J Mol Evol (2002) 54:354–364
DOI: 10.1007/s00239-001-0050-9

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary History of $4.5S_I$ RNA and Indication That It Is Functional

Irina K. Gogolevskaya, Dmitri A. Kramerov

Enhelgardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov str., Moscow, 117984, Russia

Received: 11 December 2000 / Accepted: 27 August 2001

Abstract. To date, the small nuclear $4.5S_I$ RNA has only been studied in the rat (*Rattus norvegicus*). Combining PCR and hybridization analyses, we have revealed $4.5S_I$ RNA homologues sequences in the genomes of four myomorph rodent families (Muridae, Cricetidae, Spalicidae, and Rhizomyidae), and not in other myomorph families (Dipodidae, Zapodidae, Geomyidae, and Heteromyidae) or sciuromorph and caviomorph rodents. By Northern-hybridization, $4.5S_I$ RNA has been detected in the common rat (*R. norvegicus*, Muridae), golden hamster (*Mesocricetus auratus*, Cricetidae), and Russian mole rat (*Spalax microphthalmus*, Spalacidae), but not in the related great jerboa (*Allactaga jaculus*, Dipodidae) or in four non-myomorph rodent species tested. cDNA derived from $4.5S_I$ RNA of *M. auratus* and *S. microphthalmus* has been cloned and sequenced. The hamster RNA is found to differ from rat $4.5S_I$ RNA by only one nucleotide substitution. For the mole rat, two variants of $4.5S_I$ RNA are detected: short (S) and long (L) with length 101 and 108 nt, respectively. The L variant differs from the S variant as well as from murid and cricetid $4.5S_I$ RNAs by both a 7 nt insertion and a varying number of nucleotide substitutions. The sequence similarity between the spalacid S-variant and murid/crecitid variants of $4.5S_I$ RNA is 90%. Judging from species distribution, $4.5S_I$ RNA genes emerged during the same period of time as the related short interspersed element B2 arose. This occurred after the divergence of Dipodidae lineage but before the branching of Spalicidae/Rhizomyidae lineage from a common myomorph rodent stem. S variant genes seemed to emerge in a common ancestor of spalacids and rhizomyds whereas L variant genes formed in spalacids following the divergence of these two families. The low rate of evolutionary changes of $4.5S_I$ RNA, at least, in murids and cricetids (6×10^{-4} substitutions per site per million years), suggests that this RNA is under selection constraint and have a function. This is a remarkable fact if the recent origin and narrow species distribution range of $4.5S_I$ RNA genes is taken into account. Genes with narrow species distribution are proposed to be referred to as stenogenes.

Key words: snRNA — 4.5S RNA I — Repetitive DNA — SINE — B2 — Rate of substitution — Rodentia — Muridae — Crecitidae — Spalacidae

Introduction

Currently, about a hundred of small RNAs (except tRNAs) have been found in mammalian cells (Eddy 1999; Maxwell and Fournier 1995; McKeown 1993). Many of them play an important role in such processes as pre-mRNA splicing (U1, U2, U4, U5, U6 RNAs), rRNA processing and modification (C/D box RNAs and H/ACA box RNAs) as well as protein secretion (7SL RNA). Small RNAs were also described for various eukaryotes including yeast. Most small RNAs are ubiquitous and have rather conserved nucleotide sequences, at least, among mammals. However, several species of small RNAs isolated from rodent cells were not found in mammals from other analyzed orders. The best studied among these RNAs is BC1 RNA that was isolated from different rodents including the rat, mouse and distantly

Correspondence to: D.A. Kramerov; *email:* kramerov@genome.eimb.relarn.ru

related guinea pig (DeChiara and Brosius 1987; Martignetti and Brosius 1993). However, BC1 RNA has not been detected in rabbit, cattle, and human. Primates, including humans, have an analog of BC1 RNA named BC200 (Tiedge et al. 1993). Both these RNAs are only characteristic of nervous tissue. Their function remains unknown. It is likely that BC1 and BC200 RNA genes emerged in direct ancestors of rodents and primates, respectively, and were involved in nerve cell functioning.

The other two RNAs found in rat and mouse cells, but not in human, rabbit or dog cells, were described in the seventies and named $4.5S_I$ RNA_H (Harada and Kato 1980; Harada et al. 1979) and $4.5S_I$ RNA (or 4.5S RNA_I) (Reddy et al. 1983, Ro-Choi et al. 1972). Each RNA is about 100 nucleotides in length. In contrast to BC1, both RNAs are present in many if not all rat tissues. Some molecules of 4.5S RNA_H and $4.5S_I$ RNA were detected in association with polyadenylated and nonpolyadenylated nuclear RNAs, respectively (Harada et al. 1979; Miller et al. 1984; Schoeniger and Jelinek 1986). However, the significance of these two small RNAs needs to be proved and their absence in human and rabbit may indicate the absence of functions. The rat 4.5S RNA_H gene is a part of a 5.3 kb unit, 690 copies of which are arranged as tandem repeats (Schoeniger and Jelinek 1986), whereas the $4.5S_I$ RNA genes and pseudogenes are dispersed throughout the genome and the number of pseudogenes considerably exceeds that of genes (Saba et al. 1985; Takeuchi and Harada 1986). The copy number of genes for $4.5S_I$ RNA is unknown, but the total number of its genes and pseudogenes is estimated as 500 copies per haploid rat genome.

The small RNAs considered above share two important features: they are synthesized by RNA polymerase III and demonstrate a partial sequence similarity to either rodent or primate short interspersed elements or SINEs. SINEs are 80–400 bp repetitive DNA sequences that proliferate in eukaryotic genomes via transcription followed by reverse transcription; usually the number of SINE copies ranges from 10^3 to 10^5, such copies feature age-dependent sequences variability (from a few percent to 35%) (Okada 1991). The 5′ end parts of BC1 RNA (DeChiara and Brosius 1987) and BC200 RNA (Martignetti and Brosius 1993) are virtually identical to SINEs ID and Alu, respectively. It is most likely that BC1 RNA is an evolutionary progenitor of ID elements whereas BC200 RNA originated from an Alu monomeric element (Brosius 1999), The nucleotide sequences of 4.5S RNA_H show an extensive homology to the rodent B1 element (Kramerov et al. 1982; Krayev et al. 1982). $4.5S_I$ RNA is also very similar to the first 23 nt of the B2 element and moderate similarity can be observed up to position 80 (Krayev et al. 1982; Saba et al. 1985). This suggests that there are close evolutionary relations between those small RNAs and SINEs, although it remains unclear in most cases which are progenitors. In particular, evolutionary relations between $4.5S_I$ RNA and the B2 element need to be clarified.

Here we first studied the distribution of species having $4.5S_I$ RNA which proved to be identical to that for the B2 element. The nucleotide sequences of $4.5S_I$ RNA were determined in representatives of two rodent families (Cricetidae and Spalacidae) not yet studied in this respect. It was found that $4.5S_I$ RNA was conserved during evolution, especially in certain lineages. This finding suggests that $4.5S_I$ RNA must be functional, despite its recent origin.

Materials and Methods

DNA and RNA Isolation and Electrophoresis

Genomic DNAs, ethanol-preserved tissues, or live animals were provided by various researchers (Table 1). DNA was isolated from liver, kidney or muscle by incubation with proteinase K followed by phenol/chloroform extraction. Total RNA was isolated from fresh liver by the guanidine isothiocyanate method (Chomczynski and Sacchi 1987). RNA was fractionated by electrophoresis in 1.5% agarose gels with 6% formaldehyde and transferred by capillary blotting onto a Hybond N membrane (Amersham). Alternatively, RNA electrophoresis was performed in 6% PAAG containing 7M urea. RNA was transferred onto a Hybond N membrane using an apparatus for semi-dry electroblotting (TE 70, Hoefer).

Genomic Analysis by PCR

PCR was used to detect genes and pseudogenes of $4.5S_I$ RNA. Oligonucleotides (Fig. 4) homologous to the beginning (*Mur.dir 1*) and complementary to the end (*Mur.rev*) of the rat $4.5S_I$ RNA gene were used in this assay. 27 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min were performed. PCR products were analyzed by electrophoresis in a gel containing 3% NuSeave agarose (FMC) and 1% common agarose (Sigma). To label hybridization probes, 30 μCi [α^{32}P]dATP was added in reaction mixture for PCR and 15 cycles were run as described above, but with the 72°C incubation for 3 min. Probe I was synthesized with the use of primers *Mur.dir 1* and *Mur.rey* Probe II which is more specific for detection $4.5S_I$ RNA and corresponds to its 3′ half was synthesized using primer *Mur.dir 2* (Fig. 4) instead of *Mur.dir 1*. Blot and dot hybridizations were performed as described previously (Serdobova and Kramerov, 1998).

cDNA Synthesis and Cloning.

Total liver RNA (150 μg) was fractionated by electrophoresis in 6% PAAG with 7M urea and the material from the gel zone containing $4.5S_I$ RNA (according to size markers) was transferred onto a NA45 DEAE membrane (Schleiher and Schuell) by semi-dry electroblotting. To determine the exact position of $4.5S_I$ RNA in the gel by hybridization, RNA from the analogous gel zone was transferred onto a Hybond N membrane. Then RNA was eluted from the appropriate piece of the DEAE membrane by incubation with 1M NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA at 60°C for 30 min. The isolated RNA was polyadenylated using *E. coli* poly(A) polymerase (Gibco BRL). The ^{32}P-labeled cDNA was synthesized with reverse transcriptase (Super Script RT, Gibco BRL) with primer *XbaI*$(T)_{15}$: 5′GTCGACTCTAGA$(T)_{15}$-3′. After purification by electrophoresis in 4% NuSeave agarose gel

Table 1. Rodent species[a] analyzed

Family	Code	Scientific name	Common name	Collector
Muridae	Mmu	*Mus musculus*	House mouse	EIMB[b]
	Rno	*Rattus norvegicus*	Common rat	EIMB[b]
	Ape	*Apodemus peninsula*	Wood mouse	N. Bulatova[c]
Cricetidae	Mso	*Microtus socialis*	Social vole	E. Lyapunova[d]
	Pda	*Pitymys daghestanicus*	Daghestan vole	E. Lyapunova[d]
	Ate	*Arvicola terrestris*	Water vole	T. Golovkina[e]
	Tin	*Tatera indica*	Indian gerbil	E. Lyapunova[d]
	Eta	*Ellobius tancrei*	Mole-vole	E. Lyapunova[d]
	Msc	*Myopus schisticolor*	Wood lemming	T. Golovkina[e]
	Mau	*Mesocricetus auratus*	Golden hamster	EIMB[b]
Spalacidae	Smi	*Spalax microphthalmus*	Russian mole rat	A. Puzachenko[f]
Rhizomydae	Rpr	*Rhizomyis pruinosus*	Hoary bamboo rat	MVZ[g]
Dipodidae	Abo	*Allactodipus bobrinskii*	Bobrinski's jerboa	G. Shenbrot[c]
	Apy	*Alactagulus pygmaeus*	Lesser five-toed jerboa	G. Shenbrot[c]
	Eli	*Erimodipus lichtensteini*	Lichtensten's jerboa	G. Shenbrot[c]
	Aja	*Allactaga jaculus*	Great jerboa	D. Kramerov
Zapodidae	Sti	*Sicista tianschanica*	Birch mouse	E. Ivanitskaya[c]
Geomyidea	Tbo	*Thomomys bottae*	Botta's pocket gopher	F. Catzeflis[h]
Heteromyidea	Dde	*Dipodomys deserti*	Desert kangaroo rat	R. DeBry[i]
	Ppa	*Perognathus parvus*	Grate basin pocket mouse	MVZ[g]
Sciuridae	Mbe	*Menetes berdmorei*	Palm squirrel	E. Panova[c]
	Cfu	*Citellus fulvus*	Large-toothed suslik	E. Lyapunova[d]
	Mca	*Marmota caudata*	Long-tailed marmot	E. Lyapunova[d]
	Sca	*Sciurus carolinensis*	Gray squirrel	E. Lyapunova[d]
Gliridae	Dni	*Dryomys nitedula*	Forest dormouse	E. Ivanitskaya[c]
Castoridae	Cfi	*Castor fiber*	Beaver	O. Likhnova[c]
Caviidae	Cpo	*Cavia porcelus*	Guinea pig	EIMB[b]

[a] Nonrodent species analyzed: Ocu—*Oryctolagus cuniculus*, rabbit; Cfa—*Canis familiaris*, dog; Bta—*Bos taurus*, calf; Hsa—*Homo sapiens*, man.
[b] Enhelgardt Institute of Molecular Biology, animal house.
[c] Institute of Ecology and Evolution, Moscow.
[d] Institute of Developmental Biology, Moscow.
[e] Russian Oncological Center, Moscow.
[f] Institute of Nature Conservation and Reserves, Moscow.
[g] Museum of Vertebrate Zoology at the University of California, Berkeley; USA.
[h] Institut des Sciences de l'Evolution, Universite Montpellier 2, France.
[i] University of Cincinnati, USA.

and electroelution on a DEAE membrane, the 3′ end of cDNA was tailed with poly(dG) using terminal deoxynucleotidyl transferase (Amersham). cDNA was then amplified by 30 cycles of PCR with primers *XbaI*-$(T)_{15}$ and *Eco*RI$(C)_{10}$-5′CGGAATTCGT$(C)_{10}$3′. The resultant double-stranded cDNA was digested with *XbaI* and *Eco*RI, purified by electrophoresis in 4% NuSeave agarose gel, and cloned in a plasmid that had been cut with the same enzymes. The library was screened by hybridization with $4.5S_I$ RNA specific probe II (see above). Double-stranded plasmid templates were sequenced using the dideoxynucleotide method with Sequenase 2.0 (USB-Amersham).

Results

Investigating the Species Distribution of $4.5S_I$ RNA and Its Related Genomic Sequences

To reveal how widely the DNA sequences homologous to $4.5S_I$ RNA are distributed among representatives of the order Rodentia, genomic DNAs of 20 species from nine rodent families (Table 1) were analyzed by dot hybridization using a probe corresponding to the 3′ half of rat $4.5S_I$ RNA. Under the achieved sensitivity of hybridization, the genomic sequences homologous to $4.5S_I$ RNA (genes and pseudogenes) were detected in rodents of only three families—Muridae, Cricetidae, and Spalacidae (Fig. 1). It should be noted that while the intensities of signals were comparable for all cricetids tested, a great difference in signal strength was observed for muride species: from lack of signal in the case of *Mus musculus* (Mmu) to a very intense signal in the case of *Apodemus peninsula* (Ape). This result suggests that copy numbers of the sequences studied may considerably differ in the muride genomes.

The results of dot-hybridization cannot rule out the existence of a small number of $4.5S_I$-like sequence copies in the genome of the species (e.g. *Mus musculus*) for which no hybridization signals were observed. To increase the sensitivity we used two variations of PCR analysis. In the first PCR, an constant amount (1 ng) of genomic DNAs was added to reaction mixtures regardless of the species studied. In this case (results not shown), a PCR product of proper length was observed

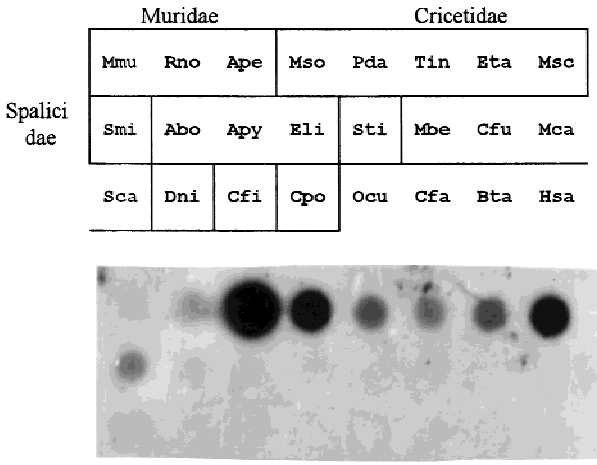


Fig. 1. Dot hybridization of genomic DNA of rodents and other mammals with a probe specific to the rat $4.5S_I$ RNA (probe II). The filter was washed at 42°C. In the scheme showing the positions of the genomic DNA dots, three-letter abbreviations for scientific names of species (see Table 1) are used, and each rodent family is framed.

only in DNA of cricetids and murids (including *M. musculus* which showed no signal in dot hybridization; see above). In the second PCR variant (Fig. 2), 200 ng of genomic DNA was used to analyze those species for which no PCR product was observed in the first PCR variant (for murids and cricetids the initial amount, 1 ng, of DNA remained). The proper PCR product was generated not only with murid and cricetid but with mole rat *Spalax microphthalmus* DNA as well. This result demonstrates the presence of $4.5S_I$ RNA homologous sequences in the spalacid genome. No proper product was detected by the PCR analysis of any other rodent families tested (Dipodidae, Zapodidae, Sciuridae, Gliridae, and Caviidae). Blot-hybridization of the PCR products with the rat $4.5S_I$ RNA gene probe (Fig. 2), which additionally enhances the sensitivity of the method, confirmed the presence of sequences studied only in murids, cricetids, and spalacids. In addition, the comparison of hybridization signals obtained after low-stringency (42°C) and high-stringency (52°C) washes suggests a high level of sequence similarity between $4.5S_I$ RNA-related sequences in the genome of murids, cricetids, and spalacids (Fig. 2). The hybridization signals observed in rodents from other families can be explained by the fact that the same short nucleotide sequences are present in both long heterogeneous PCR products and in the hybridization probe that correspond to the primers used for their synthesis. Most likely, such nonspecific PCR products are amplified on the neighboring copies of the B2-related SINEs (Serdobova and Kramerov, 1998).

The titration PCR assay showed that the copy number of $4.5S_I$ RNA-related sequences capable of acting as a template for the primers is 50–100 times less in the mole rat (spalicid) than in the common rat genome (data not shown). The dot-hybridization (Fig. 1) indicates that the number of $4.5S_I$ RNA-related sequences in the spalicid genome is comparable to the number in the genomes of murids and cricetids. The vast majority (98–99%) of $4.5S_I$ RNA-related sequences in the mole rat genome seems to be quite different from those of the common rat, at least, in the region homologous to one of the primers used in PCR.

In order to corroborate the above described taxonomic distribution of $4.5S_I$ RNA genes and pseudogenes, the northern-blot hybridization of RNA from livers of eight rodent species was performed. Total cellular RNA was separated by agarose gel electrophoresis, blotted, and then hybridized under low-stringency conditions to rat $4.5S_I$ RNA gene probe (Fig. 3A). Strong hybridization signals were observed in the case of the house mouse, golden hamster, and mole rat. Not even trace amounts of $4.5S_I$ RNA could be detected in the great jerboa (Dipodidae), long-tailed marmot, large-toothed suslik, palm squirrel (Sciuridae), or guinea pig (Caviidae), Thus, $4.5S_I$ RNA is specific for Muridae, Cricetidae, and Spalacidae. The absence of this RNA from a representative of the myomorph family Dipodidae, related to Muridae, Cricetidae, and Spalacidae (Romer, 1966) suggests that numerous (about 30) rodent families more distantly related to them also lack it. To further resolve the $4.5S_I$ RNA, total liver RNA from the species studied in the above experiment was fractionated by PAAG electrophoresis, blotted, and hybridized with the $4.5S_I$-specific probe (Fig 3B). In house mouse, golden hamster, and mole rat RNA, an appropriate band (about 100 nt) was detected. An additional hybridization band (110 nt RNA) was observed in the case of mole rat, but this signal disappeared following a higher-stringency (50°C) wash. The standard and longer RNAs from mole rat were referred to as $4.5S_I$ RNA-S and $4.5S_I$ RNA-L, respectively.

Comparative Analysis of $4.5S_I$ RNA Nucleotide Sequences

The cDNA derived from $4.5S_I$ RNA of the common rat, golden hamster and mole rat were cloned and sequenced (Fig. 4). We found that all analyzed molecules of this RNA have three G residues on the 5′ end rather than only two G residues observed in the rat $4.5S_I$ RNA nucleotide sequence from fingerprinting (Ro-Choi et al. 1972; Reddy et al. 1983). Sequencing of cDNA confirmed the existence of two variants, A and U, of rat $4.5S_I$ RNA which are distinguished by the only nucleotide substitution (position 49). Hamster $4.5S_I$ RNA-derived cDNA clones (three full-length and five truncated) were sequenced. The nucleotide sequences of these clones demonstrated homogeneity of the hamster $4.5S_I$ RNA and a high degree of similarity to common rat RNA. The only difference between the hamster and rat A-variant RNA is a substitution of a U for a C at position 51.

The short (standard) variant of mole rat $4.5S_I$ RNA

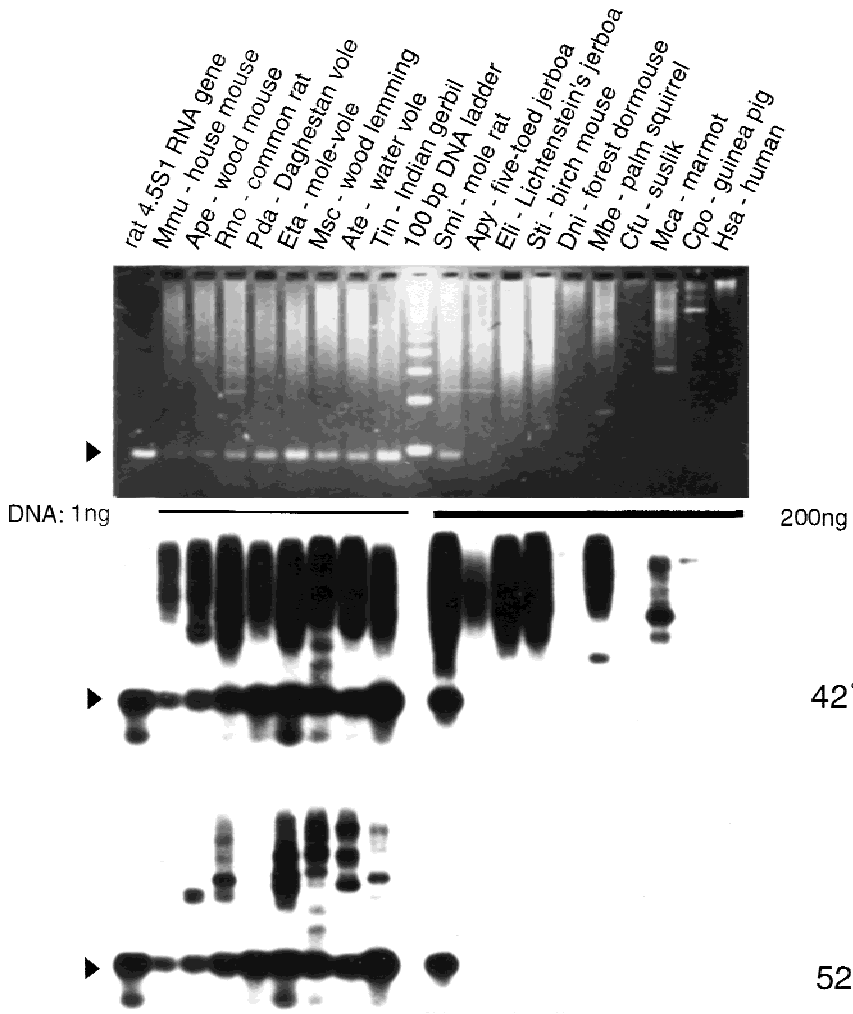


Fig. 2. Detection of $4.5S_I$ RNA-homologous sequences in rodent genomes by PCR followed by hybridization of the amplified fragments. **(Upper)** PCR products separated in a 4% agarose gel by electrophoresis and stained with ethidium bromide. Thin and thick horizontal lines mark 1 or 200 ng genomic DNA in the reaction, respectively. An arrowhead indicates the proper PCR product. **(Lower)** Blot-hybridization of PCR products with the rat $4.5S_I$ probe I. The filter was washed at 42°C and 52°C.

also showed sequence homogeneity (four full-length and six truncated cDNA clones were sequenced in this case). This $4.5S_I$ RNA-S differed from the common rat A-variant by eight substitutions and two single-nucleotide insertions (Fig. 4). The long variant of mole rat $4.5S_I$ RNA showed some sequence heterogeneity. Besides the major subvariant, $4.5S_I$ RNA-L1 (seven full-length cDNA clones), we found two other subvariants, $4.5S_I$ RNA-L2, and -L3, each of them represented by a single sequenced clone (Fig. 4). All $4.5S_I$ RNA-L molecules contained a 7-nucleotide insertion in the 3′ end region. Additionally, L1, L2, and L3 subvariants differed from rat $4.5S_I$ RNA-A by 16, 18, and 13 substitutions, respectively. There are only two mole rat specific positions (G_{17} and G_{51}) and six $4.5S_I$ RNA-L specific positions. Thus, the results of cDNA sequencing have finally proved the existence of $4.5S_I$ RNA in the hamster and mole rat. They have demonstrated the almost full identity of common rat and hamster $4.5S_I$ RNAs (99%) with rather high similarity between these RNAs in common and mole rats (90% and 83% for S- and L1- variants, respectively).

Finding of $4.5S_I$ RNA-S Related Sequences in the Rhizomyd Genome

The analysis of the mole rat $4.5S_I$ RNA nucleotide sequence revealed that the primers used for $4.5S_I$ homologous sequence detection in rodent genomes had poorly matched the mole rat $4.5S_I$ RNA genes (as shown in Fig. 4, there is a T residue on the 3′-end in primer *Mur.dir 1*, and a G residue in the corresponding position of mole rat RNA, the PCR product observed in the case of mole rat was obviously amplified from some pseudogene). We have carried out an additional experiment for PCR detection of $4.5S_I$ homologous sequences in a number of rodent genomes using primers *SpaS.dir* and *SpaS.rev* (Fig. 4) corresponding to S variant of mole rat $4.5S_I$ RNA. At the time of this experiment, the DNA of the representatives of three additional myomorph rodent families (Rhizomyidae, Geomyidae, and Heteromyidae) had become available, and were also used in the PCR analysis. Figure 5A shows that the specific ~90 bp PCR product was detected only in the case of mole rat and bamboo rat (*Rhizomyis pruniounsus*, Rhizomyidae), but

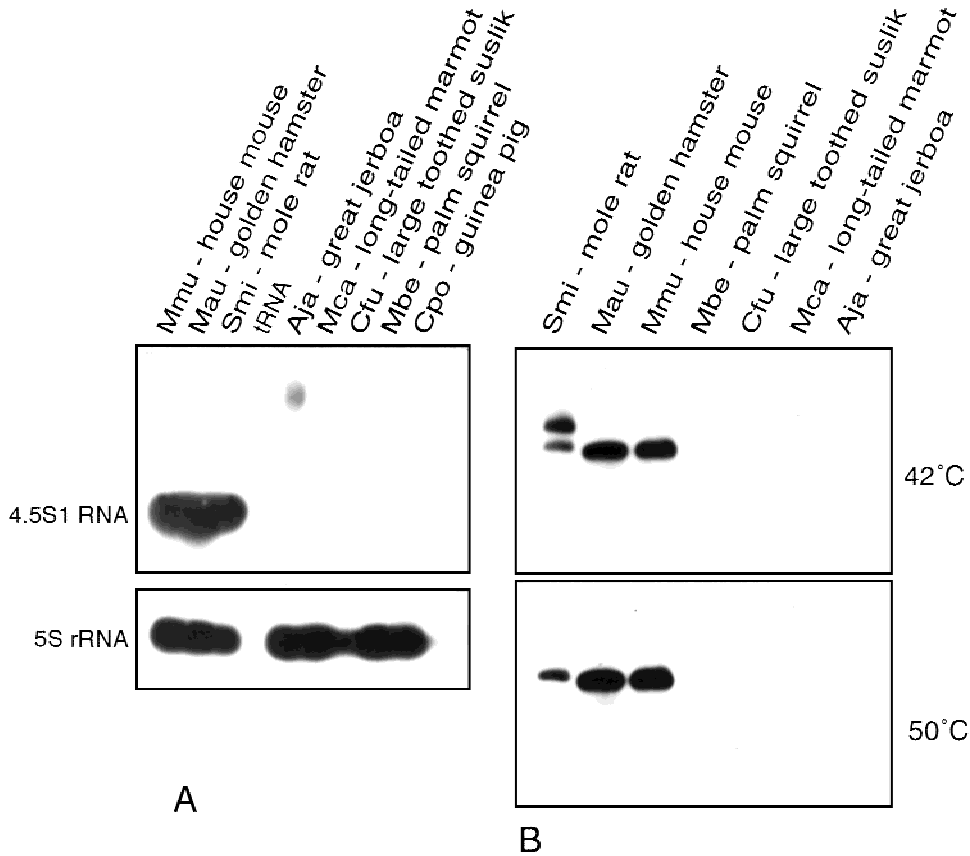


Fig. 3. Distribution of $4.5S_I$ RNA among six rodent families. **A. (Upper)** Detection of $4.5S_I$ RNA by blot-hybridization in total liver RNA separated by agarose gel electrophoresis. Filter was washed at 42°C. **(Lower)** Hybridization of the same filter with a probe specific to 5S rRNA: control for RNA amount. **B.** Detection of $4.5S_I$ RNA by blot hybridization of total RNA separated by PAAG electrophoresis. Filter was washed at 42°C and 50°C.

not in the case of other myomorph families tested (Muridae, Cricitidae, Dipodidae, Zapodidae, Geomyidae, and Heteromyidae).

To confirm that the observed PCR products do indeed correspond to $4.5S_I$ homologous sequences, we cloned and sequenced them (Fig. 5B). All sequences of both mole rat (11 clones) and bamboo rat (14 clones) demonstrated a high level of similarity to mole rat $4.5S_I$ RNA-S. Three mole rat clones (Sp21, 22, 25) were completely identical to the corresponding region of this RNA suggesting that they were amplified from the intact (functional) gene of $4.5S_I$ RNA-S. All other clones (except Sp5) of mole rat belonged to one of two kinds of sequences most likely corresponding pseudogenes. These results seem to reflect the existence of $4.5S_I$ RNA-S pseudogene subfamilies. At least four subfamilies of genes/pseudogenes were observed for the bamboo rat, but the sequences of four clones (Rh 8, 12, 15, and 33) did not refer to any of these subfamilies (Fig. 5B). The bamboo rat consensus of the PCR products was very similar to the mole rat consensus: only three nucleotide positions (22, 32, 62) differed. Moreover, these differences were not very significant. For each of these positions, there was the same nucleotide residue in several clones from both mole and bamboo rats (Fig. 5B). For example, in position 22, a C residue was in four mole rat PCR clones (Sp5, Sp21, Sp22, Sp25) as in the bamboo rat PCR consensus or in a functional $4.5S_I$ RNA-S gene. Thus, the obtained data suggest the presence of the $4.5S_I$ RNA genes/pseudogenes in the bamboo rat which are very similar to those in the mole rat. PCR has revealed no sequences homologous to $4.5S_I$ RNA-L in the bamboo rat genome (data not shown; see Fig. 4 for the primers used). This suggests that the L variant of the RNA is specific for mole rat.

Discussion

Using PCR, hybridization, and sequencing we have demonstrated that the $4.5S_I$ RNA homologous DNA sequences are characteristic of only four rodent families: Muridae (mice and rats), Cricetidae (hamsters, voles and gerbils), Spalacidae (mole rats), and Rhizomyidae (bamboo rats). These DNA sequences are not detected in the other four myomorph rodent families tested, or in more distantly related rodent families. Northern-hybridization and cDNA cloning unambiguously confirms the presence of $4.5S_I$ RNA in the common rat, golden hamster, and mole rat (the bamboo rat RNA has not yet been

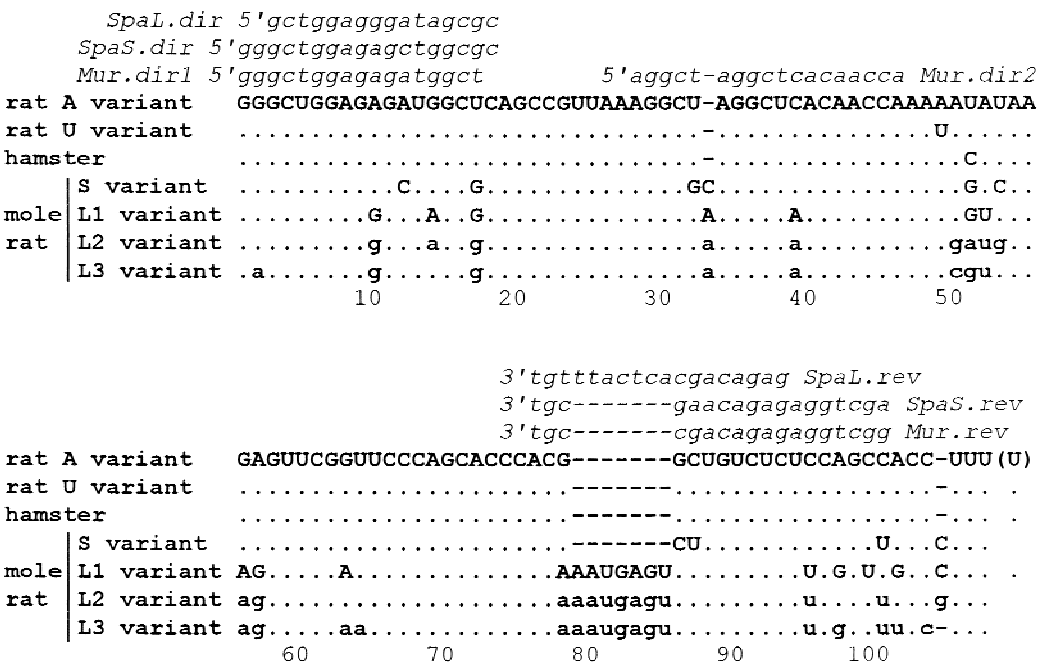


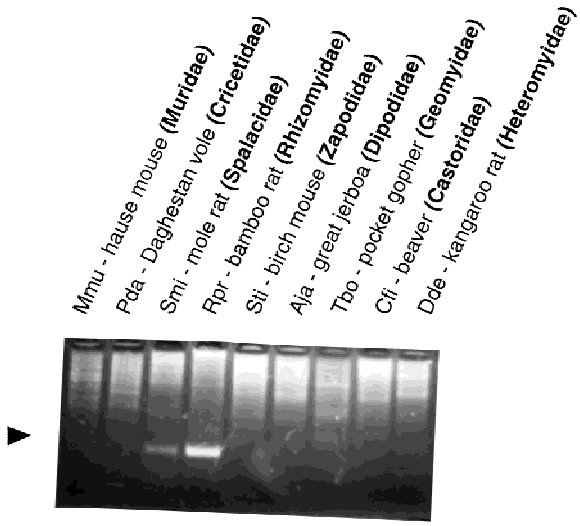
Fig. 4. Nucleotide sequences of $4.5S_I$ RNA from common rat (*Rattus norvegicus*), golden hamster (*Mesocricetus auratus*), and Russian mole rat (*Spalax microphthalmus*). Only nucleotides distinct from those in rat $4.5S_I$ RNA (A variant) are shown. Primers used for PCR detection of $4.5S_I$ RNA genes are given in italic: *Mur.dir 1* and *Mur.rev* are specific for the murid/cricetid variant, *SpaS.dir* and *SpaS.rev* for the S variant, *SpaL.dir* and *SpaL.rev* for the L variant. GenBank accession numbers: AF29888–AF298894.

analyzed because it is not available to us). This small RNA is absent even in jerboas (Dipodidae), a family most closely related to the murid/crecetid/spalacid/rhizomyd clade. These results allow us to conclude that $4.5S_I$ RNA genes first emerged after Dipodidae but before Spalacidae/Rhizomyidae clade divergence (Fig. 6). PCR analysis have revealed that bamboo rat genomic sequences are very similar to the mole rat S-variant of the $4.5S_I$ RNA gene. This suggests the emergence of the S variant gene in a common ancestor of spalacids and rhizomyds. L variant is detected only in the mole rat genome indicating its origin following Spalacidae and Rhizomyidae divergence (Fig. 6).

Interestingly, only those rodent families (Muridae, Cricetidae, Spalacidae, and Rhizomyidae) have $4.5S_I$ RNA, which are characterized by the presence of B2 SINEs (Serdobova and Krarnerov 1998; Gogolevskaya and Kramerov, unpublished data). Thus, $4.5S_I$ RNA genes and B2 elements arose at approximately the same period of rodent evolution (Fig. 6). Because $4.5S_I$ RNA genes and B2 elements are significantly similar in their 5′-end regions, they can be related in evolution as an ancestor and a descendant. Alternatively, both these sequences could have arisen independently at about the same period from ID SINEs or alanine tRNA[1], which are present in the genomes of all rodents and possess moderate sequence similarity to $4.5S_I$ RNA and the B2 element. It is possible that $4.5S_I$ RNA and B2, arose not directly from ID or $tRNA^{Ala}$, but from yet unidentified sequences, or 'intermediate links.' The above hypotheses are based on data of the species distribution of these DNA sequences and reconstruction of their phylogenetic trees (Serdobova and Kramerov 1998).

According to estimates of molecular taxonomists, Muridae and Cricetidae diverged ~17 million years (Myr) ago; the split between Spalacidae and Rhizomyidae occurred at about the same time (Robinson et al. 1997). Based on paleontological data, the divergence between Spalacidae/Rhizomyidae and Muridae/Cricetidae clades can be estimated at 23 Myr (Hugueney and Mein 1993). The rate of nonsynonymous and synonymous nucleotide substitutions calculated for the nuclear gene LCAT (lecithin:cholesterol acyltransferase) of these two rodent clades proved to be 3.3×10^{-3} and 2.05×10^{-2} substitutions/site/Myr, respectively (Robinson et al., 1997). The latter value is close to the rate of substitutions (2.2×10^{-2}) obtained by hybridization of the total single-copy DNA sequences of Spalacidae and Muridae (Catzeflis et al. 1989) (notice that most of such DNA sequences are non-coding regions of genome). The rate of synonymous substitutions estimated from comparison between numerous sequenced rat and hamster genes is also very

[1] Perhaps, the progenitor of B2 element was another tRNA, lysine (Sakamoto and Okada 1985) or serine (Daniels and Deninger 1985) one.



A

Fig. 5. **A.** PCR-detection of S variant of $4.5S_I$ RNA homologous sequences in rodent genomes by PCR. Each reaction mixture contained 10 ng genomic DNA and 35 amplification cycles were performed. **B.** Sequences of 93 bp cloned fragments obtained by PCR with genomic DNA of hoary bamboo rat (*Rhizomys pruinosus*) and Russian mole rat (*Spalax microphthalmus*). The corresponding consensus sequences are given below and above the sequences, respectively; matching nucleotides between the consensus sequences are denoted by asterisks. *SpaS.dir* and *SpaS.rev* primers are shown as > and <, respectively.

```
   Rh15 >....A...........A...T......G.............................<
   Rh32 >...T.....................................TG.....T.......<
   Rh37 >...T.....................................TG.....T.......<
   Rh12 >...T.....................................T......A.......<
   Rh38 >...A.....................................T...............<
   Rh11 >...A.....................................T...............<
   Rh9  >...A.....................................T...............<
   Rh34 >.....C.....A.............................................<
   Rh36 >.........T..A.............................................<
   Rh8  >..A.....T..A.............................................<
   Rh33 >....A.........T...........A.............G.............<
   Rh10 >..............T.............................G.............<
   Rh35 >..............T.............................G.............<
   Rh40 >..............T.............................G.............<
Rhizomys AGCCGTTAAAGGCTCAGGCTCACAACCAAAAAGACAAGAGTTCAGTTCCCAGCACCC
         *** ********* ****************************** *************
Spalax   AGCTGTTAAAGGCGCAGGCTCACAACCAAAAAGACAAGAGTTCGGTTCCCAGCACCC
   Sp21 >...C.....................................................<
   Sp22 >...C.....................................................<
   Sp25 >...C.....................................................<
   Sp3  >............T.....................G.....................<
   Sp6  >............T.....................G.....................<
   Sp7  >............T.....................G.....................<
   Sp30 >............T.....................G.....................<
   Sp5  >...C.........T.....................................T....<
   Sp2  >..............TT..A..........G...T..........A.....T.......<
   Sp26 >..............TT..A..........G...T..........A.....T.......<
   Sp27 >..............TT..A..........G...T..........A.....T.......<
         20         30         40         50         60         70
```

B.

similar to the above value (O'Huigin and Li 1992). Our sequencing data for the golden hamster $4.5S_I$ RNA revealed only one nucleotide substitution in comparison with rat $4.5S_I$ RNA. Using the divergence date of Muridae and Cricetida equal 17 Myr, we roughly estimated the rate of evolution of their $4.5S_I$ RNAs at only 6×10^{-4} substitutions/site/Myr. This value is much lower than that for sequences subject to no or weak selection (2×10^{-2}). Thus, it can be concluded that $4.5S_I$ RNA was under strong selection constraint during murid and cricetid evolution.

The conclusion that this RNA is under strong natural selection in murids is supported by the comparison data between the common rat and house mouse $4.5S_I$ RNAs. Although the complete nucleotide sequence of mouse $4.5S_I$ RNA is unknown, it can be deduced from the following data: first, the 3′ end part (from position 50) of this RNA has been sequenced (Kalinina et al. 1987); second, in Genbank one can find at least 10 mouse retropseudogenes corresponding to the 5′ end half of $4.5S_I$ RNA (up to position 57). The sequence obtained by joining these two halves is completely identical to rat $4.5S_I$

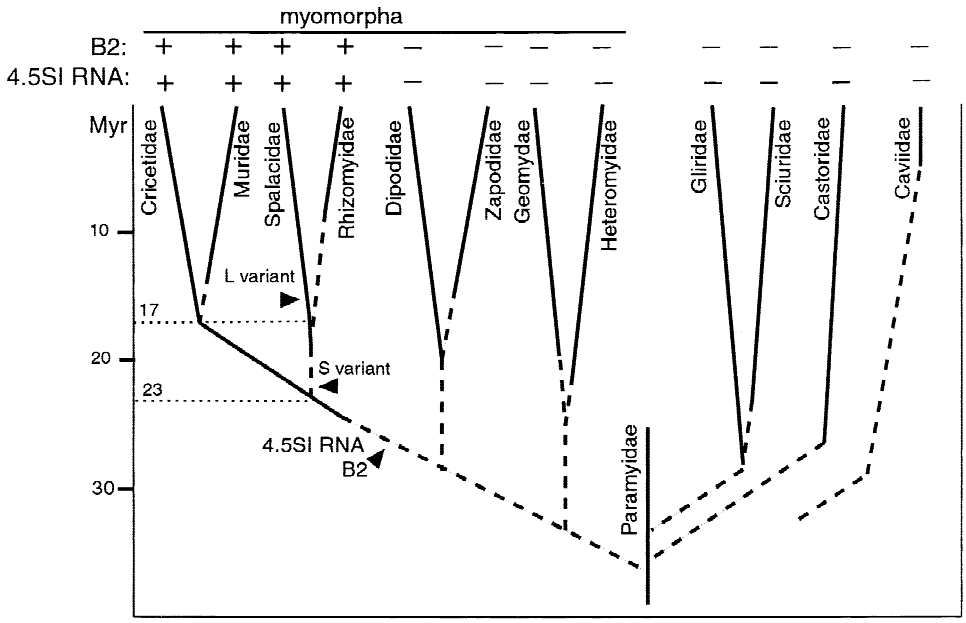


Fig. 6. The rodent phylogenetic tree and the distribution of $4.5S_I$ RNA genes and B2 element. The tree is based on the phylogeny proposed by Romer (1966) and modified according to (Carrol 1988; Robinson et al. 1997; Kramerov et al. 1999). The left-hand cluster of clades (myomorpha) is shown completely, whereas the right-hand one is significantly reduced by comparison with original: only four of 22 families are shown. Arrowheads indicate likely moments of $4.5S_I$ RNA genes and B2 element emergence. See the text for other explanations.

RNA (variant A). Thus, no nucleotide changes were fixed in $4.5S_I$ RNA genes since the divergence between *Mus* and *Rattus* that probably occurred about 12 Myr ago (Robinson et al. 1997). For comparison, mouse and rat protein-coding genes differ by 18% synonymous substitutions (O'Huigin and Li 1992).

The analysis of the S variant of mole rat $4.5S_I$ RNA and its homologs from the bamboo rat genome revealed a very high level of similarity. This suggests that the spalacid and rhizomyd $4.5S_I$ RNAs were also under strong selection constraint.

It is likely that all 10 nucleotide changes, by which the spalacid/rhizomyid S variant are distinguished from murid/cricetid $4.5S_I$ RNA (Fig. 4), appeared in the 6 Myr period which began with the branching of spalicid/rhizomyd clade (23 Myr ago) and ended with dichotomies murids/cricetids and spalacids/rhizomyds (17 Myr ago; Fig. 6). So, for this period of time, the rate of change of $4.5S_I$ RNA may be estimated as 1.7×10^{-2} changes/site/Myr (10 changes/100 nt × 6 Myr), very close to the change rate of sequences subject to no or weak selection. It is possible that $4.5S_I$ RNA had not been recruited (excepted) into a function yet, and that emergence of the function resulted in a much higher conservation of $4.5S_I$ RNA in the period following the divergence between murids and cricetids, as well as between spalacids and rhizomyds.

Evolution of $4.5S_I$ RNA genes seems to be a rather complex process. It may include duplication or amplification of the genes, the alteration of their transcriptional activity, or the conversion of the genes into pseudogenes. Perhaps the first $4.5S_I$ RNA-L gene arose in an early spalacid due to insertion of 7-nt sequence into one of $4.5S_I$ gene copies, that probably had already contained several nucleotides differing from those in the murid/cricetid variant of the gene. Then the L-variant seems to have been subjected to amplification and subsequent divergence of some copies. Thereby L1, L2, L3, and other possible existing subvariants of this gene could have emerged. Although the major subvariant L1 diverged from the ancestor gene with a significant rate (0.8×10^{-2} substitutions/site/Myr) it might have kept its function. This is less likely in regard to L2, L3, and other minor subvariants—they probably are transcribed pseudogenes.

The above considered data on the conservation of $4.5S_I$ RNA nucleotide sequences (at least, in Muridae/Cricetidae evolution) allowed us to conclude that this RNA may have a function. This is interesting considering the fact of the recent (25–30 Myr) origin and rather narrow species distribution range of $4.5S_I$ RNA genes. This RNA might possess some unusual, highly specific function. Because the species distributions of $4.5S_I$ RNA and B2 SINE coincide and their sequences are partly similar, one can suggest that the requirement for $4.5S_I$ RNA is somehow connected with the presence of the B2 elements in the genome. Following the hypothesis regarding the "selfish" and parasitic nature of SINEs (Orgel and Crick 1980; Doolittle and Sapienza 1980), the possible role of $4.5S_I$ RNA could be in suppressing of the presumable deleterious effects of B2 SINE transcripts. According to the other concepts, SINEs or their transcripts have a certain role in cell vital activity (Vidal et al. 1993; Chu et al. 1998). Therefore, $4.5S_I$ RNA might cooperate with the B2 element during their functioning,

for example, in putative regulation of expression of the B2-containing genes. However, it is quite possible that the role played by $4.5S_I$ RNA is in no way connected with B2 or some other SINEs. It was reported that the mouse cells infected with encephalomyocarditis virus are characterized by considerably increased content of $4.5S_I$ RNA in polyribosomes (Kalinina et al. 1987). It is not improbable that $4.5S_I$ RNA can be involved in the regulation of protein synthesis.

$4.5S_I$ RNA genes are not the only example of genes characterized by a recent origin and hence a narrow species distribution range. We propose to use a term *stenogenes* for such genes and *steno*RNA for their transcripts (stenos-narrow, Greek)[2]. Besides $4.5S_I$ RNA, small nuclear RNAs ($4.5S_I$ RNA, BC1, and BC200) that were considered in 'Introduction' should be referred to as stenoRNA. A small nucleolar RNA RBII-36, which is exclusively detected in the *Rattus* genus of rodents (Cavaille et al. 2001), also can be assigned to steno RNAs. Gadd7 gene is an example of a stenogene that is actively transcribed following DNA damage (Hollander et al. 1996). This gene has been detected in the Chinese hamster, but not in the mouse genome. *Drosophila melanogaster* has a heat-shock locus hsr-omega that demonstrates virtually no sequence similarity to the corresponding locus of *D. hydei* (Garbe et al. 1989) and can be ranked as a stenogene as well. Notice that both gadd7 and hsr-omega genes contain no long open reading frames, encode no polypeptides, and express only RNA products. An example of a protein-encoding stenogene is Stellate gene on *D. melanogaster* X-chromosome that is absent in the closely related species *D. simulans* (Kogan et al. 2000). Chimeric genes *jingwei* from *D. teissieri* and *D. yakuba* (Long et al. 1999) as well as *PMCHL* in hominids (Courseaux and Nahon, 2001) can also be protein-encoding stenogenes.

Although mobile genetic elements, in particular retroposons and retrotransposons, are usually characterized by narrow species distribution, they should not be assigned to stenogenes. Therefore among the features of stenogenes are a lack of: (i) short flanking repeats, (ii) A-rich sequences on the 3′-end, (iii) high heterogeneity of copies. As a rule, retroposon transcription is repressed in adult tissues (except germ line cells) (Grigoryan et al. 1985; Paulson and Schmid 1986; Bachvarova 1988), whereas stenogenes are usually actively transcribed in all or many tissues of adults. The study of stenogenes may be useful for understanding the mechanisms of new gene emergence and development of their functions.

Acknowledgments. We are grateful to Drs. E. Lyapunova, G. Shenbrot, E. Ivanitskaya, A. Puzachenko, E. Panova, N. Bulatova, O. Likhnova, T. Golovkina, F. Catzeflis, and R. DeBry for providing animals, tissues or DNA, Dr. R. Reddy for the cloned rat $4.5S_I$ RNA gene, and Jennifer Sporty for reading the manuscript. This study was supported by International Science Foundation (grant N4E000) and Russian Foundation for Basic Research (grant 99-04-49177).

[2] The term *stenonuon* (derived from *nuon* originally proposed by Brosius and Gould (1992) for any definable sequence of DNA or RNA) can also be useful. Both stenogenes and SINEs can be ranged in stenonuons.

References

Bachvarova R (1988) Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. Dev Biol 130:513–523

Brosius J, Gould SJ (1992) On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA" Proc Natl Acad Sci USA 89:10706–10710

Brosius J (1999) RNAs form all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238: 115–134

Carrol RL (1988) Vertebrate Paleontology and Evolution. Freeman and Company, New York

Catzeflis FM, Nevo E, Ahlquist JE, Sibley CG (1989) Relationships of the chromosomal species in the Eurasian mole rats of the Spalax ehrenbergi group as determined by DNA-DNA hybridization, and an estimate of the spalacid-murid divergence time. J Mol Evol 29:223–232

Cavaille J, Vitali P, Basyuk E, Huttenhofer A, Bachellerie JP (2001) A novel brain-specific box C/D small nuclear RNA processed from tandemly-repeated introns of a non-coding RNA gene in rat. J Biol Chem 276:26374–26383

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate- phenol-chloroform extraction. Anal Biochem 162:156–159

Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW (1998) Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. Mol Cell Biol 18:58–68

Courseaux A, Nahon JL (2001) Birth of two chimeric genes in the *Hominidae* lineage. Science 291:1293–1297

Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. Nature 317:819–822

DeChiara TM, Brosius J (1987) Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. Proc Natl Acad Sci USA 84:2624–2628

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Eddy SR (1999) Noncoding RNA genes. Curr Opin Genet Dev 9:695–699

Garbe JC, Bendena WG, Pardue ML (1989) Sequence evolution of the Drosophila heat shock locus hsr omega. I. The nonrepeated portion of the gene. Genetics 122:403–415

Grigoryan MS, Kramerov DA, Tulchinsky EM, Revasova ES, Lukanidin EM (1985) Activation of putative transposition intermediate formation in tumor cells. Embo J 4:2209–2215

Harada F, Kato N (1980) Nucleotide sequences of 4.5S RNAs associated with poly(A)-containing RNAs of mouse and hamster cells. Nucleic Acides Res 8:1273–1285

Harada F, Kato N, Hoshino H. (1979) Series of 4.5S RNA associated with poly(A)-containing RNAs of rodent cells. Nucleic Acids Res 7:909–917

Hollander MC, Alamo I, Fornace AJ, Jr (1996) A novel DNA damage-inducible transcript, gadd7, inhibits cell growth, but lacks a protein product. Nucleic Acids Res 24:1589–1593

Hugueney M, Mein P (1993) A comment on the earliest Spalacinae (Rodentia, Muroidae). J Mammal Evol 1:215–223

Kalinina NO, Skarlat IV, Chumakov KM (1987) 4.5S RNAI in virus-specific polyribosomes from ascitic Krebs 2 carcinoma cells infected with encephalomyocarditis virus. Biokhimiia 52:2045–2050

Kogan GL, Epstein VN, Aravin AA, Gvozdev VA (2000) Molecular

evolution of two paralogous tandemly repeated heterochromatic gene clusters linked to the X and Y chromosomes of Drosophila melanogaster. Mol Biol Evol 17:697–702

Kramerov D, Vassetzky N, Serdobova I (1999) The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. Mol Biol Evol 16:715–717

Kramerov DA, Lekakh IV, Samarina OP, Ryskov AP (1982) The sequences homologous to major interspersed repeats B1 and B2 of mouse genome are present in mRNA and small cytoplasmic poly(A) + RNA. Nucleic Acids Res 10:7477–7491

Krayev AS, Markusheva TV, Kramerov DA, Ryskov AP, Skryabin KG, Bayev AA, Georgiev, G. (1982) Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res 10:7461–7475

Long M, Wang W, Zhang J (1999) Origin of new genes and source for N-terminal domain of the chimerical gene, *jingwei*, in Drosophila. Gene 238:135–141

Mrtignetti JA, Brosius J (1993) BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. Proc Natl Acad Sci USA 90:11563–11567

Martignetti JA, Brosius J (1993) Neural BC1 RNA as an evolutionary marker: guinea pig remains a rodent. Proc Natl Acad Sci USA 90:9698–9702

Maxwell ES, Fournier MJ (1995) The small nucleolar RNAs. Annu Rev Biochem 64:897–934

McKeown M (1993) The role of small nuclear RNAs in RNA splicing. Curr Opin Cell Biol 5:448–454

Miller KS, Zbrzezna V, Pogo AO (1984) Interstrand duplexes in Friend erythroleukemia nuclear RNA. The interaction of non-polyadenylated nuclear RNA with polyadenylated nuclear RNA and with small nuclear RNAs. 177:343–368

O'HUigin C, Li WH (1992) The molecular clock ticks regularly in muroid rodents and hamsters. J Mol Evol 35:377–384

Okada N (1991) SINEs. Curr Opin Genet Dev 1, 498–504

Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607

Paulson KE, Schmid CW (1986) Transcriptional inactivity of Alu repeats in HeLa cells. Nucleic Acids Res 14:6145–6158

Reddy R, Henning D, Tan E, Busch H (1983) Identification of a La protein binding site in a RNA polymerase III transcript (4.5SI RNA). J Biol Chem 258:8352–8356

Robinson M, Catzeflis F, Briolay J, Mouchiroud D (1997) Molecular phylogeny of rodents, with special emphasis on murids: evidence from nuclear gene LCAT. Mol Phylogenet Evol 8:423–434

Ro-Choi TS, Reddy R, Henning D, Takano T, Taylor C, Busch H (1972) Nucleotide sequence of 4.5S RNA I of Novikoff hepatoma cell nuclei. J Biol Chem 247:3205–3222

Romer AS (1966) Vertebrate paleontology. 3rd ed. University of Chicago Press, Chicago

Saba JA, Busch H, Reddy R (1985) A new moderately repetitive rat DNA sequence detected by a cloned 4,5SI DNA. J Biol Chem 260:1354–1357

Sakamoto K, Okada N (1985) Rodent type 2 Alu, rat identifire sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J Mol Evol 22:134–140

Schoeniger LO, Jelinek WR (1986) 4.5S RNA is encoded by hundreds of tandemly linked genes, has a short half-life, and is hydrogen bonded in vivo to poly(A)-terminated RNAs in the cytoplasm of cultured mouse cells. Mol Cell Biol 6:1508–1519

Serdobova IM, Kramerov DA (1998) Short retroposons of the B2 superfamily: Evolution and application for the study of rodent phylogeny. J Mol Evol 46:202–214

Takeuchi Y, Harada F (1986) Cloning and characterization of rat 4.5S RNAI genes. Nucleic Acids Res 14:1643–1656

Tiedge H, Chen W, Brosius J (1993) Primary structure, neural-specific expression, and dendritic location of human BC200 RNA. J Neurosci 13:2382–2390

Vidal F, Mougneau E, Glaichenhaus N, Vaigot P, Darmon M, Cuzin F (1993) Coordinated posttranscriptional control of gene expression by modular elements including Alu-like repetitive sequences. Proc Natl Acad Sci USA 90:208–212