

## RESEARCH ARTICLES

### Evolutionary History of 4.5SH RNA

Irina K. Gogolevskaya, Anastasia P. Koval, and Dmitri A. Kramerov

Laboratory of Eukaryotic Genome Evolution, Engelhardt Institute of Molecular Biology,  
32 Vavilov Street, Moscow 19991, Russia

4.5SH RNA is a 94-nt small RNA with unknown function. This RNA is known to be present in the mouse, rat, and hamster cells; however, it is not found in human, rabbit, and chicken. In the mouse genome, the 4.5SH RNA gene is a part of a long (4.2 kb) tandem repeat (~800 copies) unit. Here, we found that 4.5SH RNA genes are present only in rodents of six families that comprise the Myodonta clade: Muridae, Cricetidae, Spalacidae, Rhizomyidae, Zapodidae, and Dipodidae. The analysis of complementary DNA derived from the rodents of these families showed general evolutionary conservation of 4.5SH RNA and some intraspecific heterogeneity of these RNA molecules. 4.5SH RNA genes in the Norway rat, mole rat, hamster and jerboa genomes are included in the repeated sequences. In the jerboa genome these repeats are 4.0-kb long and arranged tandemly, similar to the corresponding arrangements in the mouse and rat genomic DNA. Sequencing of the rat and jerboa DNA repeats containing 4.5SH RNA genes showed fast evolution of the gene-flanking sequences. The repeat sequences of the distantly related rodents (mouse and rat vs. jerboa) have no apparent similarity except for the 4.5SH RNA gene itself. Conservation of the 4.5SH RNA gene nucleotide sequence indicates that this RNA is likely to be under selection pressure and, thus, may have a function. The repeats from the different rodents have similar lengths and contain many simple short repeats. The data obtained suggest that long insertions, deletions, and simple sequence amplifications significantly contribute in the evolution of the repeats containing 4.5SH RNA genes. The 4.5SH RNA gene seems to have originated 50–85 MYA in a Myodonta ancestor from a copy of the B1 short interspersed element. The amplification of the gene with the flanking sequences could result from the supposed cellular requirement of the intensive synthesis of 4.5SH RNA. Further Myodonta evolution led to dramatic changes of the repeat sequences in every lineage with the conservation of the 4.5SH RNA genes only. This gene, like some other relatively recently originated genes, could be a useful model for studying generation and evolution of non-protein-coding genes.

#### Introduction

During the last two decades, more than a hundred small RNAs (except transfer RNAs) have been found in mammalian cells (McKeown 1993; Maxwell and Fournier 1995; Eddy 1999). Many of them play an important role in processes such as ribosomal RNA (rRNA) processing and modification (U3 RNA, C/D box, and H/ACA box RNAs), gene transcription (7SK RNA), pre-mRNA splicing (U1, U2, U4, U5, U6 RNAs), mRNA decay (small interference RNAs), and protein secretion (7SL RNA).

Most small RNAs are ubiquitous and have conserved nucleotide sequences, at least, among mammals. However, several species of small RNAs isolated from rodent cells were not found in other analyzed mammalian orders. The best studied among those RNAs is BC1 RNA that was isolated from several different rodents including rat, mouse, and distantly related guinea pig (DeChiara and Brosius 1987; Martignetti and Brosius 1993a). However, BC1 RNA has not been detected in rabbit, cattle, and human. This RNA is exclusively characteristic of nervous tissue as well as its analog BC200 RNA found in human and other primates. BC1 and BC200 RNA demonstrate a partial sequence similarity in the 3'-end parts (Martignetti and Brosius 1993b). BC1 RNA seems to contribute to the modulation of rodent behavior (Lewejohann et al. 2004).

The other two small RNA species found in mouse, rat, and hamster cells, but not in human, rabbit, and dog cells,

were described in the 1970s and named 4.5S<sub>I</sub> RNA (Ro-Choi et al. 1972; Reddy et al. 1983) and 4.5S RNA (Harada, Kato, and Hoshino 1979; Harada and Kato 1980). Harada et al. (1986) renamed the later 4.5S RNA<sub>H</sub> (we will use the term 4.5SH RNA). 4.5S<sub>I</sub> RNA is present in rodents of only four families (Muridae, Cricetidae, Spalacidae, and Rhizomyidae), and high conservation of its nucleotide sequence indicates that this RNA is functional (Gogolevskaya and Kramerov 2002). 4.5S<sub>I</sub> RNA and 4.5SH RNA are present in most, if not all rat tissues. Some of the 4.5SH RNA and 4.5S<sub>I</sub> RNA molecules were detected in association with polyadenylated and nonpolyadenylated nuclear RNAs, respectively (Harada, Kato, and Hoshino 1979; Miller, Zbrzezna, and Pogo 1984; Schoeniger and Jelinek 1986). The function of both these small RNAs remains unknown.

The small RNAs described above share several important features: (1) they are synthesized by RNA polymerase III, (2) their taxonomic distribution is relatively narrow, and (3) they demonstrate partial sequence similarity to either rodent or primate short interspersed elements (SINEs). The 5'-end parts of BC1 RNA and BC200 RNA are virtually identical to SINEs ID (DeChiara and Brosius 1987) and Alu (Martignetti and Brosius 1993b), respectively. The 5' end of the 4.5S<sub>I</sub> 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, Busch, and Reddy 1985; Serdobova and Kramerov 1998). The nucleotide sequence of 4.5SH RNA is homologous to the rodent B1 element, although the RNA is shorter and contains a specific 20-nt region (Kramerov et al. 1982; Krayev et al. 1982; Labuda and Zietkiewicz 1994).

We proposed a term “stenoRNA” (steno: narrow, Greek) for RNAs characterized by narrow species distribution

Key words: rodentia, small RNA, 4.5S RNA, 7SL RNA-derived SINEs, origin of genes, tandem repeats.

E-mail: kramerov@eimb.ru.

*Mol. Biol. Evol.* 22(7):1546–1554. 2005

doi:10.1093/molbev/msi140

Advance Access publication April 6, 2005

(Gogolevskaya and Kramerov 2002). All RNAs described above and some others are included in this RNA group. Narrow taxonomic distribution is the evidence of the recent origin of these RNAs in evolution. Therefore, stenoRNAs can serve as very useful and valuable models for studying the process of emergence of new functional RNAs and their genes.

The 4.5S<sub>I</sub> RNA is encoded in several dozen genes dispersed throughout the genome (Saba, Busch, and Reddy 1985; Takeuchi and Harada 1986). On the other hand, the mouse and rat 4.5SH RNA genes are a part of long (4.2 and 5.3 kb, respectively) units, which are arranged as tandem repeats (700–800 copies) (Schoeniger and Jelinek 1986).

In the current study, the taxonomic range of the 4.5SH RNA was investigated in detail. The data obtained allowed us to estimate the time of appearance of this RNA in evolution. In addition, it became clear that the tandem organization of the repeats containing 4.5SH RNA genes exhibits evolutionary conservatism, whereas the nucleotide sequence of the repeat itself does not do so.

## Materials and Methods

### DNA and RNA Electrophoresis and Hybridization

Sources of DNA and RNA were described previously (Gogolevskaya and Kramerov 2002). Total liver RNA was fractionated by electrophoresis in 6% polyacrylamide gel PAAG containing 7M urea and transferred onto a Hybond-N membrane using semidry electroblotting. A 72-bp polymerase chain reaction (PCR) product complementary to the 4.5SH RNA gene was used as a probe in Northern hybridization. The probe was obtained by PCR of mouse genomic DNA with primers 5'-CCGGTAGGATTTGCTGAA-3' and 5'-AAAATGTGAGCCCAGGC-3'. The aliquot (1%) of the isolated fragment was labeled by 20 cycles of PCR with  $\alpha$ [<sup>32</sup>P]deoxyadenosine triphosphate (25  $\mu$ Ci) and the same pair of primers. The hybridization was performed in 5 $\times$  standard saline citrate (SSC), 0.5% sodium dodecyl sulfate (SDS), 0.1% polyvinylpyrrolidone, 0.1% Ficoll, and 0.1 mg/ml salmon DNA at 60°C. The filter washed with 0.1 $\times$  SSC, and 0.1% SDS at 42°C (nonstringent conditions).

Genomic DNA (10  $\mu$ g) digested with restriction enzymes was fractionated by electrophoresis in 0.8% agarose gels and transferred onto a Hybond-N membrane using capillary blotting. The oligonucleotide, 5'-GCAC-(G/A)CCGGTAGGAT(T/A)TGCTGAAG-3', labeled by  $\gamma$ [<sup>32</sup>P]adenosine triphosphate with polynucleotidekinase was used as a 4.5SH RNA-specific probe in Southern hybridization. The hybridization conditions were the same as described above, except for the hybridization (42°C) and washing (37°C) temperature.

### Complementary DNA Synthesis and Cloning

Complementary DNA (cDNA) was synthesized and cloned as described previously (Gogolevskaya and Kramerov 2002). Briefly, the RNA fraction enriched with 4.5SH RNA was isolated by electrophoresis of total liver RNA in 6% PAAG. The RNA was polyadenylated using *Escherichia coli* poly(A) polymerase. cDNA was synthesized with M-MLV reverse transcriptase using the

primer *Xba*I(T)<sub>15</sub>: 5'-GTCGACTCTAGA(T)<sub>15</sub>-3'. The 3' end of cDNA was tailed with oligo(dG) using terminal deoxynucleotidyl transferase, and the reaction product was amplified by 30 cycles of PCR with primers *Xba*I(T)<sub>15</sub> and *Eco*RI(C)<sub>10</sub>—5'-CGGAATTCGT(C)<sub>10</sub>-3'. The double-stranded cDNA obtained was digested with *Xba*I and *Eco*RI and cloned in a plasmid pGEM 7Zf<sup>+</sup> that had been cut with the same enzymes. The library was screened by hybridization with the 72-bp 4.5SH RNA gene probe labeled by PCR.

### Cloning of Genomic Fragments Containing 4.5SH RNA Genes

Genomic DNA (100  $\mu$ g) was digested with *Bam*HI (Norway rat) or *Eco*RI (great jerboa and mole rats) endonucleases and separated by electrophoresis in 0.8% agarose gel. DNA from zones of the gel, which are enriched in 4.5SH RNA genes (see Results), were isolated by electroelution on a diethylaminoethyl membrane and cloned in plasmid pBS(SK+). The library was screened by hybridization with a 4.5SH RNA-specific oligonucleotide probe.

### PCR Analysis

PCR was performed to detect 4.5SH RNA genes and to investigate the arrangement of the 4.5S<sub>H</sub>-repeats. For detection of the 4.5SH RNA genes, primers 5'-CCGGTAGA(T/A)TTTGCTGAAG-3' and 5'-GTGAAAAAATGTGAG-CCCA-3' and 5 ng of genomic DNA were used. Twenty-five PCR cycles of 95°C for 1 min, 52°C for 1 min, and 72°C for 1 min were carried out. PCR products were analyzed by electrophoresis in the gel containing 3% NuSeave agarose and 1% regular agarose. To study the arrangement of the jerboa 4.5SH-repeats, primers 5'-TAGGAGGCCACCCTGAA-TTC-3' and 5'-TCTACCTGCCTCATACAA-3' were used. Twenty-five cycles of 95°C for 1 min, 54°C for 1 min, and 72°C for 2 min were performed. PCR products were analyzed by electrophoresis in 1% agarose gel.

### Sequencing

cDNA clones were sequenced with standard M13 primers. 4.5S<sub>H</sub>-repeat clones were subcloned and sequenced using the standard and specific (internal) primers. Double-stranded plasmid templates were sequenced by the dideoxynucleotide method with Sequenase 2.0 according to the manufacturer's instructions, with modifications (Redston and Kern 1994). Additionally, some of the clones were sequenced using the *Taq* DNA polymerase method (Slatko, Albright, and Tabor 1992).

## Results

### Investigating Taxonomic Distribution of 4.5SH RNA

Total liver RNA from representatives of the several rodent families was analyzed by Northern hybridization with the 72-bp probe complementary to the mouse 4.5SH RNA gene. Hybridization signals were observed in the RNA of house mouse (Muridae), golden hamster (Cricetidae), mole rat (Spalacidae), and great jerboa (Dipodidae) but not in the RNA of long-tailed marmot, palm squirrel, large-toothed suslik (Sciuridae), and guinea pig (Caviidae) (fig. 1).

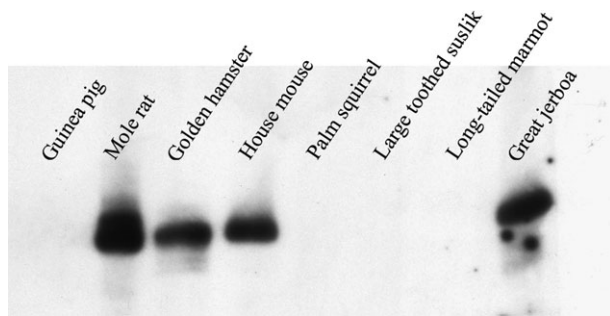


FIG. 1.—Northern blot hybridization of total liver RNA from different rodents with 72-bp probe complementary to mouse 4.5SH RNA gene. RNA was separated by electrophoresis in 6% PAAG. Hybridization and washing of the filter were carried out under nonstringent conditions.

cDNA derived from 4.5SH RNA of the Norway rat (*Rattus norvegicus*), golden hamster (*Mesocricetus auratus*), mole rat (*Spalax microphthalmus*), and great jerboa (*Allactaga major*) were cloned and six, three, five, and six clones, respectively, were sequenced (fig. 2). The nucleotide sequence analysis of the cloned cDNA revealed the following: (1) all cDNA clones represent the same small RNA type; (2) some cDNA molecules derived from the same species (in our case, single organism) differ by a limited number of substitutions, and there is a 4-nt insertion within the 3'-end region of two mole rat clones; (3) a few species-specific substitutions are observed (G is found in the position 40 in rat and hamster, whereas it is absent in most mole rat and jerboa clones; AC and TG are seen in the 61–62 positions in Norway rat and in mole rat, respectively). The rate of interspecific nucleotide substitution (2%–16%) is only slightly higher than intraspecific differences (0%–14%) between 4.5SH RNA molecules. This intraspecific heterogeneity was first discovered in the current study, because previously the nucleotide sequence of this RNA was studied by footprinting only, without sequencing cDNA clones. Most likely such heterogeneity is due to the variability of the multiple 4.5SH RNA genes. Contribution of the PCR-derived mutations in the observed 4.5SH RNA heterogeneity cannot be ruled out; however, it is not likely to be significant, as virtually no heterogeneity of the 4.5S<sub>I</sub> RNA was observed in a similar experiment (Gogolevskaya and Kramerov 2002).

A primer pair specific to all sequenced 4.5SH RNAs was designed. These primers were used to detect 4.5SH RNA genes in the genomes of the rodents of 14 families. PCR products of expected size were observed in the representatives of only six related rodent families: Muridae, Cricetidae, Spalacidae, Rhizomyidae, Zapodidae, and Dipodidae (fig. 3). The absence of the PCR product due to primer mismatch in a critical position or low copy number of the gene cannot be ruled out in some cases. However, it is unlikely because 4.5SH RNA shows obvious conservation, and the PCR conditions used were similar to those which allow the detection of a single copy of short nucleotide sequences in mammalian genomic DNA (Serdobova and Kramerov 1998). Thus, 4.5SH RNA has narrow taxonomic distribution.

#### Tandem Organization of 4.5SH RNA Genes Is Conserved in Evolution

It has been demonstrated previously that in mouse and rat 4.5SH RNA genes are a part of the long (4.2 and 5.3 kb, respectively) tandem repeats (Schoeniger and Jelinek 1986). To analyze whether such an organization is conserved among other rodents possessing 4.5SH RNA, Southern blot hybridization of the 4.5SH RNA-specific oligonucleotide probe to genomic DNA from mole rat, great jerboa, golden hamster, Norway rat, and house mouse (as a control) was performed. At least one hybridization band as strong as in the case of mouse was observed for all the species tested (fig. 4). This result suggests that the 4.5SH RNA gene is a part of repeated sequences in the mole rat, great jerboa, and golden hamster as well as in Norway rat and house mouse. We called these sequences 4.5SH-repeats.

To find out if the tandem arrangement of the 4.5SH-repeats is conserved in evolution, we studied this sequence in great jerboa, a representative of Dipodidae, a rodent

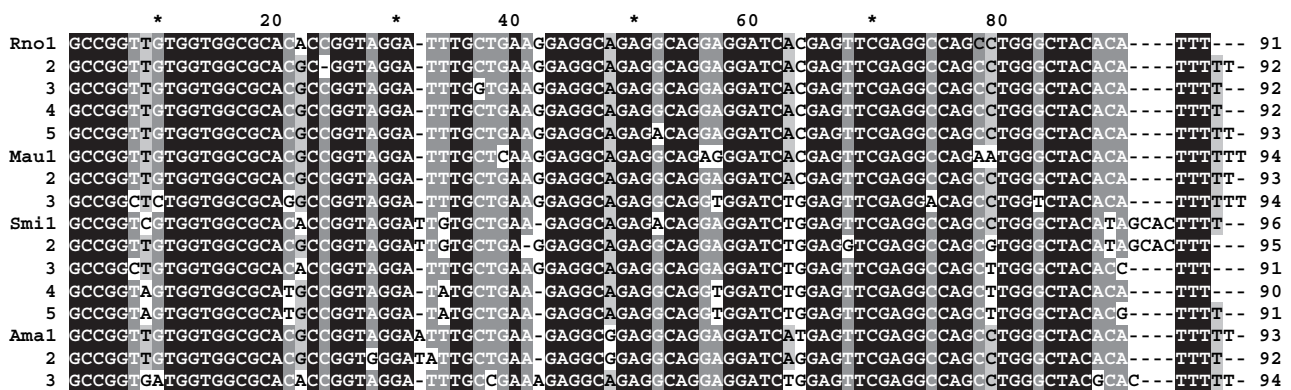


FIG. 2.—Nucleotide sequences of the different cDNA variants derived from the 4.5SH RNA of Norway rat (*Rattus norvegicus*—Rno), golden hamster (*Mesocricetus auratus*—Mau), mole rat (*Spalax microphthalmus*—Smi), and great jerboa (*Allactaga major*—Ama). Rno cDNA variant 1 is represented by two clones; Ama cDNA variant 1 and variant 2 are represented by three and two clones, respectively. All remaining cDNA variants are represented by single clones. Differences in T residue numbers at the 3' ends of the sequences could be explained with stochastic character of RNA polymerase III termination. GenBank accession numbers: Rno, AY228147–AY228151; Mau, AY228157–AY228159; Smi, AY228152–AY228156; Ama, AY228144–AY228146.

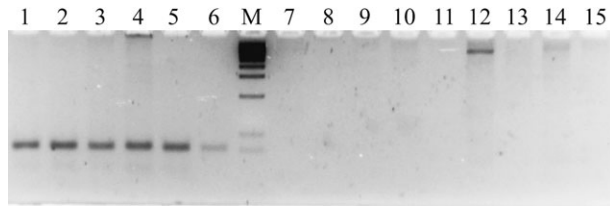


FIG. 3.—PCR detection of 4.5SH RNA genes in rodent genomes. 1: *Mus musculus*, house mouse (Muridae); 2: *Pitomis daghestanicus*, daghestan vole (Cricetidae); 3: *Spalax microphthalmus*, Russian mole rat (Spalacidae); 4: *Rhizomys pruinosus*, hoary bamboo rat (Rhizomyidae); 5: *Allactaga major*, great jerboa (Dipodidae); 6: *Sicista tianschanica*, birch mouse (Zapodidae); 7: *Thomomys bottae*, Botta's pocket gopher (Geomyidea); 8: *Dipodimus deserti*, desert kangaroo rat (Heteromyidea); 9: *Perognathus parvus*, grate basin pocket mouse (Heteromyidea); 10: *Marmota caudate*, long-tailed marmot (Sciuridae); 11: *Hystrix leucura*, Indian porcupine (Hystricidae); 12: *Castor fiber*, beaver (Castoridae); 13: *Anomalurus species*, scaly-tailed flying squirrel (Anomaluridae); 14: *Myoprocta acoushy*, green acoushy (Dasiproctidae); 15: *Cavia porcellus*, guinea pig (Caviidae). M: 100-bp DNA ladder; the bottom band is the 80-bp fragment.

family that is the most distant from Muridae among those possessing 4.5SH RNA. Jerboa genomic DNA was subjected to limited or complete *Bgl*III digestion, separated by electrophoresis, and hybridized with the 4.5SH RNA-specific probe (fig. 5A). After complete digestion the 4.0-kb band was observed, whereas after limited digestion the second 8.0-kb band appeared. Those bands were

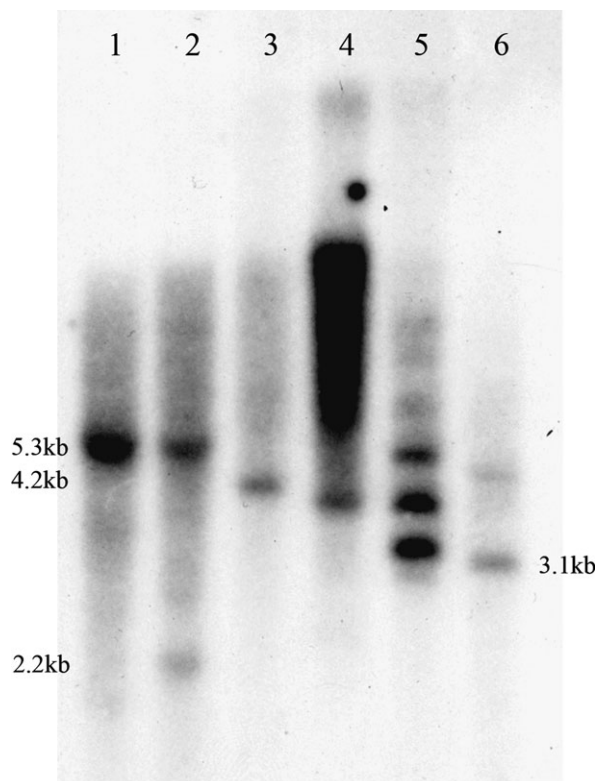


FIG. 4.—Southern blot hybridization of the 4.5SH RNA-specific probe to genomic DNA of different rodents. Genomic DNA of Norway rat (1, 2), house mouse (3), golden hamster (4), mole rat (5), and great jerboa (6) was digested with *Hind*III (1), *Bam*HI (2, 5), and *Eco*RI (3, 4, 6).

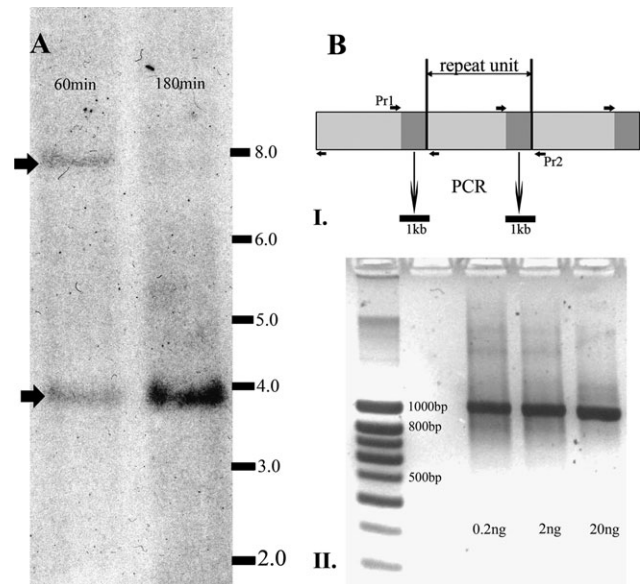


FIG. 5.—Tandem organization of the 4.5SH-repeat in the great jerboa genome. (A) Southern blot hybridization of the 4.5SH RNA-specific probe to the partially *Bgl*III-digested great jerboa genomic DNA. The time of reaction was 60 and 180 min for partial and complete digestion, respectively. The 4.0- and 8.0-kb bands correspond to the monomer and the dimer of the 4.5SH-repeat, respectively. (B) PCR analysis of great jerboa 4.5SH-repeat organization. (I) The diagram of the PCR. Primers (arrowheads) are directed toward the ends of the 3.1-kb *Eco*RI fragment (light gray-shaded boxes). The 1-kb *Eco*RI fragment is dark gray shaded. Three tandem repeat units are shown. (II) The electrophoresis of the PCR products in agarose gel. The amount of great jerboa genomic DNA used in the reaction is shown at the bottom.

interpreted as the monomer and the dimer of the tandemly arranged repeats. Thus, these data are in accord with the hypothesis postulating tandem arrangement of the 4.5SH-repeats in the jerboa genome. However, due to the large size of the repeat, it is difficult to observe a ladder with a larger number of bands. An alternative approach was used to further support this hypothesis. The 3.1-kb *Eco*RI fragment of the jerboa 4.5SH-repeat was cloned and sequenced. Then, primers complementary to the end regions and directed toward the ends of the fragment (fig. 5B) were designed. In such an experiment, PCR products can be produced only if repeat units are arranged tandemly. As expected, the 1-kb-long PCR product was observed after PCR amplification using different amounts of jerboa genomic DNA template, confirming the tandem organization of the 4.5SH-repeats in the jerboa genome.

Thus, one may strongly suggest, that tandem arrangement of the repeated sequence containing the 4.5SH RNA gene is characteristic not only of mouse and rat but of the phylogenetically distant species great jerboa as well. Therefore, the most likely scenario is that 4.5SH-repeats are also tandemly arranged in the genomes of the rodents from the families, which are closer to Muridae than Dipodidae.

#### Analysis of 4.5SH-Repeat Nucleotide Sequences

The complete nucleotide sequence of the 4.5SH-repeat was determined for great jerboa and partial sequences for mole rat and Norway rat. To isolate jerboa 4.5SH-repeats,

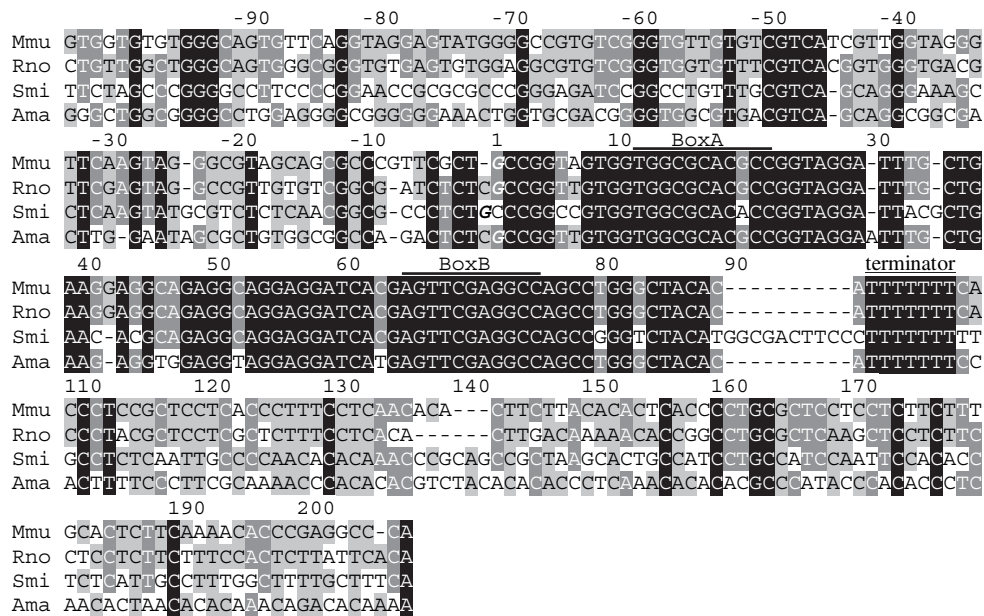


FIG. 6.—The alignment of the mouse (Mmu), rat (Rno), mole rat (Smi), and great jerboa (Ama) cloned 4.5SH RNA genes with 100-bp flanking sequences. The transcription start site is marked “1.” Box A, box B, and RNA polymerase III terminator ( $T_7$ ) are shown. The accession numbers are X60026, AY228160, AY828231, and AY828230 for Mmu, Rno, Smi, and Ama, respectively.

a library was prepared by cloning of the 3.0- to 3.5-kb fragment of the *EcoRI*-digested genomic DNA. Using colony hybridization, two clones containing the 4.5SH RNA gene were selected. The complete nucleotide sequence of one of the clones (3,073-bp fragment) was determined. The second clone that was partially sequenced differed from the first one approximately by 1% of nucleotides. Judging from Southern blot hybridization experiments (see fig. 5A), the jerboa 4.5SH-repeat seems to be 4.0 kb in length and seems to have two *EcoRI* sites. The remaining not sequenced region was PCR amplified with primers complementary to the ends of the 3.1-kb *EcoRI* fragment mentioned above. The 1-kb PCR product (see fig. 5B) was cloned and sequenced, thus, completing the nucleotide sequence of the jerboa 4.0-kb 4.5SH-repeat (accession number AY828230).

The nucleotide sequences of the 4.5SH RNA gene are quite similar in jerboa and mouse; however, other sequences of the jerboa 4.5SH-repeat demonstrated virtually no similarity to the mouse sequences. Nevertheless, very short regions of similarity to the mouse sequence can be observed upstream of the jerboa 4.5SH RNA gene (fig. 6). Like in the case of mouse, the jerboa 4.5SH-repeat contains many simple sequence repeats (SSRs) (fig. 7).

In the mole rat genome, the length of the restriction fragments, containing the 4.5SH RNA gene, are 3.3 and 2.0 kb for *BamHI* and *EcoRI*, respectively. The *EcoRI* fragment has been cloned and partially sequenced (accession number AY828231). The nucleotide sequence of the 4.5SH RNA gene in this clone differed from the mole rat cDNA sequences by several single-nucleotide substitutions and a 10-nt insertion in the 3' part of the gene (fig. 6). However, the internal split promoter and terminator for RNA polymerase III were not changed. Perhaps this gene is a member of the particular 4.5SH RNA gene subfamily.

The flanking sequences contain only short regions of similarity to the mouse and jerboa sequences (fig. 6). The mole rat repeat unit like the mouse one seems to be rich in SSRs, as two simple sequences [(TG)<sub>16</sub> and (GA)<sub>17</sub>] were found in the relatively small (960 bp) sequenced region.

Schoeniger and Jelinek (1986) assessed the length of the rat 4.5S<sub>H</sub>-repeat at 5.3 kb, but they did not determine its nucleotide sequence. In the current study, the rat 2.2-kb *BamHI* fragment containing the 4.5SH RNA gene was cloned and sequenced (accession number AY228160). It appears that the rat 4.5SH RNA gene-flanking sequences demonstrate significantly higher sequence similarity to the mouse (80%), then to the corresponding jerboa and

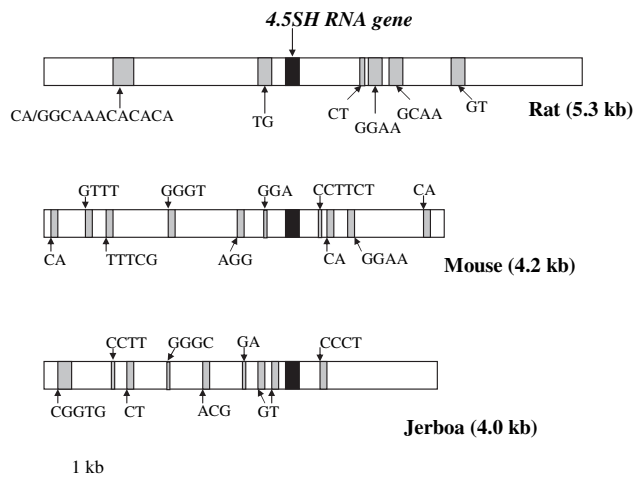


FIG. 7.—Location of the simple sequences in the mouse, rat, and jerboa 4.5SH-repeats. 4.5SH RNA genes are shown as black boxes; shaded boxes correspond to simple sequences. Nucleotide sequences of SSR units are indicated.

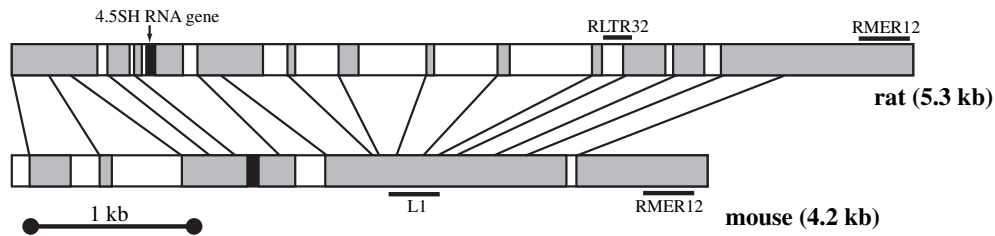


FIG. 8.—The diagram of distribution of homologous regions in the rat and mouse 4.5SH-repeats. 4.5SH RNA genes are shown as black boxes; shadowed boxes represent homologous regions; white boxes are species-specific sequences, lines connect the beginning of corresponding homologous sequences. Regions with sequence similarity to three mobile elements are indicated by thick lines.

mole rat sequences (<50%) (fig. 6). A search in the Rat Genome Database produced three nonannotated full-size tandem units containing the 4.5SH RNA gene on the rat four chromosome (NW 047691.1). It allowed us to compare the full-size 4.5SH-repeats for two closely related species, rat and mouse, as well (fig. 1 in Supplementary Material online).

The alignment of the mouse and rat 4.5SH-repeats showed that the regions of high sequence similarity are alternated by the species-specific sequences (fig. 8). Perhaps such species-specific sequences appeared as a result of accumulation of insertions and deletions after the divergence of those two rodent species. The rat and mouse 4.5SH RNA gene sequences are almost identical. Other homologous regions show about 80% sequence identity that is characteristic of neutrally evolving DNA of these species (O'HUigin and Li 1992). There are many simple sequences (SSRs) in the 4.5SH-repeat of rat as well as mouse, mole rat, and jerboa. Some of the rat SSRs are similar in location and their nucleotide motifs are similar to the mouse SSRs but often differ in their lengths (fig. 7). Obviously, not only

nucleotide substitutions, but also long insertions, deletions, and simple sequence amplification play an important role in the 4.5SH-repeat evolution.

## Discussion

The results of this study demonstrated that 4.5SH RNA genes are distributed among six related rodent families (Muridae, Cricetidae, Spalacidae, Rhizomyidae, Zapodidae, and Dipodidae) and are most likely absent from the representatives of the other rodent families. Thus, 4.5SH RNA is characterized by relatively narrow taxonomic distribution and, therefore, belongs to the stenoRNA group.

Like some other RNAs of this group, 4.5SH RNA demonstrates the visible homology to one of the SINES, the B1 element. However, while the B1 element is characteristic for all Rodentia (Zietkiewicz and Labuda 1996; Vassetzky, Ten, and Kramerov 2003), 4.5SH RNA occurs only in the representatives of six families that comprise the group Myodonta. Thus, 4.5SH RNA arose in evolution much later than B1. Judging by the taxonomic distribution

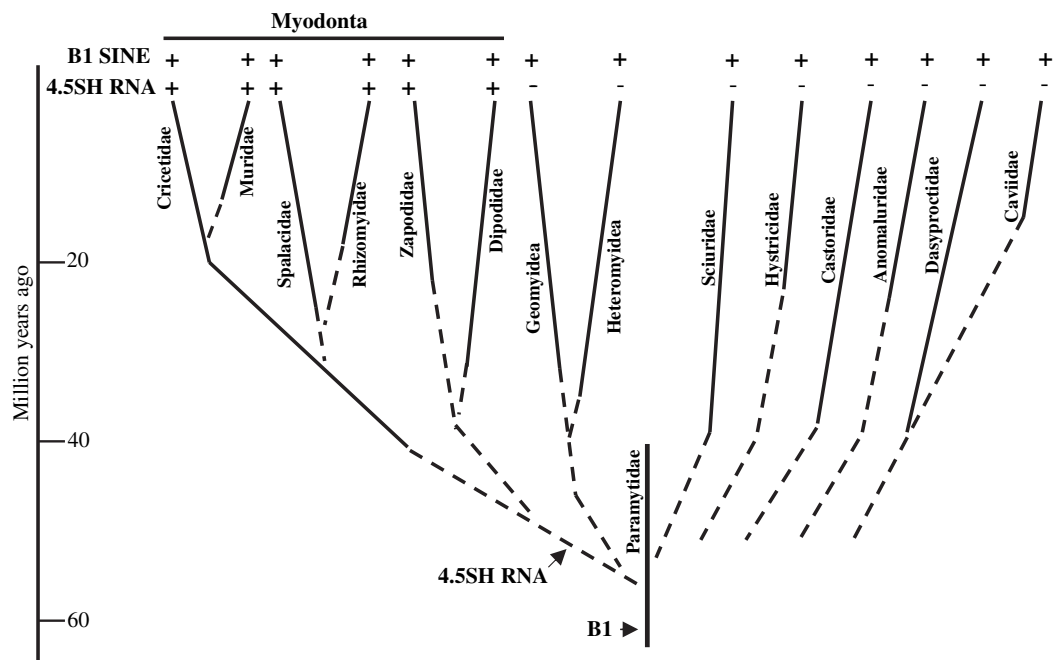


FIG. 9.—Rodent evolutionary phylogenetic tree and the distribution of the 4.5SH RNA genes and B1 SINE. The tree is based on the phylogeny proposed by Romer (1966) and modified according to Kramerov, Vassetzky, and Serdobova (1999). The left-hand cluster of clades is shown completely, while the right-hand one is significantly reduced by comparison with the original; only four of 22 families are shown. Arrowheads indicate likely time points of the emergence of the 4.5SH RNA genes and B1 SINE.

of 4.5SH RNA, its ancestral genes first evolved in a common ancestor of Myodonta (fig. 9). According to paleontological and molecular data, Myodonta began to diverge 50 (Carroll 1988) to 85 (Adkins, Walton, and Honeycutt 2003) MYA, and 4.5SH RNA has to have the similar age. One of the B1 variants (pB1d10) could be the direct precursor of the 4.5SH RNA gene as both pB1d10 and 4.5SH RNA show the same 10-nucleotide deletion in comparison with their common ancestor, small cytoplasmic 7SL RNA (Quentin 1994).

Analysis of cDNA derived from the 4.5SH RNA of different rodents shows the high level of nucleotide sequence identity (84%–98%). Thus, 4.5SH RNA is rather conserved in evolution. However, while displaying high interspecies evolutionary conservation, 4.5SH RNA demonstrates significant heterogeneity in the same species (organism). This heterogeneity may be due to a large number of the 4.5SH RNA genes. The conservation of the 4.5SH RNA nucleotide sequence suggests that this RNA is under natural selection and seems to have a function. Considering its relatively recent origin, the 4.5SH RNA could be involved in some optional mechanism, for example, an additional defense against stresses and parasitic agents.

According to the widely accepted concept, SINEs are “selfish” DNA sequences that have no specific functions in the cell life (Doolittle and Sapienza 1980; Orgel and Crick 1980), although an alternative view exists (Schmid 1998; Allen et al. 2004). As a typical SINE, pB1d10 does not seem to be functional. Therefore, the emergence of the presumably functional 4.5SH RNA from the pB1d10 may be considered as the example of molecular exaptation or recruiting (see Makalowski, Mitchell, and Labuda [1994] for other examples of these processes).

4.5SH RNA genes in the mole rat, hamster, and jerboa genomes are included in the repeat sequences. Like the mouse and rat 4.5SH-repeats, these repeats are tandemly arranged in jerboa DNA. Dipodids are the most distantly related to murids among those rodent families, whose representatives have the 4.5SH RNA. Therefore, 4.5SH-repeats are also most likely to be arranged tandemly in the genomes of rodents of other Myodonta families. It is not clear why the tandem arrangement of the 4.5SH-repeats is so conservative. This type of organization is also characteristic of another small RNA, 5S rRNA (Stambrook 1976; Suzuki, Moriwaki, and Sakurai 1994). The tandem arrangement is believed to be necessary to produce the great quantity of 5S rRNA, coordinate expression of its multiple genes, and maintain their structure homogeneity. The same may be applicable to 4.5SH RNA.

Although the exact size of the 4.5SH-repeat was determined for only three rodent species, it appears to be generally conserved. Even in such distant relatives as mouse and jerboa, the repeat lengths are quite similar, 4.2 and 4.0 kb, respectively. The rat 4.5SH-repeat is only 20% longer than the mouse one (5.3 kb). As the size of the 4.5SH-repeat is conserved in evolution, it seems to be important for a 4.5SH-repeat function.

Nucleotide sequences of the 4.5SH-repeats of the distantly related species (mouse, mole rat, and jerboa) show no apparent similarity, except for the gene itself. A similar phenomenon was shown for 5S rRNA and some sno RNA genes

**Table 1**  
The Percentage<sup>a</sup> of Different Types of Species-Specific Sequences in Rat 4.5SH-Repeat and Rat Euchromatic Genome Determined by Comparison with the Mouse Sequences

	Rat-Specific Sequences			
	In Total (%)	SSRs (%)	Mobile Elements (%)	Nonrepeated (%)
Genome	34	1	25	8
4.5SH-repeat	40	20	3.5	16.5

<sup>a</sup> Presented values are given as percentage of total rat euchromatic genome (Gibbs et al. 2004) or 4.5SH-repeat.

(Cavaille et al. 2000). However, one should notice several short conserved motifs upstream of the 4.5SH RNA genes (fig. 6) that might play a role in their expression. In the closely related species, mouse and rat, long regions with high levels of similarity are interrupted with long species-specific indels. In the homologous regions, average nucleotide sequence identity is 70%–80%, that is typical for sequences that evolved neutrally in the rat and mouse genomes (O’Hugin and Li 1992). The species-specific sequences represent 20% and 40% of the 4.5SH-repeat in mouse and rat, respectively. The latter percentage proved similar to the percentage of the rat-specific sequences (34%) determined while comparing the rat and mouse total euchromatic genomes (table 1) (Gibbs et al. 2004). In the rat euchromatic genome, the species-specific sequences are mainly represented by mobile genetic elements, whereas the contribution of simple and unique sequences is not very significant. On the contrary, in the 4.5SH-repeat, half of the rat-specific regions are SSRs and the rest mainly consist of unique sequences, that is, they are not mobile elements (table 1). We found partial sequences of only three mobile elements in the 4.5SH-repeat: L1 in the mouse, RLTR32 in the rat, and RMER12 in both species (fig. 8). Thus, insertion of the nonmobile sequences and amplification of the simple sequences apparently make greater contribution in the evolution of the 4.5SH-repeat than in the evolution of the euchromatic part of the genome in general.

SSRs also represent a major component of spacers of the rat and mouse 5S rDNA repeats that mainly consist of GC-rich regular and irregular simple sequences. In the 4.5SH-repeat, the SSR contribution is essentially lower, however, their density (SSR per kb) is higher than in the entire genome. Some of such sequences in the rat and mouse 4.5SH-repeat seem to be the homologues, however, others might have emerged de novo after the divergence of these species.

The role of the SSRs in the genome is not quite clear yet. The simple sequences were shown to be involved in the transcription and splicing regulation of some genes (Majewski and Ott 2002). By forming the noncanonical DNA structure (hairpins, triplexes, and quadruplexes) and taking part in the non-Watson-Crick interactions, simple sequences may play a role in maintaining the chromatin organization (Catasti et al. 1999). Despite the nonconserved pattern of distribution along the 4.5SH-repeat, SSRs might be important in the 4.5SH RNA gene expression. In addition, one can speculate that SSR amplification may be

involved in the hypothetical mechanism of the 4.5SH-repeat size maintenance.

In conclusion, we can propose the following evolutionary scenario. The 4.5SH RNA evolution began most likely with the alteration of a copy of pB1d10. Then, the transcript of the new gene could be recruited by cell to help performing one of the existing cellular functions. Probably, a function of this RNA required the high level of its synthesis, and that could lead to further changes of the nucleotide sequence and amplification of the gene. The amplification might involve sequences flanking the gene that resulted in the appearance of the tandemly arranged repeats. The subsequent divergence of the rodents allowed the independent evolution of the 4.5SH-repeat in every rodent lineage. Three genetic mechanisms were likely involved in this evolutionary process: nucleotide substitution, long insertions and deletions, and changes of the SSR lengths in the 4.5SH-repeat. Thereby, the nucleotide sequence similarity of the 4.5SH-repeats of different rodent families had been gradually lost, while the gene structure remained conserved. Similar to other tandem repeats, a high level of nucleotide sequence identity is maintained in different 4.5SH-repeat copies in the same species by the mechanisms of intraspecific homogenization of repetitive-sequence arrays, such as unequal crossing-over between repeat units and gene conversion (Liao 1999). However, these mechanisms are not hundred percent efficient, which provided for heterogeneity of the 4.5SH-repeat including the 4.5SH RNA gene. Functions of the sequences flanking the 4.5SH RNA gene remain unclear and will be the subject of our future research.

### Supplementary Material

Supplementary figure 1 is available at *Molecular Biology and Evolution* online ([www.mbe.oupjournals.org](http://www.mbe.oupjournals.org)).

### Acknowledgments

We are grateful to E. Lyapunova, E. Ivanitskaya, A. Puzachenko, E. Potapova, F. Catzeflis, and R. DeBry for providing animals, tissues, or DNA and to S. Morzunov for helpful discussions. This study was supported by the Russian Foundation for Basic Research (grant 05-04-49553).

### Literature Cited

- Adkins, R. M., A. H. Walton, and R. L. Honeycutt. 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol. Phylogenet. Evol.* **26**:409–420.
- Allen, T. A., S. Von Kaenel, J. A. Goodrich, and J. F. Kugel. 2004. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.* **11**:816–821.
- Carroll, R. L. 1988. *Vertebrate paleontology and evolution*. Freeman and Company, New York.
- Catásti, P., X. Chen, S. V. Mariappan, E. M. Bradbury, and G. Gupta. 1999. DNA repeats in the human genome. *Genetica* **106**:15–36.
- Cavaille, J., K. Buiting, M. Kieffmann, M. Lalande, C. Brannan, B. Horsthemke, J. P. Bachelier, J. Brosius, and A. Huttenhofer. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. USA* **97**:14311–14316.
- DeChiara, T. M., and J. Brosius. 1987. Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. *Proc. Natl. Acad. Sci. USA* **84**:2624–2628.
- Doolittle, W. F., and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**:601–603.
- Eddy, S. R. 1999. Noncoding RNA genes. *Curr. Opin. Genet. Dev.* **9**:695–699.
- Gibbs, R. A. G. M., M. L. Weinstock, D. M. Metzker et al. (231 co-authors). 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**:493–521.
- Gogolevskaya, I. K., and D. A. Kramerov. 2002. Evolutionary history of 4.5SI RNA and indication that it is functional. *J. Mol. Evol.* **54**:354–364.
- Harada, F., Y. Takeuchi, and N. Koto. 1986. Molecular cloning and in vitro transcription of rat 4.5 S RNAH genes. *Nucleic Acids Res.* **14**:1629–1642.
- Harada, F., and N. Kato. 1980. Nucleotide sequences of 4.5S RNAs associated with poly(A)-containing RNAs of mouse and hamster cells. *Nucleic Acids Res.* **8**:1273–1285.
- Harada, F., N. Kato, and H. Hoshino. 1979. Series of 4.5S RNA associated with poly(A)-containing RNAs of rodent cells. *Nucleic Acids Res.* **7**:909–917.
- Kramerov, D. A., I. V. Lekakh, O. P. Samarina, and A. P. Ryskov. 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.
- Kramerov, D., N. Vassetzky, and I. Serdobova. 1999. The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. *Mol. Biol. Evol.* **16**:715–717.
- Krayev, A. S., T. V. Markusheva, D. A. Kramerov, A. P. Ryskov, K. G. Skryabin, A. A. Bayev, and G. P. Georgiev. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucleic Acids Res.* **10**:7461–7475.
- Labuda, D., and E. Zietkiewicz. 1994. Evolution of secondary structure in the family of 7SL-like RNAs. *J. Mol. Evol.* **39**:506–518.
- Lewejohann, L., B. V. Skryabin, N. Sachser et al. (14 co-authors). 2004. Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice. *Behav. Brain Res.* **154**:273–289.
- Liao, D. 1999. Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* **64**:24–30.
- Majewski, J., and J. Ott. 2002. Distribution and characterization of regulatory elements in the human genome. *Genome Res.* **12**:1827–1836.
- Makalowski, W., G. A. Mitchell, D. Labuda. 1994. Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet.* **10**:188–193.
- Martignetti, J. A., and J. Brosius. 1993a. Neural BC1 RNA as an evolutionary marker: guinea pig remains a rodent. *Proc. Natl. Acad. Sci. USA* **90**:9698–9702.
- . 1993b. BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. *Proc. Natl. Acad. Sci. USA* **90**:11563–11567.
- Maxwell, E. S., M. J. Fournier. 1995. The small nucleolar RNAs. *Annu. Rev. Biochem.* **64**:897–934.
- McKeown, M. 1993. The role of small nucleolar RNAs in RNA splicing. *Curr. Opin. Cell Biol.* **5**:448–454.
- Miller, K. S., V. Zbrzezna, and A. O. Pogo. 1984. Interstrand duplexes in Friend erythroleukemia nuclear RNA. The interaction of non-polyadenylated nuclear RNA with polyadenylated



- nuclear RNA and with small nuclear RNAs. *J. Mol. Biol.* **177**:343–368.
- O’Hugin, C., and W. H. Li. 1992. The molecular clock ticks regularly in murid rodents and hamsters. *J. Mol. Evol.* **35**:377–384.
- Orgel, L. E., and F. H. Crick. 1980. Selfish DNA: the ultimate parasite. *Nature* **284**:604–607.
- Quentin, Y. 1994. Emergence of master sequences in families of retroposons derived from 7sl RNA. *Genetica* **93**:203–215.
- Reddy, R., D. Henning, E. Tan, and H. Busch. 1983. Identification of a La protein binding site in a RNA polymerase III transcript (4.5 I RNA). *J. Biol. Chem.* **258**:8352–8356.
- Redston, M. S., and S. E. Kern. 1994. Klenow co-sequencing: a method for eliminating “stops”. *Biotechniques* **17**:286–288.
- Ro-Choi, T. S., R. Reddy, D. Henning, T. Takano, C. Taylor, and H. Busch. 1972. Nucleotide sequence of 4.5S RNA I of Novikoff hepatoma cell nuclei. *J. Biol. Chem.* **247**:3205–3222.
- Romer, A. S. 1966. *Vertebrate paleontology*. University of Chicago Press, Chicago.
- Saba, J. A., H. Busch, and R. Reddy. 1985. A new moderately repetitive rat DNA sequence detected by a cloned 4.5SI DNA. *J. Biol. Chem.* **260**:1354–1357.
- Schmid, C. W. 1998. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* **26**:4541–4550.
- Schoeniger, L. O., and W. R. Jelinek. 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, I. M., and D. A. Kramerov. 1998. Short retroposons of the B2 superfamily: evolution and application for the study of rodent phylogeny. *J. Mol. Evol.* **46**:202–214.
- Slatko, B. E., L. M. Albright, and S. Tabor. 1992. DNA sequencing by the dideoxy method. 1992. Pp. 740–741 in F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Mooko, J. G. Seidman, J. A. Smith, and K. Struhl, eds. *Short protocols in molecular biology*. Wiley, New York.
- Stambrook, P. J. 1976. Organisation of the genes coding for 5S RNA in the Chinese hamster. *Nature* **259**:639–641.
- Suzuki, H., K. Moriwaki, and S. Sakurai. 1994. Sequences and evolutionary analysis of mouse 5S rDNAs. *Mol. Biol. Evol.* **11**:704–710.
- Takeuchi, Y., and F. Harada. 1986. Cloning and characterization of rat 4.5S RNAI genes. *Nucleic Acids Res.* **14**:1643–1656.
- Vassatzky, N. S., O. A. Ten, and D. A. Kramerov. 2003. B1 and related SINEs in mammalian genomes. *Gene* **319**:149–160.
- Zietkiewicz, E., and D. Labuda. 1996. Mosaic evolution of rodent B1 elements. *J. Mol. Evol.* **42**:66–72.

David Irwin, Associate Editor

Accepted March 21, 2005