Nucleotide sequences of B1 SINE and 4.5S RNA support a close relationship of zokors to blind mole rats (Spalacinae) and bamboo rats (Rhizomyinae)

Irina K. Gogolevskaya, Natalia A. Veniaminova, Dmitri A. Kramerov

Abstract

Until recently, zokors (Myospalacinae) were assigned to the Cricetidae family. However, analysis of mitochondrial and nuclear genes suggests a sister relationship between zokors and subterranean rodents of the Spalacidae family, namely blind mole rats (Spalacinae) and bamboo rats (Rhizomyinae). Here, we cloned and sequenced copies of the B1 short interspersed element (SINE) from the genome of zokor Myospalax psilurus. The consensus nucleotide sequence of zokor B1 was very similar to spalacids and rhizomyids, but not to cricetids. Similar to spalacids (Spalax microphthalmus) and rhizomyids (Tachyoryctes splendens), zokor contained two variants of the 4.5S small nuclear RNA. The longer variant (L-variant, 104 nucleotides) was found only in zokor, spalacids and rhizomyids. The short, or S-variant (98 nucleotides), had a wider distribution; however, analysis of the nucleotide sequences of S-variants of 4.5S RNA confirmed that zokors are closely related to spalacids and rhizomyids, but not to cricetids. The evolution of the 4.5S RNA genes and pseudogenes is discussed.

Keywords: Retroposon Small RNA 4.5S RNA Rodents Phylogeny

1. Introduction

Zokors are Asiatic burrowing rodents that comprise two genera, Myospalax and Eospalax. Their taxonomic position is controversial. Tullberg (1899) first suggested a relationship between zokors and two families of subterranean muroid rodents, Spalacidae and Rhizomyidae. Later, other morphologists also assigned zokors to Spalacidae (Miller and Gidley, 1918; Chaline et al., 1977). Gradually, however, an alternative classification that assigned zokors (subfamily Myospalacinae) to Cricetidae became accepted (Simpson, 1945; Chaline et al., 1977; Gromov and Polyakov, 1977; Carleton and Musser, 1984; Carrol, 1988; Pavlinov, 2003). Zokors could also be regarded as a separate branch within arvicoline rather than an independent derivative of the ancestral cricetid stock. Thus, Kretzoi (1955) amended arvicolines to the family level, placed myospalacines as a subfamily, and assigned the same rank to true lemmings (Lemminae) and mole voles (Ellobii). This taxonomic affiliation of zokors was recently supported by Agadjanian (2009). Molecular studies initially confirmed the relationship between zokors and cricetids (Michaux and Catzefflis, 2000; Michaux et al., 2001). However, the results of Norris et al. (2004) as well as Jansa and Weksler (2004) suggested that Myospalax is a sister group to the clade containing the Spalacinae and Rhizomyinae subfamilies. These discrepancies have been attributed to the misidentification of a single sample of zokor in the study by Michaux and Catzefflis.

To date, phylogenetic data on zokor has been obtained primarily through analysis of mitochondrial and nuclear genes. The possibility remains that genetic similarities among zokors, mole rats and bamboo rats could be the result of convergence caused by a similar (subterranean) mode of life. Additionally, phylogenetic analysis of gene nucleotide sequences requires complex computer programs, and the use of different algorithms can result in different trees. Ideally, gene-based phylogenetic trees should be confirmed by analysis of other kinds of molecular markers (Shedlock and Okada, 2000; Murphy et al., 2004; Kriegs et al., 2007; Kramerov and Vasetskii, 2009). We have developed several markers in our laboratory that can potentially shed light on the phylogeny of zokors. One of these markers, the B1 short interspersed element (SINE) is likely not under selection pressure (Veniaminova et al., 2007b). A second marker, the 4.5S small nuclear RNA, has yet to be functionally defined (Gogolevskaya and Kramerov, 2002). Both sequences often exhibit distinct features among rodent families. Here, we investigated the feasibility of using these markers to determine whether zokors are a sister group to hamsters (Cricetidae), or to mole rats (Spalacinae) and bamboo rats (Rhizomyinae).

Abbreviations: SINE, short interspersed element; LINE, long interspersed element; Pol III, RNA polymerase III; PAGE, polyacrylamide gel electrophoresis.

* Corresponding author. Fax: +7 499 135 1405.
E-mail address: kramerov@eimb.ru (D.A. Kramerov).

1 Present address: Department of Internal Medicine, University of Michigan, 109 Zina Pitcher Place, 888 2468, Ann Arbor, MI 48109-2200, USA.

0378-1119/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
SINEs or short retroposons are repetitive 80 to 400 basepair (bp) sequences that are interspersed over the eukaryotic genome and are amplified through reverse transcription (Kramerov and Vassetzky, 2005; Ohshima and Okada, 2005). The mammalian genome contains 65–90% similar SINE families are derived from a cellular RNA; however, some SINE families also descended from 7SL RNA or 5S rRNA. SINES are transcribed by RNA polymerase III (Pol III) from an internal promoter in their 5’ region. The classical SINE promoter consists of two boxes (A and B) that are spaced 30–40 bp apart. SINES are considered nonautonomous mobile elements, since they do not encode proteins and utilize the reverse transcriptase of LINE-1.

**Fig. 1.** Alignment of nucleotide sequences of cDNA clones from golden hamster (Mesocricetus auratus), brown rat (R. norvegicus) and Russian mole rat (S. microphthalmus) (Gogolevskaya and Kramerov, 2002). Four variants of mole rat 4.5S, RNA are shown: S variant, major L1 variant, and minor L2 and L3 variants. Only nucleotides that differ from the hamster RNA sequence are indicated. Dashes indicate gaps. The nucleotide sequences of the PCR primers are shown above the alignment.

In the cited article, we followed the traditional classification considering mole rats and bamboo rats as two different families, Spalacidae and Rhizomyidae, respectively. 4.5S RNA was not studied in Rhizomyidae, but its genes/pseudogenes were found in the Rhizomyidae pruinosa genus.

2 SINE (Serdobova and Kramerov, 1998). In

2.1. DNA and RNA isolation and electrophoresis

Genomic DNA, ethanol-preserved tissue and live animals were kindly provided by the researchers listed in Supplementary Table 1. DNA was isolated from liver, kidney or muscle by incubation of the tissue with proteinase K followed by phenol/chloroform extraction. Total RNA was isolated from the liver by the guanidine isothiocyanate method (Chomczynski and Sacchi, 1987).

2.2. Cloning of B1 SINE

Zokor genomic DNA (5.0 μg) was digested with EcoRI and HindIII and then separated by 1% agarose gel electrophoresis. DNA fragments of 0.5–1.2 kb were collected by reverse electrophoresis on a DEAE membrane. DNA was eluted from the membrane and precipitated by ethanol using 10 μg of glycerol as a carrier. The isolated genomic fragments (0.5 μg) were ligated into 0.3 μg of pGEM3Z digested with EcoRI and HindIII and used to transform XL-1 Blue Escherichia coli. Colony hybridization was carried out at 60 °C in 4× SSC containing 0.5% SDS, 5× Denhardt’s solution, 0.1 mg/ml boiled herring sperm DNA and a 32P-labeled murine B1 probe (Vassetzky et al., 2003). Nitrocellulose filters were washed in 0.1× SSC containing 1% SDS at 42 °C and positive colonies were identified by autoradiography. B1-containing E. coli clones were purified by two additional rounds of colony hybridization.

2.3. Northern hybridization

RNA (30 μg) was separated by electrophoresis on 6% polyacrylamide gels containing 7 M urea. RNA was transferred onto a Hybond X1 membrane using a semi-dry electrophoretic apparatus (TE 77 PVR, Amersham Bioscience). Hybridization was carried out at 42 °C as described for cloning of B1 SINE, with the addition of 50% formamide to the hybridization solution. Preparation of the 32P-labeled 4.5S RNA-

<table>
<thead>
<tr>
<th>Hamster</th>
<th>Rat, Mouse</th>
<th>Mole Rat L-variant Type L1</th>
<th>Mole Rat L-variant Type L2 Ag</th>
<th>Mole Rat L-variant Type L2 Ag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GGGCCUGAGGAGGCUCAACCUCUAA</td>
<td>GGGCCUGAGGAGGCUCAACCUCUAA</td>
<td>GGGCCUGAGGAGGCUCAACCUCUAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 20 30 40 50 60</td>
<td>10 20 30 40 50 60</td>
<td>10 20 30 40 50 60</td>
</tr>
</tbody>
</table>

3 In the cited article, we followed the traditional classification considering mole rats and bamboo rats as two different families, Spalacidae and Rhizomyidae, respectively. 4.5S RNA was not studied in Rhizomyidae, but its genes/pseudogenes were found in the Rhizomyidae pruinosa genus.
Fig. 2. Alignment of B1 consensus sequences from different rodent families. The top row represents the common rodent B1 consensus sequence. The nucleotides that coincide with those in the formal consensus are marked by dots. Dashes indicate gaps. The alignment demonstrates the similarity of B1s from Spalacinae (Spal), Rhizomyinae (Rhiz), and Myospalacinae (Myos) and the difference from cricetid B1s (represented by subfamilies Cri-A0, Cri-A2, and Cri-B1). Eight positions that distinguish cricetid B1s and spalacine/rhizomyine/myospalacine B1s are indicated by arrows and are shaded grey and black, respectively. Other sequences are designated according to the three first letters of the family name as follows: Dipodidae (Dip), Zapodidae (Zap), Anomaluridae (Ano), Pedetidae (Ped), Geomyidae (Geo), Castoridae (Cas), Ctenodactylidae (Cte), Hystricidae (Hys), Thryonomyidae (Thr), Dasyproctidae (Das), Hydrochoeridae (Hyd), Caviidae (Cav), Chinchillidae (Chi), Octodontidae (Oct), and Myocastoridae (Myo). Other consensus sequences: Mus, M. musculus; Rat-A and Rat-B, R. norvegicus; and subfamily Gerbillinae (Ger). All consensus sequences except B1_Myos are from (Veniaminova et al., 2007b).
specific probe was performed as previously described (Gogolevskaya and Kramerov, 2002).

2.4. PCR, cloning, and data analysis

For the detection of 4.5S RNA cDNAs, genes and pseudogenes, PCR was carried out using rodent genomic DNA (10 ng) or cDNAs as the template and oligonucleotide primers specific for the S- and L-variants of mole rat 4.5S RNA (Fig. 1). cDNAs were synthesized from the 90–110 nucleotide fraction of liver RNA isolated as described for Northern hybridization. For PCR, the thermocycle parameters were as follows: 35 cycles of 95 °C for 30 s, 53 °C for 30 s, and 72 °C for 30 s. Amplified products were isolated by electrophoresis on gels containing 3% NuSieve agarose (Promega) and then cloned into pGEM-T (Promega). All clones were verified by sequencing with standard M13 primers using a BigDye Terminator sequencing kit and an ABI Prism 3100-Avant sequencer (Applied Biosystems). The nucleotide sequences of the cloned DNA fragments were deposited in GenBank under the following accession numbers: B1 sequences, FJ184041–FJ184056; 4.5S RNA sequences, GU066799–GU066805.

Multiple alignments using the ClustalW program were manually adjusted in GeneDoc. Sequence identity was determined using the GeneDoc program. Phylogenetic trees were constructed by the maximum parsimony method with a randomized order of sequences (10 times) and the bootstrap procedure (1000 replications) with the help of the MEGA4 program (Tamura et al., 2007).
were constructed using the Bayesian method (MrBayes 3.1, model GTR+I+Γ, 10^7 generations without the first 17%).

3. Results

3.1. Structure of zokor genomic B1 SINE

A zokor (Myospalax psilurus) genomic library was screened by hybridization with a radiolabeled probe specific for mouse B1 SINE. Seventeen randomly chosen positive clones were sequenced. B1 sequences were identified by alignment with the mouse B1 consensus sequence. From this multiple alignment (Supplementary Fig. 1), we were able to deduce a consensus sequence for M. psilurus B1. Fig. 2 shows the consensus M. psilurus B1 sequence (B1_Myo) in comparison to the B1 consensus sequences from rodents of various other families (Veniaminova et al., 2007b). Zokor B1 SINE was similar to B1 elements from Spalax microphthalmus (Spalacidae) and Tachyoryctes splendens (Rhizomyidae), the B1 sequences of these three rodents differing from the cricetid B1 subfamilies (Cri-A0, Cri-A2, and Cri-B1) by at least eight positions. Bayesian analysis of the consensus B1 sequences supported a monophyly of Myospalax, Spalax and Tachyoryctes (Supplementary Fig. 2).

3.2. 4.5S RNA and zokor phylogeny

Previously, we identified two variants of 4.5S RNA, the S- and L-variants (98 and 104 nucleotides, respectively), in the mole rat S. microphthalmus (Gogolevskaya and Kramerov, 2002). The L-variant (4.5S, RNA-L) was absent from house mouse and golden hamster. Northern blot analysis of liver RNA isolated from T. splendens and M. psilurus showed that both 4.5S-L and -S are present not only in S. microphthalmus, but in T. splendens and M. psilurus as well (Fig. 3), confirming that there is a relationship between these species.

The variants of 4.5S RNA were further studied in T. splendens and M. psilurus using the following experiment. cDNAs were synthesized using 90–110 nt RNA fraction isolated from the liver of S. microphthalmus, T. splendens, and M. psilurus as template and oligonucleotides complementary to 3′–5′end region of two 4.5S RNA variants as primers. This cDNA was used in PCR with the primers specific for the 5′-ends of S and L 4.5S RNAs. Amplified DNA fragments were cloned and then 15–17 clones per rodent species were sequenced.

Sequence analysis revealed that both the short and long 4.5S RNA variants were present in all three species. Most of the 4.5S-S variants were identical (Fig. 4a), with the following exceptions: (i) clone cTsp_49_S, which contained seven single-nucleotide substitutions; (ii) clone cTsp_38_S, which contained 4 additional A residues in a five-A residue block; and (iii) eight of nine clones of cSmi_N_S (S. microphthalmus), which contained a G instead of T at position 32 (98% identity).

The 4.5S-L RNA sequences exhibited more variability than the S-variant. This variability mapped to five positions (nucleotides 19, 50, 53, 63, and 80) (Fig. 4b). Based on the identity of the nucleotides at these positions, the L-variant was divided into two subvariant classes, type I and type II. The major/minor nucleotide distributions of the five diagnostic positions was as follows (upercase and lowercase letters represent major and minor nucleotides, respectively): subvariant I: (G/a)19, (A/g/t)50, (G/t)53, (G/a)63, and (A)80; subvariant II: (A/g)19, C50, (A/t)53, A63, and (–)80. Of note, both types of 4.5S-L RNA were present in all three rodent species (Fig. 4b), further supporting a sister relationship.

3.3. Genomic sequences similar to 4.5S RNA confirm the relation of zokors with mole and bamboo rats

Since RNA from exotic animals is often unavailable and can be more difficult to analyze than DNA, we examined whether genomic DNA could be used to detect specific nucleotide sequences with a narrow distribution (in this case, different variants of 4.5S RNA genes/pseudogenes) as phylogenetic markers. To this end, PCR using primers specific for 4.5S-S and -L RNA was performed. Amplified fragments of the expected length were detected only in S. microphthalmus, T. splendens and M. psilurus (Fig. 5). The absence of 4.5S-L RNA genes/pseudogenes in M. musculus and R. norvegicus was also supported by screening of available genomic databases.

The amplified genomic fragments from S. microphthalmus, T. splendens and M. psilurus were cloned and sequenced. S-variant sequences amplified from genomic DNA exhibited greater nucleotide sequence heterogeneity (80–100% identity) than those amplified from cDNAs (Fig. 6a). This higher level of heterogeneity could be due to the amplification of 4.5S RNA pseudogenes. In M. psilurus, there were two groups of genomic S-variant sequences: type 1 corresponded to the 4.5S-S RNA, whereas type 2 sequences contained nine specific single-nucleotide substitutions. Type 2 sequences may represent non-transcribed or poorly transcribed sequences. One of the T. splendens cDNA clones (cTsp_49_S, Fig. 4a) was a type 2 sequence, which supported the idea that these sequences can be transcribed.

The consensus sequences of the genomic 4.5S-S RNAs of S. microphthalmus, T. splendens, and M. psilurus were very similar to each other and to the S-variant of S. microphthalmus 4.5S RNA (98–100% identity), but were clearly different from hamster 4.5S RNA at three positions (marked by arrows in Fig. 7a; 92–94%).

The genomic L-variant sequences were characterized by an even higher heterogeneity (73–100% identity) as compared to the corresponding cDNAs (91–100%) and more so to S-variant sequences.

Fig. 5. PCR analysis of rodent genomic DNA using primers specific for the S (a) and L (b) variants of mole rat 4.5S RNA. S-specific primers: SpaSdir and SpaSrev; L-specific primers: SpaLdir and SpaLrev1 (see Fig. 1). Arrows indicate amplified products of the expected length. Mmm, M. musculus (house mouse); Mau, M. auratus (golden hamster); Pda, Pitymys daghestanicus (Daghestan vole); Eta, Ellobius talpinus (mole-vole); Smi, S. microphthalmus, (Russian mole rat); Tsp, T. splendens (root rat); Mps, M. psilurus (zokor); Ama, A. major (great jerboa).
These results could reflect a high level of diversity of the L-variant and/or by the presence of pseudogenes. The L-variant genomic consensus sequences of \textit{T. splendens} and \textit{M. psilurus} were much more similar to each other and to the L-variant of mole rat than to hamster 4.5SI RNA (Fig. 7b). The differences in RNA structure in the particular region analyzed in this study from the hamster sequence included nine single-nucleotide positions (marked by arrows in Fig. 7b), as well as a 6–7 nucleotide insertion. Parsimony analysis of the 4.5SI RNA sequences also supported a monophyletic relationship for \textit{Myospalax}, \textit{Spalax} and \textit{Tachyoryctes} (Supplementary Fig. 3). Taken together, the results of the current study, using both genomic DNA and cDNAs, clearly support a close relationship of Myospalacinae with Spalacinae and Rhizomyinae, rather than with Cricetidae.

4. Discussion

4.1. \textit{B1 SINE} as a molecular marker for rodent phylogenetic studies

We studied \textit{B1 SINE} nucleotide sequences to clarify the evolutionary relationships of zokors (Myospalacinae). The approach was based on the hypothesis that \textit{B1} subfamilies evolved congruently with their host species. That is, the emergence of new rodent groups was accompanied by the emergence of successful new \textit{B1} subfamilies that evolved from previously existent subfamilies. These new subfamilies, often with old diagnostic traits, acquired new ones. Thus, tracing the evolution of \textit{B1} subfamilies allows one to draw conclusions about rodent phylogeny. Mobile elements such as \textit{B1} sequences appear to be
nonfunctional, which makes them particularly good phylogenetic markers.

Comparison of the consensus B1 sequence from *M. psilurus* with consensus sequences from other rodent families (subfamilies) showed that *M. psilurus* B1 SINE was most similar to *T. splendens* and *S. microphthalmus* (Fig. 2). There were at least three single-nucleotide substitutions (at positions 108, 109, and 121) that were specific to these three species. Five other positions (8/9, 70, 115, 116, and 124) distinguished B1 of these three species from cricetids. Cricetidae was represented by three B1 subfamilies, two of which (Cri-A0 and Cri-A2) had a unique trinucleotide insertion (CCA). None of the 17 B1 sequences from *M. psilurus* contained this insertion. These results argue against a close relationship between Myospalacinae and Cricetidae, but confirm recent results (Norris et al., 2004; Jansa and Weksler, 2004) supporting a relationship between Myospalacinae and Spalacinae and Rhizomyinae.

4.2. The structure of 4.5SI RNA, its genes, and pseudogenes suggests a sister relationship between Myospalacinae, Spalacinae and Rhizomyinae

It has been shown previously that 4.5S RNA genes are absent from the majority of rodent families, with the exception of Muridae, Cricetidae and Spalacidae, and most likely Rhizomyidae (Gogolevskaya and Kramerov, 2002). Due to the narrow distribution pattern of 4.5S RNA, it is referred to as a stenoRNA (from the Greek “stenos” meaning narrow), along with other RNAs such as 4.5Sr, BC1 and BC200 (Gogolevskaya and Kramerov, 2002). Such RNAs arose relatively recently (some tens of millions years ago) and their genes were passed on from a progenitor to all descendant species. As such, these RNAs and the genes that encode them can serve as phylogenetic markers (Martignetti and Brosius, 1993; Gogolevskaya and Kramerov, 2002; Kramerov and Vastskii, 2009).

Northern blot analysis of liver RNA and sequencing of PCR-amplified cDNAs revealed the presence of two variants of 4.5S RNA, 4.5S-S and 4.5S-L, in Russian blind mole rat (Spalacidae) and root rat (a member of the bamboo rat family Rhizomyidae), as well as in zokor. The L-variant was absent from house mouse (Muridae), golden hamster (Cricetidae) and great jerboa (Dipodidae), whereas the S variant was absent only from Dipodidae. The nucleotide sequences of the S-variants from Spalacidae, Rhizomyidae and Myospalacinae were very similar; however, they differed considerably from those of mouse and hamster (Fig. 7a). This distribution of 4.5S RNA among rodent species was supported by PCR analysis of genomic DNA (Fig. 5) and sequencing of cloned amplified PCR products (Figs. 6 and 7). The data support a sister relationship between zokors and mole rats and bamboo rats (Norris et al., 2004; Jansa and Weksler, 2004), but not hamsters (Simpson, 1945; Michaux and Catzeffis, 2000; Michaux et al., 2001). These relationships are depicted as a tree in Fig. 8. Thus, zokors can be considered a subfamily (Myospalacinae) together with the subfamilies of Spalacinae and Rhizomyinae in the Spalacidae family. The similarities in many morphological traits among zokors, mole rats and bamboo rats is most likely due to their common origin rather than convergence caused by a common subterranean way of life.

4.5SI RNA

<table>
<thead>
<tr>
<th>S variant</th>
<th>L variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muridae</td>
<td>+</td>
</tr>
<tr>
<td>Cricetidae</td>
<td>+</td>
</tr>
<tr>
<td>Spalacinae</td>
<td>+</td>
</tr>
<tr>
<td>Rhizomyina</td>
<td>+</td>
</tr>
<tr>
<td>Myospalacinae</td>
<td>+</td>
</tr>
<tr>
<td>Dipodidae</td>
<td>–</td>
</tr>
</tbody>
</table>

Fig. 7. Comparison of the nucleotide sequences of 4.5S RNAs of golden hamster (cMau4.5SI) and mole rat (cSmi4.5S_S and cSmi4.5S_L) with consensus sequences derived from cDNAs and genomic DNA fragments amplified using primers specific for the S (a) and L (b) variants. «c» and «g» refer to cDNA and genomic DNA, respectively. Smi, Tsp, and Mps are the sequences of *S. microphthalmus*, *T. splendens*, and *M. psilurus*, respectively. The arrows indicate the nucleotide positions that distinguish hamster 4.5S RNA from the S variant and L1 subvariant of the mole rat.

Fig. 8. The distribution of S and L variants of 4.5S RNA among Myodonta rodents confirms the sister relationship of Spalacinae, Rhizomyinae, and Myospalacinae. “+” and “−” denote the presence and absence of these RNAs in representatives of the families/subfamilies, respectively. The arrows indicate the putative time of emergence of the two 4.5S RNA variants. The relationships among Myodonta families have been confirmed by a number of studies (for example, Huchon et al., 2002; Steppan et al., 2004; Veniaminova et al., 2007b). The divergence order of Spalacidae subfamilies is depicted in accordance with Norris et al. (2004) and Jansa and Weksler (2004).
Appendix A. Supplementary data


References


Mishin, A., 1969. The phylogenetic position of the zokors (Myospalacinae) and comments on the phylogenetic position of the zokors (Myospalacinae) and comments on the phylogenetic position of the zokors (Myospalacinae). Mol. Biol. Evol. 19, 1053–1065.

