a 9-bp deletion in the central part of murine B1. The genomes of these rodents were shown to harbor a small number of B1 copies lacking these duplication and deletion. These copies, termed pB1 (proto-B1) are thought to be evolutionary precursors of B1 [15]. In the mouse and rat genomes, pB1 variants with a specific deletion of 7, 10, or, less frequently, 9 bp were found (pB1d7, pB1d10, and pB1d9, respectively; see Fig. 1).

Rodents (Rodentia) are the largest mammalian order consisting of at least 30 families [16, 17]. B1 in mice and rats (family Muridae) was fairly extensively studied [18–20], in particular, owing to sequencing the *M. musculus* [21] and *R. norvegicus* [22] genomes, but this SINE was scarcely studied in the genomes of other rodents. Zietkiewicz et al. [12] have found B1 in chipmunk (Sciuridae) and guinea pig (Caviidae), but using

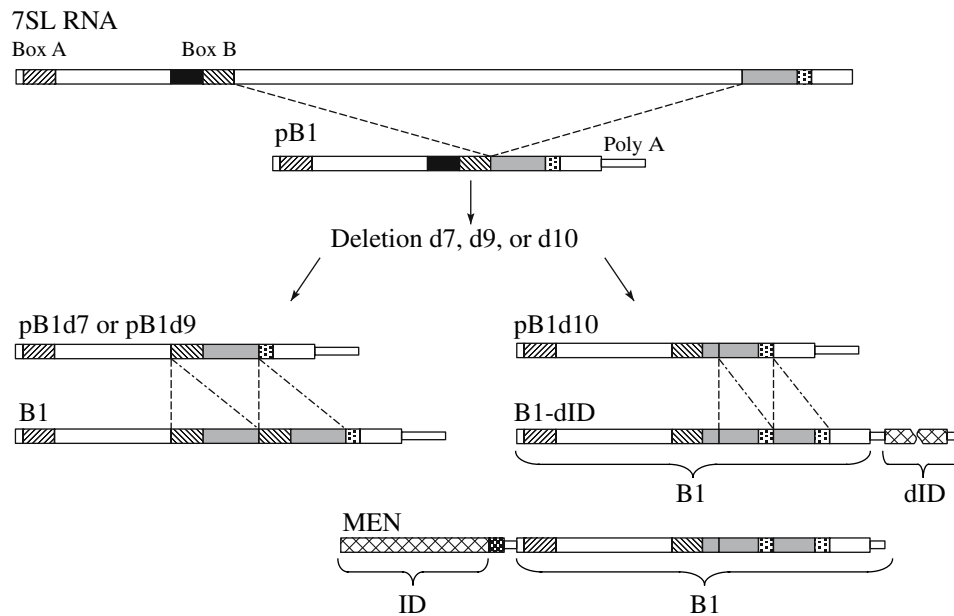


Fig. 1. Schematic diagram illustrating the origin, evolution, and structure of element B1 and related SINEs in rodents. Retroseudogene 7SL RNA carrying a 182-bp deletion appeared in the common ancestor of rodents, primates, and tree shrews, which yielded element pB1. A subsequent loss of a specific region of 7, 9, or 10 bp in size generated novel pB1 families. A tandem duplication of a 29-bp region produced the canonical B1 element known in the *Mus musculus* and *Rattus norvegicus* genomes. The formation of a 20-bp tandem duplication in the pB1d10 element preceded the appearance of dimeric SINEs B1-dID and MEN. ID is the tRNA-related SINE that formed these dimeric SINEs through fusion with B1. A 19-bp deletion is characteristic of the dID sequences. Boxes A and B of RNA polymerase III promoter are hatched. Regions of tandem duplications are shown in gray and point shading.

sequencing of the PCR products these authors could not examine the total nucleotide sequence of this SINE. We have described two dimeric SINEs having as monomers B1 and ID-related element (ID is known as a rodent SINE derived from alanine tRNA [23]). The first of them, MEN, isolated from the squirrel *Menetes berdmorei* genome, had the left ID-like monomer and B1 as the right monomer [24]. The second SINE, detected in the genomes of squirrels (Sciuridae) and dormice (Gliridae) had the opposite monomer location and was termed B1-dID [25]. Note that the B1 monomer of squirrels and dormice contained an internal duplication of 20, rather than 29 bp, as the B1 of mice and rats (Fig. 1).

Recently, using hybridization with the *M. musculus* B1 probe, we have shown the presence of 7SL RNA-related SINEs not only in primates (Alu), but also in rodents of all families examined [26] as well as in tree shrews (Scandentia) [26, 27], but not in members of other mammalian orders. Preliminary experiments on cloning and sequencing genomic DNA fragments confirmed the presence of B1 copies in such rodents as jerboas, birch mice, squirrels, beavers, and guinea pigs [26].

It was shown that SINEs could be used as reliable markers for studying phylogenetic relationships. In particular, one can establish the presence of a certain SINE families in the genome of the organisms studied [28, 29]. The presence of a shared SINE family in the

genomes of different organisms suggests that these organisms are related.

In the present work, we have used for phylogenetic analysis SINE (B1) subfamilies rather than families. Many B1 copies of rodents from 22 families were sequenced, which revealed a number of B1 subfamilies (variants), differing by specific nucleotide substitutions, duplications, and deletions. Based on distribution analysis of such B1 variants in rodents from different families, we have tested the current schemes of phylogenetic relationships within this mammalian order.

MATERIALS AND METHODS

DNA samples. Sources of animals, their tissues or DNA are listed in the table. DNA was isolated from fresh, frozen, or fixed in ethanol tissues (kidney, liver, or muscle) by incubation of the homogenate with proteinase K followed by phenol-chloroform (1 : 1) deproteinization.

Dot hybridization. Genomic DNA (500 ng) was incubated in 10 μ l of 0.5 M NaOH for 1 h at 37°C, after which 200 μ l of a solution 6 \times SSC–6% formaldehyde–25 mM NaH₂PO₄ were added. DNA was applied onto a Hybond N nylon membrane using a standard device for dot blotting. Then DNA immobilized on the membrane was hybridized to ³²P-labeled B1 mouse probe in a 4 \times SSC–0.5% SDS–5 \times Denhardt solution–0.1 mg/ml denatured herring DNA mixture at 60°C [26]. The

Rodents examined in this study

Family*, subfamily	Species**	Source
Muridae (mice)		
Murinae (mice)	<i>Mus musculus</i> (house mouse) D	Vivarium (Engelhardt Institute of Molecular Biology, Moscow)
	<i>Rattus norvegicus</i> (Norwegian rat) D	Vivarium (Engelhardt Institute of Molecular Biology, Moscow)
Gerbillinae (gerbils)	<i>Tatera indica</i> (Indian gerbil) D, C	E.A. Lyapunova (Koltzov Institute of Developmental Biology, Moscow)
Cricetidae (hamsters)	<i>Microtus socialis</i> (social vole) D, C	E.A. Lyapunova (Koltzov Institute of Developmental Biology, Moscow)
Spalacidae (mole rats)	<i>Spalax microphthalmus</i> (Russian mole rat) D, C	A. Puzachenko (Institute of Geography, Moscow)
Rhizomyidae (bamboo rats)	<i>Tachyoryctes splendens</i> (African bamboo rat) D, C	Provided by the author (L.L.)
Zapodidae (birch mice)	<i>Sicista tianschanica</i> (Tyan-Shan birch mouse) D, C	E.Yu. Ivanitskaya (Severtsov Institute of Ecology and Evolution, Moscow)
Dipodidae (jerboas)	<i>Eremodipus lichtensteini</i> (Lichtenstein's jerboa) D, C	G.I. Shenbrot (Severtsov Institute of Ecology and Evolution, Moscow)
	<i>Allactaga major</i> (greater jerboa) C	Acquired from a private person
Anomaluridae (scaly-tailed flying squirrels)	<i>Anomalurus</i> sp. (scaly-tailed squirrel) D, C	F. Catzeflis (Universite Montpellier 2, France)
Pedetidae (spring hares)	<i>Pedetes capensis</i> (spring hare) D, C	T. Robinson (University of Stellenbosch, Matieland, South Africa)
Castoridae (beavers)	<i>Castor fiber</i> (common beaver) D, C	O.P. Likhnova (Severtsov Institute of Ecology and Evolution, Moscow)
Geomyidae (gophers)	<i>Thomomys bottae</i> (western pocket gopher) D, C	F. Catzeflis (Universite Montpellier 2, France)
Heteromyidae (kangaroo mice)	<i>Chaetodipus californicus</i> (Californian kangaroo mouse) C	R. DeBry (University of Cincinnati, Cincinnati, United States)
Ctenodactylidae (gundis)	<i>Ctenodactylus gundi</i> (gundi) C	F. Catzeflis (Universite Montpellier 2, France)
Hystriidae (porcupines)	<i>Hystrix indica</i> (Indian porcupine) D, C	Moscow Zoo
Thryonomyidae (African cane rats)	<i>Thryonomys gregorianus</i> (lesser cane rat) D, C	Provided by the author (L.L.)
Dasyproctidae (agouties)	<i>Myoprocta acouchy</i> (acouchi) D, C	Moscow Zoo
Hydrochoeridae (capybaras)	<i>Hydrochoerus hydrochaeris</i> (water hog) D, C	Moscow Zoo
Caviidae (guinea pigs)	<i>Cavia porcellus</i> (guinea pig) D	Vivarium (Engelhardt Institute of Molecular Biology, Moscow)
Chinchillidae (chinchillas)	<i>Chinchilla laniger</i> D, C	R. DeBry (University of Cincinnati, Cincinnati, United States)
Octodontidae (degus)	<i>Octodon degus</i> (degu) D, C	A.P. Koval' (Engelhardt Institute of Molecular Biology, Moscow)
Myocastoridae (nutrias)	<i>Myocastor coypus</i> (nutria) D, C	Acquired from a private person
Gliridae (dormice)	<i>Dryomys nitedula</i> (wood dormouse) D, C	E.Yu. Ivanitskaya (Severtsov Institute of Ecology and Evolution, Moscow)
Sciuridae (squirrels)	<i>Sciurus carolinensis</i> (gray squirrel) D, C	E.A. Lyapunova (Koltzov Institute of Developmental Biology, Moscow)
Aplodontidae (mountain beavers)	<i>Aplodontia rufa</i> (mountain beaver) D, C	F. Catzeflis (Universite Montpellier 2, France)

* Taxonomy at the family level is given according to [16].

** D, species examined by dot hybridization; C, species, examined by cloning and sequencing.

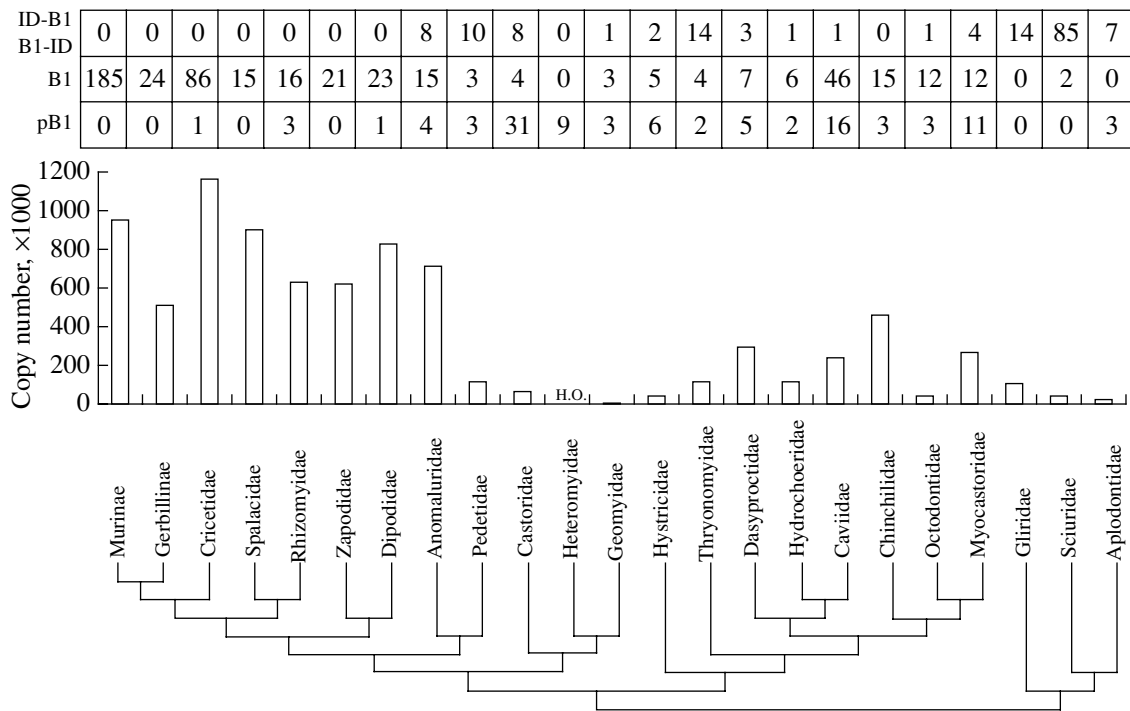


Fig. 2. B1 copy number in the genomes of rodents from various families. The column height shows the number of B1 copies estimated by dot hybridization of genomic DNA (the species examined are marked by D in the table). A phylogenetic tree of rodents showing relationships between the families is presented below. Muridae are represented by two subfamilies, Murinae and Gerbillinae. In species of the family Heteromyidae, the B1 copy number was not determined (n.d.) because of very small DNA samples. The copy number of pB1, B1, and dimeric SINES, B1-ID and ID-B1, sequenced or found in databases in the course of this study, is shown at the top of the figure.

membrane was washed in $0.1 \times \text{SSC}-0.1\%$ SDS at 42°C . Radioactivity bound to the genomic DNA was measured using a Cyclon phosphorimager (Packard, United States).

Library construction and screening. Rodent genomic DNAs (1.5–5.0 μg) were digested with restriction endonucleases *EcoRI* and *HindIII* (DNA of *Eremodipus lichtensteini* was digested with *KpnI* and *HindIII* because most of the jerboa B1 element copies contain an internal *EcoRI* site) and fractionated in 1% agarose gel. Then 500–1200-bp DNA fragments were electrophoretically transferred onto a DEAE membrane. DNA was eluted by incubating the membrane in 400 μl 1 M NaCl–10 mM Tris-HCl (pH 8.0)–1 mM EDTA at 60°C for 30 min and precipitated by ethanol, using 10 μg of glycogen as carrier. DNA fragments (0.1–0.5 μg) were ligated with plasmid vector pGEM3Z (0.1–0.3 μg) digested by the same restriction enzymes. The ligate was used to transform competent *Escherichia coli* XL-1 Blue cells. The colonies were transferred onto nitrocellulose membranes and hybridized under the same conditions as in the dot hybridization. Positive colonies were detected by radioautography. Clones of *E. coli* containing B1 fragments were purified from contaminating clones using two additional hybridization rounds and plating.

As the *C. gundi* genomic DNA was greatly degraded, after its electrophoresis in 3% agarose gel the 150–250-bp fragments were collected on the DEAE membrane and eluted as indicated above. To produce blunt-end DNA, the isolated fraction was incubated with the Klenow fragment (15 units) and 25 μM dNTP for 20 min at 30°C . Then, to add adenosine residues at 3' ends of the DNA fragments, the latter were incubated with *Taq* polymerase and 0.5 mM dATP for 30 min at 72°C . To create a *C. gundi* genomic library, the resultant DNA fragments were ligated to the pGEM-T vector (Promega, United States) according to the manufacturer instructions. Hybridization and screening of the libraries were conducted using radioactively labeled B1 probe as described above.

DNA sequencing and computer-aided analysis. The cloned DNA fragments containing B1 were sequenced using standard M13 primers, BigDye sequencing kit, and an AI Prism 3100-Avant sequencer (Applied Biosystems). The nucleotide sequences of the cloned DNA fragments were deposited to GenBank under accession nos. EF042308–EF042578.

B1 sequences were also identified by means of screening nucleotide databases using the FASTA procedure. Multiple alignments were conducted manually or using Clustal W and optimized using GeneDoc. Consensus sequences were created using GeneDoc on the

basis of alignments; hypervariable sites CG/CA/TG were manually replaced by CG.

To construct a tree, we used the continuous characters maximum likelihood procedure [30] from the PHYLIP package [31], which allows analyzing quantitative characters, because the information on the frequency of a particular character (i.e., the degree of its expression) may be lost in the consensus sequence. As characters, we used internal duplications, insertions, deletions, and single-nucleotide substitutions (except hypervariable sites CpG), marked by dots in Fig. 3. Out of 39 characters, 33 were used as clade I and its characters were excluded from analysis. A character was assigned a numerical index depending on its proportion in the B1 copies of the given family (e.g., +1.00, if it was presented in all sequences; 0.00, in half of them; -0.50, in a quarter of them, and so on).

RESULTS

Using dot hybridization of the labeled B1 *M. musculus* probe to genomic DNA of rodents from various families, we have estimated the number of B1 copies in their genomes. For this, we measured radioactivity of each spot including that containing house mouse DNA. Based on the number of B1 copies in the *M. musculus* genome (955 000), which had been estimated by sequencing of the total genome of this species [21], we computed the B1 copy number in other rodents. This number proved to significantly differ among rodent families. Families Muridae, Cricetidae, Spalacidae, Rhizomyidae, Zapodidae, and Dipodidae, constituting clade Myodonta, have very high (630 000–1 200 000) B1 copy numbers (Fig. 2). Two families of the clade Anomaluroomorpha showed a considerable difference in the B1 copy numbers, which were 720 000 in Anomaluridae and 130 000 in Pedetidae. Only 13 000 B1 copies were observed in *Thomomys bottae* (Geomyidae). Most of the families examined (Thryonomyidae, Dasyproctidae, Hydrochoeridae, Caviidae, Chinchilidae, and Myocastoridae) of the clade Hystricognathi had a moderate B1 copy number (130 000–460 000), although in *Hystric indica* (Hystricidae) and *Octodon degu* (Octodontidae) these values were only 39 000 and 47 000, respectively. Squirrel-related rodents had relatively small (28 000–45 000: Aplodontidae and Sciuridae) or moderate (114 000: Gliridae) B1 copy number in the genomes (Fig. 2).

To elucidate the specific structure of B1 elements in DNA of rodents from various families, we developed genomic libraries for 23 species of 22 Rodentia families. After screening the libraries by hybridization of colonies with labeled B1 element, plasmids from the clones that produced hybridization signal were sequenced. In this manner, sequences of 305 B1 copies were established. In addition, some B1 copies were found in database GenBank/EMBL for rodents. Most B1 copy sequences of rodents from families Cricetidae,

Caviidae, and Sciuridae were identified by means of computer-aided screening.

All of the identified B1 sequences were analyzed and systematized by classing them into three categories: (I) canonical B1 variant containing internal tandem duplication; (II) pB1 variant lacking this duplication and generally considered as the precursor of the other B1 variants; (III) complex SINEs containing B1 as one of two monomers. In Fig. 2, the number of identified B1 copies of each of the three categories is presented for each rodent family. The classical B1 variant prevailed over pB1 in most of the rodent families. However, in Pedetidae, Geomyidae, and Hystricidae these two B1 variants seemed to occur in approximately equal numbers. Moreover, in the genome of beaver (Castoridae), pB1 content was eightfold higher than that of duplication-bearing B1. In the case of kangaroo mouse (Heteromyidae), nine pB1 versus none of canonical B1 copies were cloned, which suggests at least significant predominance of pB1 over B1 in this rodent genome. Note that rodent families that did not show B1 predominance over pB1, occupy more basal positions in the phylogenetic tree than most of the other families (Fig. 2). Interestingly, according to the dot hybridization results, the above families have the lowest number of pB1/B1 elements in the genomes of their members. Gliridae, Sciuridae, and Aplodontidae presumably occupy even more basal position. In these rodents, copies of the complex B1-containing element B1-dID prevail: only two B1 copies in Sciuridae and three pB1 copies in Aplodontidae were found (Fig. 2).

pB1 copies in all species examined showed strong divergence, but their consensus sequences were very similar in different rodents (data not shown). Most of the among-consensus differences observed were related to hypervariable (methylated) CpG sites. No specific subfamilies were found among pB1 elements. An exception was presented by pB1 and complex elements (pB1-ID and ID-pB1) from the Pedetidae and Anomaluridae genomes, which contained a specific insertion AGAGG(C/G), absent in pB1 of all other species examined. This common feature indicates that these two rodent families are related.

In contrast to pB1, canonical (i.e., having an internal tandem duplication) B1 from the genomes of rodents of various families very often exhibit specific structural features. Consensus sequences were established for each rodent family. Figure 3 presents alignment of these consensus sequences, including those for B1 monomers of complex SINEs. The family Muridae is represented in Fig. 3 by four consensus sequences: B1_Mus (*Mus musculus*, subfamily Murinae), variants B1_Rat-A and B1_Rat-B (*Rattus norvegicus*, Murinae), and B1-Ger (subfamily Gerbillinae). The family Cricetidae is also represented by four consensus sequences: three variants of monomer B1 (B1_Cri-A0, B1_Cri-A2, and B1_Cri-B1) and one variant included in the dimeric SINE (B2-B1_Cri).

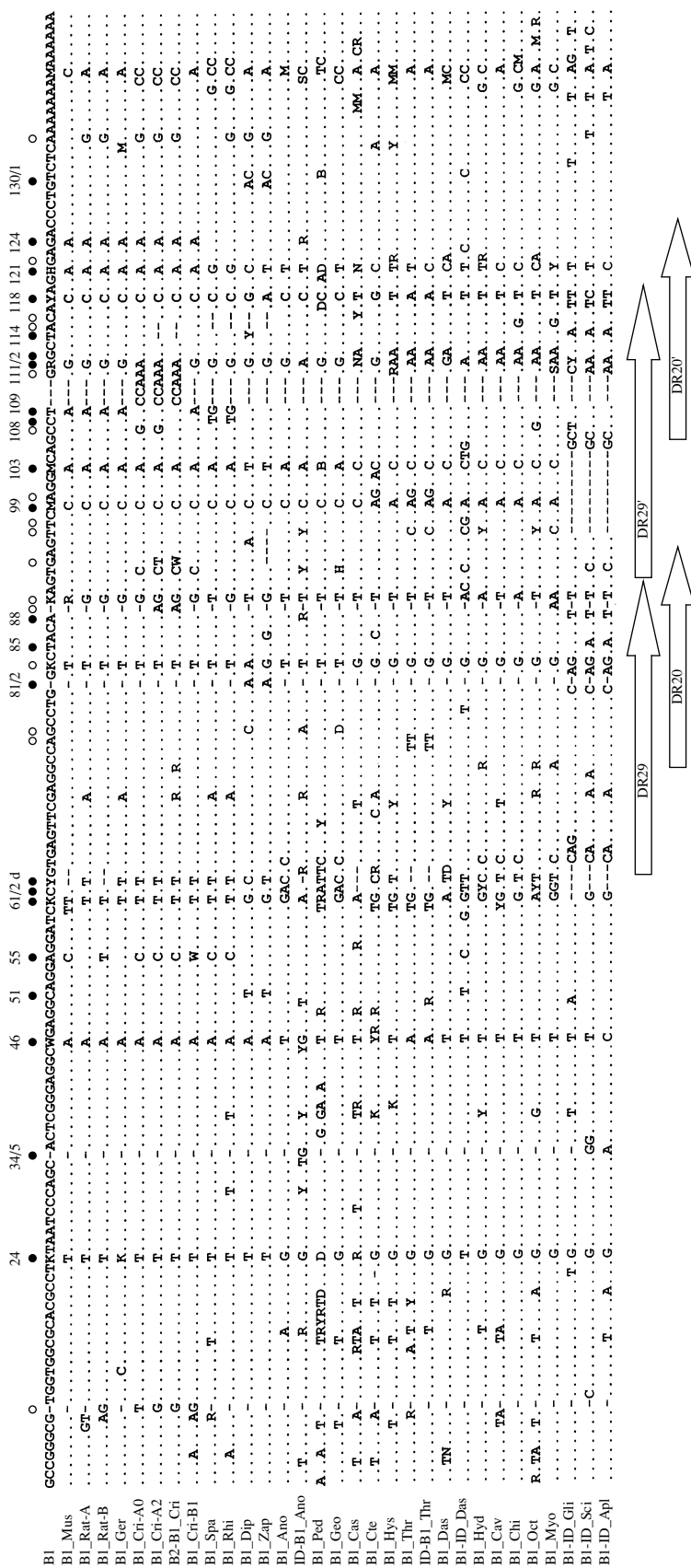


Fig. 3. Alignment of consensus B1 sequences from the genomes of rodents of various families. The top row is the formal B1 consensus based on all family-specific consensus sequences. Nucleotides coinciding with those in the formal consensus are marked by dots. The absent nucleotides are shown by a “-” sign. The sequences are termed by three first letters of the family name (for full names see Fig. 4). An exception is the family Muridae represented by four consensus sequences: B1_Mus (*Mus musculus*, subfamily Murinae), variants B1_Rat-A and B1_Rat-B (*Rattus norvegicus*, Murinae), and B1_Ger (subfamily Gerbillinae). The family Cricetidae is also represented by four consensus sequences: three variants of multimeric B1 (B1_Cri-A0, B1_Cri-A2, and B1_Cri-B1) and one variant from the dimeric SINE (B2-B1_Cri). B1 sequences that form dimeric SINEs with the ID element are also presented. Top dots mark all significant characters used for the rodent phylogenetic reconstruction (Fig. 5); the characters supporting the current phylogenetic tree of rodents (see Fig. 4) are shown by solid dots and position numbers according to the murine B1 consensus.

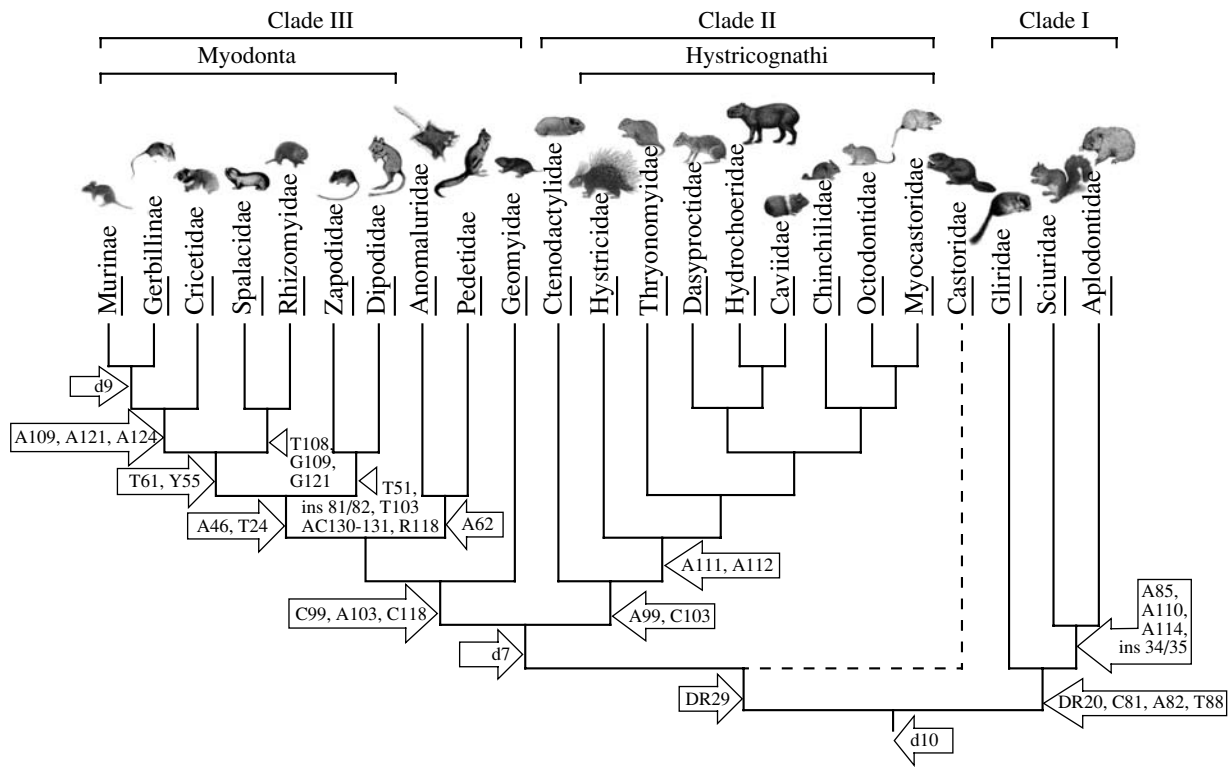


Fig. 4. Phylogenetic tree of the family Rodentia. The probable times of the appearance of substitutions, deletions, and duplications in the B1 family, estimated by analysis of their absence/presence in various rodent families, are given. The observed character distribution generally supports the topology of this tree, constructed by combining the data from several studies [34–38, 44]. The family Castoridae, whose position had to be changed, is an exception (compare to tree in Fig. 2). The boundaries of clades I, II, and III, as well as groups Myodonta and Hystricognathi, are shown on top. These clades correspond to those described earlier: clade I, the squirrel and dormouse-related clade; clade II, the Ctenohystrica clade (Ctenodactylidae + Hystricognathi); clade III, the mouse-related clade [33, 35].

In recent years, based on analysis of nuclear gene sequences, considerable advances have been made in understanding relationships among rodent families. Nevertheless, some unresolved issues still remain. Figures 2 and 4 presents the phylogenetic tree of rodent families, constructed by us on the basis of molecular phylogenetic data from several articles [29, 32–37] (only the families examined in the present work are shown).

Comparison of B1 consensus sequences identified 39 characters (long duplications, insertions, deletions) by which differed at least several consensus sequences. Of these, 23 characters supported the rodent tree (Fig. 4) whereas the remaining characters either were uninformative, i.e., occurred only in one family (10) or did not correspond to the tree (6). Let us consider some of the synapomorphic characters. The presence of a 29-bp duplication (DR29) in B1 of all rodents except Gliridae, Scuridae, and Aplodontidae (clade I), which are characterized by a 20-bp duplication (DR20), supports relatedness between clades II and III and suggests earlier divergence of clade I from the other rodents (Fig. 4). Deletion d10, characteristic of clade I, also shows its isolation from clades II and III, in which B1s carry deletions d7 or d9 (with Castoridae having deletion d10 as the single exception). Monophyletic origin of clade I is

supported also by specific nucleotide substitutions in positions 81, 82, and 88. Clades II and III are differentiated from one another owing to differences in positions 99, 103, and 118 of their B1 elements (Fig. 4). In addition, all families from the group Hystricognathi, which together with Ctenodactylidae forms clade II, are characterized by substitutions A111 and A112. In different families of clade II, B1 are very similar; an exception is the family Thyronomyidae, which had deletion d9 (rather than d7 as in other clade II families) and nucleotide substitutions in positions 75, 76, 100, and 118 (Fig. 3). In clade III, particularly in the Myodonta family group, B1 elements evolved much faster. Myodonta monophyly is supported by substitutions T24 and A46. Close relationships between Muridae (Murinae + Gerbillinae), Cricetidae, Spalacidae, and Rhizomyidae are shown by nucleotide positions Y55 and T61. Sister relationships of Muridae and Cricetidae are supported by nucleotide substitutions in positions 109, 121, and 124 on A. High similarity between B1 of mole rats and bamboo rats and the presence in them of specific nucleotide substitutions (T108, G109, G121) testify to close phylogenetic relationships between Spalacidae and Rhizomyidae. B1s of jerboas (Dipodidae) and birch mice (Zapodidae) contain insertion A81/A82 and five shared nucleotide substitutions (T51,

T103, R118, A130, and C131), which confirm relatedness of these families.

In the tree constructed on the basis of literature data [32, 33, 35–37], Castoridae (beavers) form a common clade with Geomyidae and Heteromyidae (Fig. 2). However, the results of analysis of B1 elements suggested that Castoridae occupies a different, basal position with regard to clades II and III. This is reflected in Fig. 4, where a dashed line shows the alternative position of Castoridae in the tree. This conclusion is based on the fact that beavers are an a sole family whose B1 combine deletion d10 with duplication DR29. The following scenario of the evolution of rodents and their B1, corresponding to the basal tree part, can be suggested. In clade I, element pB1d10 after the appearance of 20-bp duplication yielded the SINE family B1-dID. In the common ancestor of other rodents, B1 with 29-bp duplication (the element of this type was found in the beaver genome) could evolve from pB1d10. Then in this B1 element, deletion d10 was truncated via tandem duplication of three nucleotides (CGC) positioned upstream to this deletion, which produced B1 with deletion d7 and duplication DR29. The newly arisen element B1d7DR29 abundantly amplified in rodent genomes of group Myodonta proved to be more successful than its predecessor.

In addition to the above testing of the phylogenetic tree, based on nuclear gene analysis, we made an attempt to construct an independent phylogenetic tree of the rodent family on the basis of B1 consensus sequences of rodent families. Employing standard methods of tree construction (for instance, maximum likelihood or maximum parsimony) for B1 element sequences encounters difficulties, related, in particular, to different expression of characters in consensus sequences of rodents from different families. Because of this, we used the continuous characters maximum likelihood method, designed for quantitative characters and thus accounting for the different expression of the characters (e.g., the proportion between nucleotides in the given position of the B1 sequences, see Materials and Methods). Figure 5 presents such unrooted tree constructed for B1 elements of rodent families (B1 monomers of all complex SINEs were excluded from analysis). This tree clearly supports monophyly of clades II and III, as well as group Myodonta. However, note that the branch position within these clades do not always conform to that depicted in Figs. 2 and 4.

DISCUSSION

Major Variants of B1 Elements

In the present study, we have conducted analysis of B1 elements in the genomes of rodents from most (22) of the known (30–35) families. More than 300 copies of these SINEs were sequenced. In addition, we have analyzed a significant number (269) of B1 sequences obtained by screening bases of these copies from the

genomes of hamsters (Cricetidae), guinea pig (Caviidae) and squirrels (Sciuridae). Analysis of sequences of all of the copies permitted classifying them into three types: (1) pB1, (2) canonical B1, and (3) complex SINEs containing B1. In contrast to pB1, B1 carry an internal 29-bp tandem duplication. pB1 is considered to be an evolutionary precursor of B1 [38]. Complex SINEs generally consist of two monomers, one of which is pB1 or B1 and the other, an ID element. Depending on the family of such SINEs, the B1 sequence can constitute the right or the left monomer.

Estimation of the total copy number has shown that the number of B1 elements is the highest in rodents of the group Myodonta, lower in the species from the sub-order Hystricognathi, and relatively low in Pedetidae, Castoridae, Geomyidae, and Hystricidae. Interestingly, predominance of the copy number of B1 over pB1 was not observed in these four families (Fig. 2). Furthermore, these families showed an early divergence in the rodent evolution. Apparently, in the genomes of rodents from these families, canonical B1 elements could not become effective retroposons as analogous elements in other rodents. The number of pB1 copies in the genomes of all rodents could expected to be commensurable, but significant differences in the canonical B1 number may explain the fact that pB1 were more readily detected in the rodent genomes with a low number of canonical B1 elements (Fig. 2).

B1 and Rodent Phylogeny

Analysis of the pB1 element structure proved to be poorly efficient for estimating phylogenetic relationships among the rodent families. Only in the case of Pedetidae and Anomaluridae, these SINEs were informative, because in these rodents pB1 elements carrying a specific 6-bp insertion were found, which supported relatedness of these families.

The structure of B1 elements proves to be more informative. We have detected a number of nucleotide substitutions, as well as deletions and duplications, characteristic for particular rodent families. The distribution of these characters in the phylogenetic tree (Fig. 4) generally supported the current views on the relationships among the rodent families. Until recently, these relationships were studied exclusively on the basis on morphological traits [16, 39]. Later these issues were examined using mitochondrial DNA analysis [40], but because of some paradoxical conclusions made from these data (for instance that guinea pig is not a rodent [41]), the phylogenetic tree based on the results of these studies were not universally accepted. In the last years, exon sequences of several nuclear genes were examined, which, together with using new procedures for tree construction led to considerable advances in this line [32–36, 42]. However, positions of some branches are still not quite clear. In particular, this holds for the mutual positions of clade I (Sciuridae, Aplodontidae, and Gliridae), clade II (Hystricognathi + Ctenodactyl-

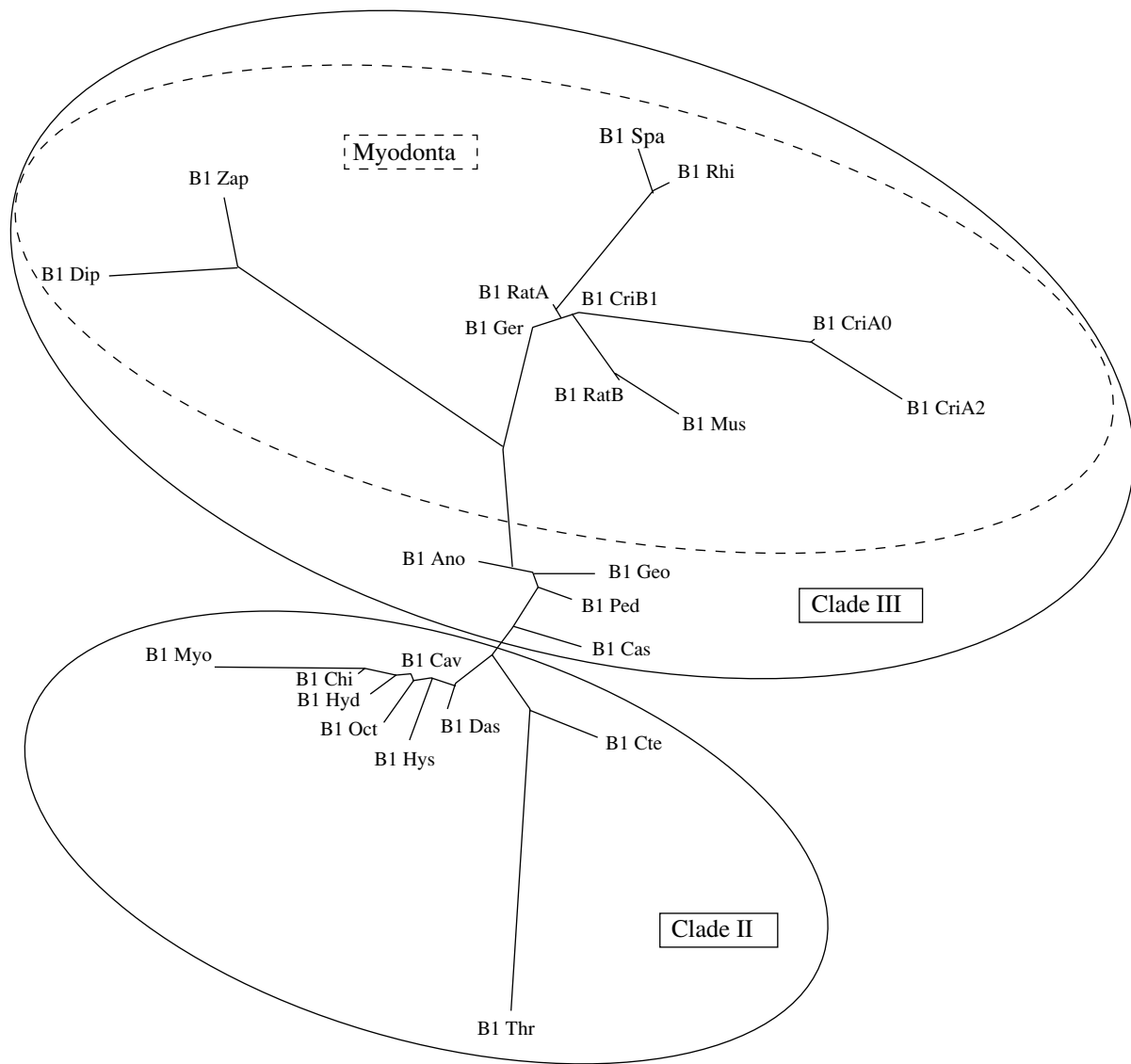


Fig. 5. Tree based on analysis of the B1 structure characters of rodents from various families (except Gliridae, Sciuridae, and Aplodontidae). All significant characters of B1 sequences (marked by dots in Fig. 3) were used. In B1 designations, three first letters from the rodent family names are used (for full names, see Fig. 4). Clades II and III, and group Myodonta are circled.

idae), and clade III (the remaining rodents) (Fig. 4). Moreover, the localization of the Geomyidae–Heteromyidae–Castoridae branch in the tree is doubtful. Note that the positions of most other families in the tree still require additional support. The presence of a 20-bp duplication in B1 of Gliridae, Sciuridae, and Aplodontidae as opposed to a 29-bp duplication in other rodents testifies to relatedness of clades II and III as well as earlier clade I divergence. In addition, the presence of a 10-bp deletion (d10) in B1 virtually only in rodents of the families Gliridae, Sciuridae, and Aplodontidae also suggests a basal position of clade I in the tree. This is supported by the presence of the same deletion in tree shrews that represent another order [26, 27].

At least three nucleotide substitutions (C99, A103, and C118) confirm that the family Geomyidae belongs to clade III. At the same time, our results cast doubt on the relatedness of Geomyidae and Castoridae, suggesting that the divergence of Castoridae occurred prior to the separation of clades II and III. Finally, note that a significant number of nucleotide substitutions as well as insertions are in a good agreement with the tree topology within clade III, shown in Fig. 4.

Comparison of SINE-Based Methods of Phylogeny Analysis

Two SINE-based approaches to studying species phylogeny have been developed. The first approach

deals with the presence or absence of SINE copies in definite genome loci of different species (the method of copies). In the second approach, the presence or absence of the whole SINE families is established (the method of families). The presence of SINE copies in particular loci or the presence of SINE families in the genomes indicate that the species in question are related. The main advantage of these methods lies in the fact that SINEs, unlike morphological characters, do not tend to homoplasy—phenomenon that includes convergence and parallelism. Note that morphology of rodents is especially rich in homoplasy cases [16]. The absence of homoplasy in SINEs is based on the following. SINE integration in the same genome site in different species is a very improbable event because of enormous genome size and the lack of marked preference for integration sites. The appearance of two identical SINE families in two different species also has extremely low probability.

Methods based on analysis of coding gene sequences, which are today widespread, also may suffer from homoplasy, because convergent and parallel evolution of functional DNA sequences is quite plausible. The presence of a high proportion of positions with nucleotide substitutions in coding sequences also creates problems for using these methods, since this is accompanied by frequent reverse mutations. Moreover, the results of these methods may depend on the procedures used for the phylogenetic tree construction.

SINE-based methods find increasing use for constructing phylogenetic trees as well as testing trees based on analysis of morphological traits or coding DNA sequences. For instance, using the method of copies, it was shown that whales (Cetacea) are a branch within the even-toed ungulates (Artiodactyla), which allowed uniting them into one taxonomic group, Cetartiodactyla [43]. This method is successfully used for investigating relations among various human populations and races [44, 45]. Using the methods of families, the family of dormice (Gliridae) was shown to be close to squirrels (Scuridae) rather than myomorph rodents, as thought earlier [29]. This method was also used to conclusively support the monophyletic origin of clade Afrotheria, which includes elephants, hyraxes, sea cows, armadillos, golden moles, tenrecs, and elephant shrews [46]. Moreover, this approach was used to demonstrate the relatedness of primates, tree shrews, and rodents [26, 27].

Although SINE-based methods seem very reliable, their use may also present some problems. In the case of the method of copies, the following difficulties may be encountered. Independent integration of a SINE copy into the same genome site in different species do occur, albeit extremely rarely [47]. It is very important for these methods that the copy integration is irreversible and preserved during long periods of evolution. However, recently it has been found that 0.5–1.0% of Alu copies are lost from the genome because of recom-

bination between short direct flanking repeats [48]. It seems that such excision concerns mostly young Alu copies, because they are flanked by perfect direct repeats. This finding highlights the fact that the position of a tree branch should be supported by analysis of several rather than one SINE copy. Studies of taxonomic groups undergoing very fast speciation, such as African cichlid fishes, was particularly connected with difficulties [49]. Note that many “young” SINE copies are not fixed at the population level, i.e., they occur not in all individuals of the population. As a result, rapid speciation entails a situation in which one novel species may descend from an individual carrying a given SINE copy, while another novel species stems from an ancestor lacking this copy. In such cases employing the copy method for elucidation of phylogenetic relationships between the species is problematic [49]. The problems described above are not inherent to the methods of families. However, the efficiency of SINE amplification may greatly differ for different organisms; consequently, their detection may be hindered by a small copy number and strong divergence.

In the present work, we for the first time used structure analysis of SINE subfamilies from one family for phylogenetic reconstructions. These B1 element subfamilies are characterized by a definite set of diagnostic nucleotide positions. In essence, this approach is a variant of the family method and has the same advantages and potential drawbacks.

What is the explanation of the fact that distribution analysis of B1 subfamilies yields information on the rodent phylogeny? The evolution of B1 subfamilies likely went in parallel with the evolution of their “hosts.” This means that the appearance of new rodent groups was accompanied by the formation of new successful B1 subfamilies from those that had already existed. These new subfamilies, retaining their diagnostic characters, required new, additional ones. Thus, tracing the evolution of B1 families, one can draw conclusions on the phylogeny of the rodents themselves.

All diagnostic positions presented in Fig. 4 behave as synapomorphic characters, supporting monophyly of certain branches in the tree. However, we would like to note that several positions (characters) in Fig. 3 did not allow us to make unambiguous phylogenetic conclusions. For instance, dinucleotide deletion AC in position 115–116, occurring in B1 of rodents from the families Spalacidae, Rhyzomyidae, Dipodidae, and Zapusidae, indicated a monophyletic origin of the whole group. However, this is at variance with the distribution of other diagnostic positions, which we detected, and with other molecular phylogenetic data (Fig. 4). It cannot be excluded that this deletion appeared independently in clade Spalacidae/Rhyzomyidae and clade Dipodidae/Zapusidae. Yet, it seems more plausible that this deletion have appeared in the common Myodonta ancestor, but potent amplification of such B1 variants occurred only in these four families.

The nucleotide substitution in position 83 also leads to somewhat unclear conclusions. It seems likely, however, that the G for T substitution occurred in the common ancestor of clade III, whereas in jerboas (Dipodidae) and birch mice (Zapodidae) it was again replaced by a purine base.

These two examples demonstrate that the method used in the present study has its pitfalls. However, most of the described diagnostic characters are synapomorphic and can be used for testing phylogenetic trees constructed by other methods. Hopefully this method will find application in phylogenetic studies of other groups of organisms.

ACKNOWLEDGMENTS

We thank R. DeBry, E. Ivanitskaya, F. Catzeflis, O. Likhnova, E. Lyapunova, A. Puzachenko, T. Robinson, and G. Shenbrot for providing animals, their tissues, and DNA, and M. Grashchuk for assistance in some experiments.

This work was supported by the Russian Foundation for Fundamental Research (grant nos. 05-04-49553 and 07-04-00462).

REFERENCES

- Deininger, P.L. and Batzer, M.A., Mammalian Retroelements, *Genome Res.*, 2002, vol. 12, no. 10, pp. 1455–1465.
- Kramerov, D. and Vassetzky, N., Short Retroposons in Eukaryotic Genomes, *Int. Rev. Cytol.*, 2005, vol. 247, pp. 165–221.
- Okada, N., SINEs, *Curr. Opin. Genet. Dev.*, 1991, vol. 1, no. 4, pp. 498–504.
- Kapitonov, V.V. and Jurka, J., A Novel Class of SINE Elements Derived from 5S rRNA, *Mol. Biol. Evol.*, 2003, vol. 20, no. 5, pp. 694–702.
- Nishihara, H., Smit, A.F., and Okada, N., Functional Noncoding Sequences Derived from SINEs in the Mammalian Genome, *Genome Res.*, 2006, vol. 16, no. 7, pp. 864–874.
- Kramerov, D.A., Grigoryan, A.A., Ryskov, A.P., et al., 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.*, 1979, vol. 6, no. 2, pp. 697–713.
- Krayev, A.S., Kramerov, D.A., Skryabin, K.G., 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.*, 1980, vol. 8, no. 6, pp. 1201–1215.
- Deininger, P.L., Jolly, D.J., Rubin, C.M., et al., Base Sequence Studies of 300 Nucleotide Renatured Repeated Human DNA Clones, *J. Mol. Biol.*, 1981, vol. 151, no. 1, pp. 17–33.
- Haynes, S.R., Toomey, T.P., Leinwand, L., et al., 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.*, 1981, vol. 1, no. 7, pp. 573–583.
- Ullu, E. and Tschudi, C., Alu Sequences Are Processed 7SL RNA Genes, *Nature*, 1984, vol. 312, no. 5990, pp. 171–172.
- Daniels, G.R. and Deininger, P.L., A Second Major Class of Alu Family Repeated DNA Sequences in a Primate Genome, *Nucleic Acids Res.*, 1983, vol. 11, no. 21, pp. 7595–7610.
- Zietkiewicz, E., Richer, C., Sinnett, D., et al., Monophyletic Origin of Alu Elements in Primates, *J. Mol. Evol.*, 1998, vol. 47, no. 2, pp. 172–182.
- Roos, C., Schmitz, J., and Zischler, H., Primate Jumping Genes Elucidate Strepsirrhine Phylogeny, *Proc. Natl. Acad. Sci. USA*, 2004, vol. 101, no. 29, pp. 10650–10654.
- Labuda, D., Sinnett, D., Richer, C., et al., Evolution of Mouse B1 Repeats: 7SL RNA Folding Pattern Conserved, *J. Mol. Evol.*, 1991, vol. 32, no. 5, pp. 405–414.
- Quentin, Y., A Master Sequence Related to a Free Left Alu Monomer (FLAM) at the Origin of the B1 Family in Rodent Genomes, *Nucleic Acids Res.*, 1994, vol. 22, no. 12, pp. 2222–2227.
- Hartenberger, J.-L., The Order Rodentia: Major Questions on Their Evolutionary Origin, Relationships and Suprafamilial Systematics, in *Evolutionary Relationships among Rodents*, Lockett, W. and Hartenberger, J.-L., Eds., New York: Plenum, 1985, pp. 1–33.
- Pavlinov, I., *Sistematika sovremennykh mlekopitayushchikh* (Systematics of Contemporary Mammals), Moscow: Mosk. Gos. Univ., 2003.
- Krayev, A.S., Markusheva, T.V., Kramerov, D.A., et al., Ubiquitous Transposon-Like Repeats B1 and B2 of the Mouse Genome: B2 Sequencing, *Nucleic Acids Res.*, 1982, vol. 10, no. 23, pp. 7461–7475.
- Den Dunnen, J.T. and Schoenmakers, J.G., Consensus Sequences of the *Rattus norvegicus* B1 and B2 Repeats, *Nucleic Acids Res.*, 1987, vol. 15, no. 6, p. 2772.
- Bains, W. and Temple-Smith, K., Similarity and Divergence among Rodent Repetitive DNA Sequences, *J. Mol. Evol.*, 1989, vol. 28, no. 3, pp. 191–199.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., et al., Initial Sequencing and Comparative Analysis of the Mouse Genome, *Nature*, 2002, vol. 420, no. 6915, pp. 520–562.
- Gibbs, R.A., Weinstock, G.M., Metzker, M.L., et al., Genome Sequence of the Brown Norway Rat Yields Insights into Mammalian Evolution, *Nature*, 2004, vol. 428, no. 6982, pp. 493–521.
- Kim, J., Martignetti, J.A., Shen, M.R., et al., Rodent BC1 RNA Gene as a Master Gene for ID Element Amplification, *Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, no. 9, pp. 3607–3611.
- Serdobova, I.M. and Kramerov, D.A., Short Retroposons of the B2 Superfamily: Evolution and Application for the Study of Rodent Phylogeny, *J. Mol. Evol.*, 1998, vol. 46, pp. 202–214.
- Kramerov, D.A. and Vassetzky, N.S., Structure and Origin of a Novel Dimeric Retroposon B1-dID, *J. Mol. Evol.*, 2001, vol. 52, no. 2, pp. 137–143.

26. Vassetzky, N.S., Ten, O.A., and Kramerov, D.A., B1 and Related SINES in Mammalian Genomes, *Gene*, 2003, vol. 319, pp. 149–160.
27. Nishihara, H., Terai, Y., and Okada, N., Characterization of Novel Alu- and tRNA-Related SINES from the Tree Shrew and Evolutionary Implications of Their Origins, *Mol. Biol. Evol.*, 2002, vol. 19, no. 11, pp. 1964–1972.
28. Serdobova, I.M. and Kramerov, D.A., Usage of Short Retroposons as Phylogenetic Markers, *Dokl. Akad. Nauk SSSR*, 1994, vol. 335, no. 5, pp. 664–667.
29. Kramerov, D., Vassetzky, N., and Serdobova, I., The Evolutionary Position of Dormice (Gliridae) in Rodentia Determined by a Novel Short Retroposon, *Mol. Biol. Evol.*, 1999, pp. 715–716.
30. Felsenstein, J., Phylogenies from Gene Frequencies: A Statistical Problem, *Systematic Zool.*, 1985, vol. 34, no. 3, pp. 300–311.
31. Felsenstein, J., PHYLIP—Phylogeny Inference Package (Version 3.2), *Cladistics*, 1989, vol. 5, pp. 164–166.
32. Huchon, D. and Douzery, E.J., From the Old World to the New World: A Molecular Chronicle of the Phylogeny and Biogeography of Hystricognath Rodents, *Mol. Phylogenet. Evol.*, 2001, vol. 20, no. 2, pp. 238–251.
33. Huchon, D., Madsen, O., Sibbald, M.J., 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.*, 2002, vol. 19, no. 7, pp. 1053–1065.
34. Adkins, R.M., Gelke, E.L., Rowe, D., et al., Molecular Phylogeny and Divergence Time Estimates for Major Rodent Groups: Evidence from Multiple Genes, *Mol. Biol. Evol.*, 2001, vol. 18, no. 5, pp. 777–791.
35. Adkins, R.M., Walton, A.H., and Honeycutt, R.L., Higher-Level Systematics of Rodents and Divergence Time Estimates Based on Two Congruent Nuclear Genes, *Mol. Phylogenet. Evol.*, 2003, vol. 26, no. 3, pp. 409–420.
36. Stepan, S., Adkins, R., and Anderson, J., Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes, *Syst. Biol.*, 2004, vol. 53, no. 4, pp. 533–553.
37. Opazo, J.C., A Molecular Timescale for Caviomorph Rodents (Mammalia, Hystricognathi), *Mol. Phylogenet. Evol.*, 2005, vol. 37, no. 3, pp. 932–937.
38. Quentin, Y., Emergence of Master Sequences in Families of Retroposons Derived from 7sl RNA, *Genetics*, 1994, vol. 93, nos. 1–3, pp. 203–215.
39. Carroll, R., *Vertebrate Paleontology and Evolution*, New York: Freeman, 1988.
40. Reyes, A., Pesole, G., and Saccone, C., Long-Branch Attraction Phenomenon and the Impact of Among-Site Variation on Rodent Phylogeny, *Gene*, 2000, vol. 259, nos. 1–2, pp. 177–187.
41. D’erchia, A.M., Gissi, C., Pesole, G., et al., The Guinea-Pig Is Not a Rodent, *Nature*, 1996, vol. 381, no. 6583, pp. 597–600.
42. Murphy, W.J., Eizirik, E., O’Brien, S.J., et al., Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics, *Science*, 2001, vol. 294, no. 5550, pp. 2348–2351.
43. Shimamura, M., Yasue, H., Ohshima, K., et al., Molecular Evidence from Retroposons that Whales Form a Clade within Even-Toed Ungulates, *Nature*, 1997, vol. 388, no. 6643, pp. 666–670.
44. Stoneking, M., Fontius, J.J., Clifford, S.L., et al., Alu Insertion Polymorphisms and Human Evolution: Evidence for a Larger Population Size in Africa, *Genome Res.*, 1997, vol. 7, no. 11, pp. 1061–1071.
45. Watkins, W.S., Rogers, A.R., Ostler, C.T., et al., Genetic Variation Among World Populations: Inferences from 100 Alu Insertion Polymorphisms, *Genome Res.*, 2003, vol. 13, no. 7, pp. 1607–1618.
46. Nikaido, M., Nishihara, H., Hukumoto, Y., et al., Ancient SINES from African Endemic Mammals, *Mol. Biol. Evol.*, 2003, vol. 20, no. 4, pp. 522–527.
47. Rothenburg, S., Eiben, M., Koch-Nolte, F., et al., Independent Integration of Rodent Identifier (ID) Elements into Orthologous Sites of Some RT6 Alleles of *Rattus norvegicus* and *Rattus rattus*, *J. Mol. Evol.*, 2002, vol. 55, no. 3, pp. 251–259.
48. Van De Lagemaat, L.N., Gagnier, L., Medstrand, P., et al., Genomic Deletions and Precise Removal of Transposable Elements Mediated by Short Identical DNA Segments in Primates, *Genome Res.*, 2005, vol. 15, no. 9, pp. 1243–1249.
49. Terai, Y., Takahashi, K., Nishida, M., et al., Using SINES to Probe Ancient Explosive Speciation: “Hidden” Radiation of African Cichlids?, *Mol. Biol. Evol.*, 2003, vol. 20, no. 6, pp. 924–930.