4.5S$_I$ RNA genes and the role of their 5′-flanking sequences in the gene transcription

Irina K. Gogolevskaya, Dmitri A. Kramerov *

Laboratory of Eukaryotic Genome Evolution, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov Street, Moscow 119991, Russia

ARTICLE INFO

ABSTRACT

4.5S$_I$ RNA is a small nuclear RNA synthesized by RNA polymerase III and detected in rodents of only four families. Hundreds of copies of this RNA retropseudogenes are interspersed throughout the mouse (*Mus musculus*) and rat (*Rattus norvegicus*) genomes. We found a single locus containing 4.5S$_I$ RNA genes in the genomes of these rodents. The locus harbors three genes and occupies 80 kb on *M. musculus* chromosome 6 and 44 kb on *R. norvegicus* chromosome 4. Two long duplications seem to have taken place during evolution of this locus. Two mouse 4.5S$_I$ RNA genes were used for a study of the role of 5′-flanking sequences in transcription *in vitro* and *ex vivo*. We found that removal of these DNA sequences resulted in a dramatic reduction of transcription though an internal promoter for RNA polymerase III was preserved in 4.5S$_I$ RNA genes. Thus, 5′-flanking sequences (from −1 to −90) containing conserved regions are important for 4.5S$_I$ RNA gene expression.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

4.5S$_I$ RNA is a 98 nt nuclear RNA with unknown function which is synthesized by RNA polymerase III (pol III) (Ro-Choi et al., 1972; Reddy et al., 1983). This RNA is characteristic of rodents of only four related families: Muridae (mice and rats), Cricetidae (hamsters and voles), Spalacidae (mole rats), and Rhizomyidae (bamboo rats) (Gogolevskaya and Kramerov, 2002).

It was suggested that 4.5S$_I$ RNA genes are interspersed throughout the rat genome (Takeuchi and Harada, 1986), but their copy number in genome was not determined exactly. Additionally in the mouse and rat genomes, numerous retropseudogenes (446 and 1453, respectively) corresponding to 5′-half of 4.5S$_I$ RNA sequence were found (Saba et al., 1985; Schmitz et al., 2004). Probably 4.5S$_I$ RNA originated from one of the short interspersed elements (SINEs) of the rodent genome: B2 or ID (Serdobova and Kramerov, 1998).

4.5S$_I$ RNA genes contain an internal promoter, which consists of A and B boxes (each 11 bp long) and is recognized by pol III. Such a type 2 pol III promoter is present in viral genes (VAI and II, EBERI and II), tRNAs and 7SL RNA genes, as well as in SINEs derived from them (Schramm and Hernandez, 2002; Kramerov and Vassetzky, 2005). Box B is recognized by TFIIIC (a transcription factor of pol III), which directly interacts with DNA binding boxes B and A. TFIIIC recruits another transcriptional factor, TFIIIB, that involves pol III to promoter complex. In many genes, the internal promoter is enough for efficient transcription both *in vitro* (Guilfoyle and Weinmann, 1981; Hofstetter et al., 1981; Bhat et al., 1983; Perez-Stable et al., 1984) and *in vivo* (Galli et al., 1981; Hofstetter et al., 1981; Ciliberto et al., 1982). However, there are sequences in the upstream regions of some genes that are necessary for their transcription (e.g. 7SL RNA, (Ullu and Weiner, 1985) and BC1 RNA (Martignetti and Brosius, 1995)) and 5′-flanking regions can modulate pol III transcription in many other genes (Chesnokov and Schmid, 1996; Dieci et al., 2007).

Takeuchi and Harada (1986) cloned several genes and pseudogenes of 4.5S$_I$ RNA and estimated the copy numbers of the genes in the haploid rat genome to be 50–300. They also noted a similarity of the 14 bp sequences located immediately upstream of the two 4.5S$_I$ RNA genes. This could point to the role of these 5′-flanking DNA sequences in transcription of these genes.

Here, we analyzed the mouse and rat genomes and concluded that they contain only three 4.5S$_I$ RNA genes located in a single locus. By removing regions of 5′-flanking sequences we showed their importance for transcription both *in vitro* and in experiments with transfection of the human cells by 4.5S$_I$ genes.

2. Materials and methods

2.1. Search for of 4.5S$_I$ RNA genes and pseudogenes in genome databases

Sequences similar to 4.5S$_I$ RNA in the genomes of mouse and rat were searched using the BLAST-like alignment tool (BLAT) in the

UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgBlat). Multiple alignments were produced by the ClustalW program and manually adjusted in GeneDoc.

2.2. Plasmid constructs

4.5S$_I$ RNA genes Mmu1′ and Mmu2 with ~400 bp upstream sequences were isolated by PCR using mouse genomic DNA as a template and oligonucleotides as primers (Supplementary Table 1). PCR products were purified by electrophoresis in 2% agarose gel and cloned into plasmid pGEM-T (Promega, Madison, WI, USA) following the manufacturer's protocol.

4.5S$_I$ RNA genes with partially or completely deleted upstream sequences were obtained by PCR using cloned Mmu1′ and Mmu2 genes (see above) as templates and specific primers (Supplementary Table 1). PCR products were isolated by electrophoresis in 3% agarose gel and cloned into pGEM-T. Plasmids were isolated using a variant of alkaline method (Lee and Rasheed, 1990). All constructs were sequenced in order to avoid nucleotide substitutions introduced during PCR. A strong transcriptional dependence on the orientation of pol III-transcribed genes in this vector has recently been reported (Koval and Kramerov, 2009). Hence, the transcription and transfection experiments were conducted using plasmids with the same orientation of DNA insertions: the 5′ end of 4.5S$_I$ RNA genes faced the site complementary to the M13 Reverse primer in pGEM-T.

2.3. Transcription in vitro and RNA analysis

Transcription *in vitro* was performed as described by Weil et al. (1979), using whole EAC cell extract. Reaction mixture contained 0.5 μg plasmid DNA, 14 μl cell extract, and 12.5 μCi [α^{32}P]GTP in a final volume of 50 μl of 10 mM Hepes, pH 7.9, 70 mM KCl, 2.5 mM MgCl$_2$, 0.5 mM DTT, 3 μmol of ATP, CTP, and UTP, and 1.25 μmol GTP. The mixture was incubated at 30 °C for 1 h . After incubation, RNA from the reaction mixture was purified by phenol-chloroform extraction and precipitated with ethanol. RNA was fractionated by electrophoresis in 6% PAAG containing 7 M urea. The gel was dried, exposed against an X-ray film and quantitated by Cyclone phosphoroimager (Perkin-Elmer Life Sciences Inc., Boston, MA).

2.4. Cell transfection and Northern blot analysis

Transfections were carried out in HeLa cells. Monolayer was grown to 80% confluency in 60 mm Petri dishes. Cells were transiently transfected with 4.5 μg of plasmid DNA (4.5S$_I$ RNA gene construct) using ExGen 500 *in vitro* Transfection Reagent (Fermentas, Vilnius, Lithuania) following the manufacturer's protocol. Co-transfection with 0.5 μg of SOR SINE-containing construct was used to control the efficiency of transfection. (SOR is tRNA-derived SINE from the common shrew genome, family Soricidae (Borodulina and Kramerov, 2001). The construct was prepared by Olga Borodulina from Sar4SA copy of this SINE by adding pol III terminator at the 3′ end of the SINE). RNA was isolated 20 h post-transfection using the guanidinium thiocyanate method (Chomczynski and Sacchi, 1987). Following ethanol precipitation, RNA was treated with 100 μg/ml RNase-free DNase I at 37 °C for 30 min and purified by chloroform extraction. RNA was resolved by denaturing electrophoresis in 6% PAAG and transferred onto Hybond-XL membrane (GE Healthcare UK Ltd., Buckinghamshire, England) by semidry electroblotting at 5 V for 1 h. The 4.5S$_I$ RNA was detected by hybridization with ^{32}P-labeled probe obtained by PCR (Gogolevskaya and Kramerov, 2002). The blot was incubated overnight with the probe in 50% formamide, 5× Denhardt solution, 4× SSC, 1% SDS, and 0.1 mg/mL salmon sperm DNA at 42 °C. The membrane was washed in 0.1× SSC and 0.1% SDS at 52 °C, exposed against an X-ray film, and scanned by Cyclone phosphoroimager.

3. Results

3.1. There are only three 4.5S$_I$ RNA genes in the mouse and rat genomes

We searched for DNA regions similar to 4.5S$_I$ RNA sequences in the mouse and rat genomes using UCSC database. Most of the regions found corresponded to the sequence of the 5′-half of this RNA. (We suggested to denote it as 4.5S$_I$-5′). Such DNA sequences were previously revealed by Saba et al. (1985), and Schmitz et al. (2004) classified them as tailless retroposons and estimated their copy numbers as 446 and 1453 in the mouse and rat genomes, respectively. However, the structure of this DNA repeat has not been analyzed in detail.

Using Blat program, we discovered 183 and 187 copies of such 4.5S$_I$-5′ elements in the mouse and rat genomes, respectively (Table 1), that demonstrated a high level (96–100%) of sequence similarity to 4.5S$_I$ RNA. Though there is some heterogeneity in the length of 4.5S$_I$-5′ elements (from 48 to 59 bp), there are the major variants: 68% (mouse) and 76% (rat) 4.5S$_I$-5′ copies have a length of 56–57 bp (Supplementary Tables 2a, b). 56 bp element terminates with the …AAAAATATAAGA sequence whereas 57 bp element has additional G at the 3′ end. All 4.5S$_I$-5′ copies analyzed are flanked by perfect short direct repeats (or TSD, target sequence duplication); this, together with sequence similarity of the copies, suggests their recent origin.

Additionally, both several truncated (Supplementary Tables 2a, b) and full-length 4.5S$_I$ RNA-derived DNA sequences were found that have a poly(A)-tail and TSD (Table 1; Supplementary Figs. 1a, b, 2, and 3a, b). Such features suggest that these DNA regions are retroelements that emerged due to LINE1 (L1) activity (Jurka, 1997; Deininger and Batzer, 2002). The full-length retroelements found fall into two groups: (i) simple elements containing only sequences very similar to 4.5S$_I$ RNA (Supplementary Figs. 1a, b) and (ii) "chimeric" elements composed of two monomers: the left one is derived from 4.5S$_I$ RNA and the right one is either SINE B1 (Supplementary Fig. 2) or L1 sequences (Supplementary Figs. 3a, b). A single copy of 4.5S$_I$-L1 element was found both in the mouse and rat genome but in different location. In the majority of both the simple and the "chimeric" retroelements, 4.5S$_I$ RNA-derived sequences had nucleotide substitutions. However, one copy of 4.5S$_I$-B1 element comprised 4.5S$_I$ RNA sequence with no nucleotide substitutions (copy 1 in Supplementary Fig. 2).

Finally, we found a locus on the mouse chromosome 6 and the rat chromosome 4 which contained three regions exactly corresponding to 4.5S$_I$ RNA sequence. They seem to be the genes coding for 4.5S$_I$ RNA. Three mouse genes are spaced by 39–40 kb from each other (Fig. 1). Two of them (genes Mmu1 and Mmu1′) have the same orientation and are flanked by similar 25 kb sequences interrupted by mobile elements in several sites (Supplementary Fig. 4). Evidently, these sequences arose by DNA duplication in an ancestor of *M. musculus*. There is no evidence of

Table 1
Types of 4.5S$_I$ RNA-derived sequences.

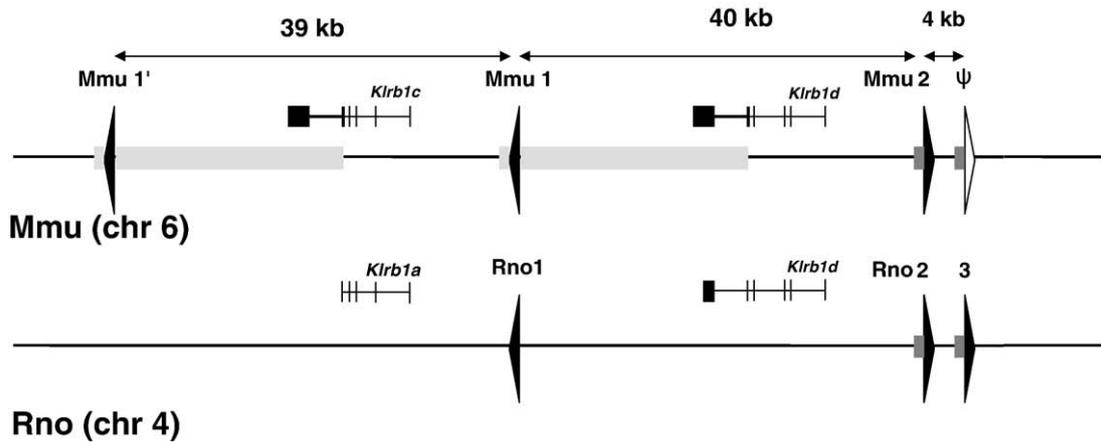| DNA sequence type			Copy number (*Mus musculus*)	Copy number (*Rattus norvegicus*)
4.5S$_I$ RNA genes			3	3
Full-length 4.5S$_I$ RNA-derived DNA	Simple elements		4	6
	"Chimeric" elements	4.5S1-B1	4	0
		4.5S1-L1	1	1
Truncated 4.5S$_I$ RNA retropseudogenes (from 61 to 65 bp)			8	6
4.5S$_I$-5′ elements (from 48 to 59 bp)			183	187

Fig. 1. The arrangement of 4.5S$_I$ RNA genes in the mouse (Mmu) and rat (Rno) genomes. 4.5S$_I$ RNA genes are marked by filled arrowheads and a pseudogene (Ψ) is depicted as an empty arrowhead. The arrowheads directed right and left correspond to the genes in direct and reverse orientation, respectively. The regions with similar nucleotide sequences flanking the mouse (Mmu1′ and Mmu1) and rat (Rno2 and Rno3) 4.5S$_I$ RNA genes are shown by grey rectangles. The mouse and rat protein-coding _Krlb1_ genes are shown above the locus of Mmu chr6 and Rno chr4, respectively. The _Krlb1_ genes are transcribed from right to left. The positions of the identified 4.5S$_I$ RNA genes: on mouse chromosome 6 — 128710094–128710194 (Mmu1′), 128748878–128748978 (Mmu1), 128789067–128789167 (Mmu2), and 128793605–12879363 (φ), assembly, July 2007; on rat chromosome 4 — 165683814–165683914 (Rno1), 165727937–165728037 (Rno2), 165732555–165732655 (Rno3), assembly, July, 2004.

such duplication in the rat genome and the gene 1′ is absent from _R. norvegicus_. In the genomes of both species there is gene 2 which is in opposite orientation to the gene 1 (Fig. 1). There is 4.5S$_I$ RNA gene 3 located 4 kb from the gene 2 in the rat genome. In mouse, the gene 3 contains long deletions indicating that it is a non-functional pseudogene. The rat genes 2 and 3 as well as the mouse gene 2 and the pseudogene (Ψ) are 5′-flanked by similar sequences about 700 bp long (Fig. 1 and Supplementary Fig. 5), suggesting that DNA duplication took place in the common ancestor of _Mus_ and _Rattus_.

We aligned all three mouse and rat genes with their 5′-flanking sequences (Fig. 2). The nucleotide sequences of 4.5S$_I$ RNA genes are identical in all 6 analyzed copies except Rno1 gene, which has a nucleotide substitution (T in position 48). This result is in agreement with our previous data (Gogolevskaya and Kramerov, 2002) on existence of two rat 4.5S$_I$ RNA variants with A or U in this position. On the contrary, we found only a short CA-motif located immediately

downstream of the gene terminator. On the other hand, the conserved sequences are revealed in 100 bp long 5′-flanking region (Fig. 2). The most conserved sequences are the boxes AGAAT (from −14 to −18) and GC(C/A)ACGCCT (from −36 to −46).

3.2. Influence of 5′-flanking sequences on the 4.5S$_I$ RNA gene transcription

Although boxes A and B forming the internal promoter play a key role in gene transcription by pol III both _in vitro_ and _in vivo_, 5′-flanking sequences can sometimes influence the transcription level (see Section 1 for references). Therefore we studied the significance of these sequences for 4.5S$_I$ RNA gene transcription using five constructs obtained for each of the two mouse genes (Mmu1′ and Mmu2), in which only the part of 5′-flanking sequences was preserved: from the first nucleotide to position −26, −44, −65, −87, or about −400. (In
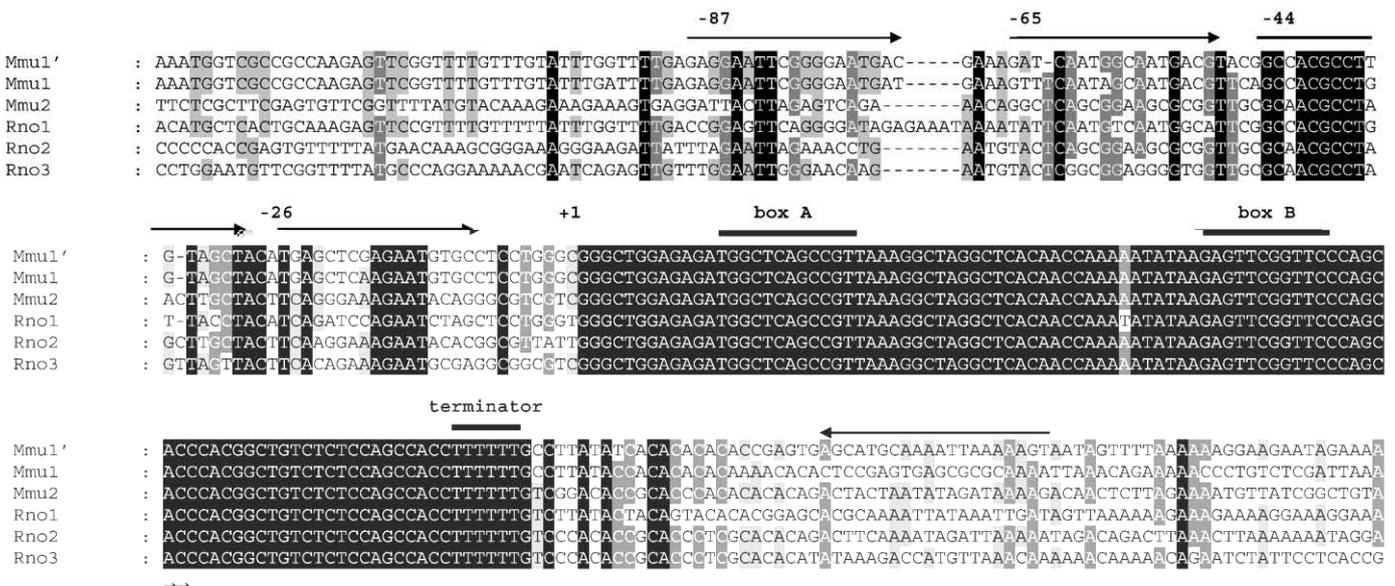

Fig. 2. The alignment of three mouse (Mmu) and three rat (Rno) 4.5S$_I$ RNA genes with 90 bp upstream sequences. The transcription start site is marked "+1". Promoter (box A and box B) and terminator (T$_6$) for pol III are shown. Conserved sequences are highlighted by black background. Two mouse 4.5S$_I$ RNA genes (Mmu1′ and Mmu2) were used in transcription _in vitro_ and transfection experiments. Constructs for these experiments were obtained by PCR (positions of primers are marked with arrows; numbers above the alignment indicate positions of the beginning of each constructs).

this experiment only two of three mouse genes were used, because upstream sequences of Mmu1 and Mmu1′ genes are almost identical).

Fig. 3A shows the result of transcription of these constructs by pol III using the extract of EAC cells. Graphs were plotted using the quantitative estimations of these data (Fig. 3B). Patterns of transcription for the two genes were somewhat different. In the case of Mmu2 gene, the transcriptional maximum was observed for the −87 construct; further truncation of the 5′-flanking sequences led to a gradual decrease of transcription. The result of Mmu1′ gene transcription *in vitro* was rather surprising: a very low transcription level was observed for −87 construct, whereas transcription of −65 construct was maximal. (It should be noted that such transcriptional patterns were reproduced when we used even the plasmids isolated in different experiments). For both genes, +1 constructs (i.e. without native 5′-flanking sequences) were transcribed most ineffectively: the level of the transcription was 8 and 9 times lower than the maximal level of the transcription for −65 construct in Mmu1′ gene and for −87 construct in Mmu2 gene, respectively. In general, the data obtained indicate that 5′-flanking sequences contribute to the

efficient transcription of $4.5S_I$ RNA. However, more complicated interactions seem to take place: for example, the region from −65 to −87 in Mmu1′ gene could repress transcription *in vitro*.

For further investigation of the 5′-flanking sequence function in $4.5S_I$ RNA transcription, the human HeLa cell transient transfections by the above mentioned constructs were carried out. (Mouse or rat cells could not be used in such experiment since all of them contain a lot of endogenous $4.5S_I$ RNA). A plasmid that contained SINE SOR from the shrew genome also transcribed by pol III was used as an internal control of transfection. The results of Northern hybridization of $4.5S_I$ RNA and SOR RNA from transfected cells are shown on Fig. 4A, whereas quantitative estimations of $4.5S_I$ RNA expression normalized on SOR RNA are presented on Fig. 4B. The obtained results indicate that step-by-step removal of 5′-flanking sequences leads to the gradual decline of transcription level of both $4.5S_I$ RNA genes (Mmu1′ and Mmu2). The constructs without 5′-flanking "native" sequences were transcribed at the very low level (the transcription was 30 and
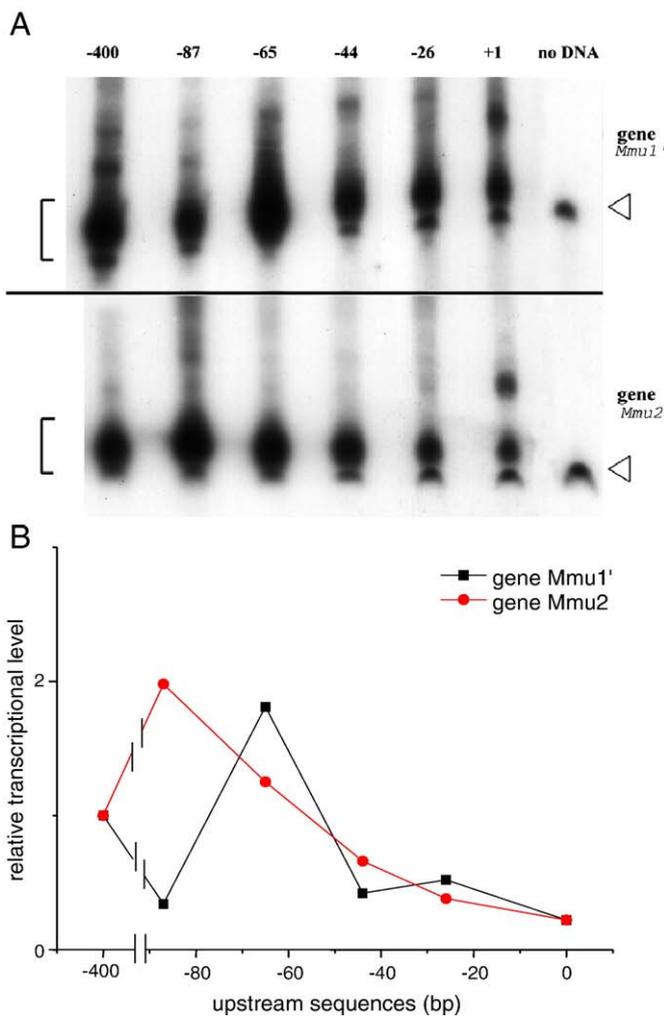

Fig. 3. Transcription of two mouse $4.5S_I$ RNA genes by the extract of EAC cells. (A) The autoradiograph of the transcription products separated by PAGE. The numbers above the lanes indicate from which position an upstream sequence begins in each of the transcribed constructs. "No DNA" — the result of the reaction without adding template (the arrow indicates a band generated by nontemplate synthesis). $4.5S_I$ RNAs are marked with square brackets. The longer RNA observed in the case of the construct +1 probably originated as a result of transcription from the site located upstream of the normal start nucleotide. A similar case was described by Koval and Kramerov (2009). (B) The graphs showing the relative transcription levels of the constructs with Mmu1′ and Mmu2 genes. The transcription level for the −400 upstream nucleotide sequence construct was set as one unit.
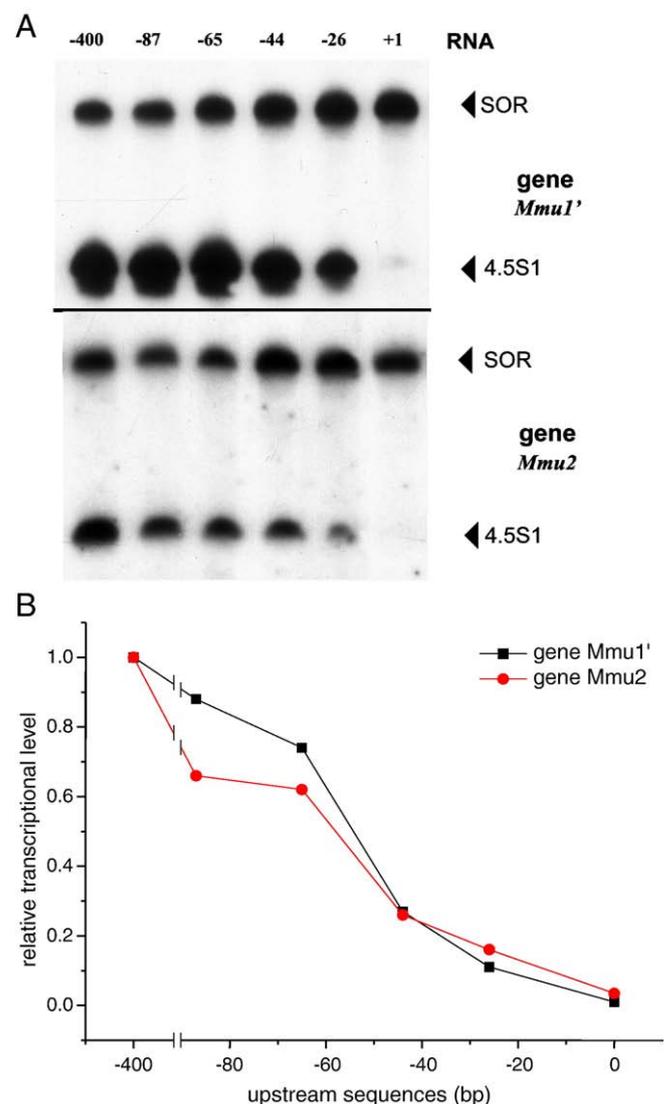

Fig. 4. The expression of two $4.5S_I$ RNA genes in HeLa cells transfected by the genes with upstream sequences of various lengths. (A) Northern blot hybridization of RNA isolated from transfected cells. The numbers above the lanes indicate the position from which the upstream sequence begins in each of the transcribed constructs carrying $4.5S_I$ RNA gene. Co-transfection with SINE SOR was carried out for the control of transfection efficacy. (B) The graphs illustrated the relative level of the Mmu1′ and Mmu2 gene expression in the transfection experiments. The quantity of $4.5S_I$ RNA was normalized by the quantity of SOR RNA. The transcription level for the −400 upstream nucleotide sequence construct was assigned one unit.

80 times lower as compared to maximal transcription level in −400 construct for Mmu1′ and Mmu2 genes, respectively). Hence, native 5′-flanking sequences are important for effective transcription of 4.5S$_I$ RNA genes in transfected cells.

4. Discussion

4.1. Genes and pseudogenes of 4.5S$_I$ RNA and their nomenclature

Contrary to the earlier estimations of 4.5S$_I$ RNA gene copy numbers (50–300) in the rat genome (Takeuchi and Harada, 1986), we showed that there were only three copies of these genes in the mouse and rat genomes. For both species they are at the same locus. In the mouse genome, 4.5S$_I$ RNA gene locus resides on chromosome 6 and occupies about 80 kb, whereas in the rat genome the locus is on chromosome 4 and extends for 44 kb (Fig. 1). Undoubtedly, the mouse genes M1 and M2 are orthologues of the rat genes R1 and R2, respectively. On the other hand, the rat gene R3 is an orthologue of the mouse pseudogene (Ψ) severely damaged by internal deletions. Presumably, there was a duplication in the common ancestor of *Mus* and *Rattus* that resulted in emergence of genes 2 and 3 located at a distance 4 kb from each other (Figs. 1 and 5). (Judging from nucleotide substitution percentage in extended upstream sequences, the duplication occurred about 19 Mya). Later, another genomic duplication (25 kb long) in line *Mus* led to the emergence of two gene 1 copies, thus the number of functional genes equal three was maintained (Fig. 1). The upstream sequences of genes 1 and 1′ differed only by 3.7% nucleotide substitutions; it suggests that the duplication took place about 2 Mya (Fig. 5). Thus, this event happened much more recently than the *Mus* and *Rattus* split (12 Mya). Notably, in the mouse genome the same event resulted in the appearance of the sixth exon in the *Klrb1c* gene encoding a member of subfamily B killer cell lectin-like receptor (Fig. 1). Perhaps homologous recombination between *Klrb1a* and *Klrb1d* had played a key role in the 25 kb duplication.

It is necessary to address an issue of the nomenclature of the DNA sequences studied here. Three 4.5S$_I$ RNA genes which we identified in each of the mouse and rat genomes were annotated on website of UCSC Genome Browser as chimeric SINEs composed of two parts:
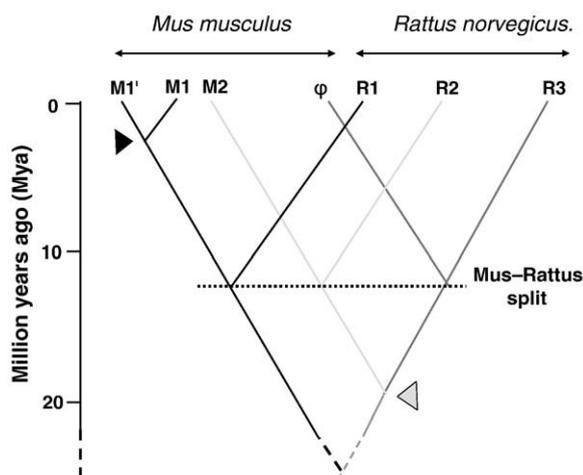
Repeat B4A (Family B4) and Repeat B3 (Family B2). Probably, it resulted from automatic annotation: there are short nucleotide similarities between 4.5S$_I$ RNA and B4A/B3 repeats. We think it is a mistake to consider these genes as SINEs; hence we strongly suggest they must be annotated as short nuclear 4.5S$_I$ RNA genes.

Like other short nuclear and cytoplasmic RNAs (Brosius, 1999), 4.5S$_I$ RNA can generate retropseudogenes. We found 4 and 6 full-length retropseudogenes of 4.5S$_I$ RNA in the mouse and rat genomes, respectively. Several pseudogenes were truncated. All these retropseudogenes had an A-rich tail and were flanked by TSD. Interestingly, besides such simple retropseudogenes we discovered some chimeric retroelements composed of two parts: a 4.5S$_I$ RNA-derived sequence and SINE B1 or 3′-region of LINE L1. Previously, Buzdin (2004) reported a number of chimeric retroelements consisting of small RNA sequences (mostly U6 RNA) and a 3′-end region of L1. Obviously, the 4.5S$_I$ RNA-containing chimeric elements should be ranged in the same category of retroelements. These chimeric elements are supposed to arise as a result of RNA-template switch during reverse transcription. Interestingly, one copy of the mouse chimeric 4.5S$_I$–B1 element comprised a 4.5S$_I$ RNA-derived sequence with no nucleotide substitutions. Active transcription of this sequence resulting in normal 4.5S$_I$ RNA cannot be ruled out. However, the upstream region of this element has no similarity to the 5′-flanking sequences of canonic 4.5S$_I$ RNA genes, which are crucial for their transcription (see Sections 3.2 and 4.2). Therefore, active transcription of this element seems unlikely.

For an unknown reason the 5′-end half of 4.5S$_I$ RNA (retroelements 4.5S$_I$-5′) multiplied much more effectively than the full-length (or slightly truncated) 4.5S$_I$ RNA retropseudogenes in the mouse and rat genomes. The copy number of 4.5S$_I$-5′ elements runs into some hundreds and, as a rule, these elements have no A-rich tail (Schmitz et al., 2004). However, it should be noted that there is a 10-nt A-rich conserved region at the 3′-end of 4.5S1-5′ retroelements; this 10-nt region is located approximately in the middle of the full-length 4.5S$_I$ RNA sequence and forms a loop in the secondary structure of this RNA (Fig. 2 and Supplementary Tables 2a, b). Probably, the 10-nt A-rich region can function in retroposition process like a long variable A-rich tail, a usual feature of most SINEs, LINEs, and retroelements in mammals (Deininger and Batzer, 2002).

4.2. Significance of 5′-flanking sequences for 4.5S$_I$ RNA gene transcription

As the transcribing sequence of 4.5S$_I$ RNA genes contains canonical boxes A and B, the promoter of these genes can be ranged in the type 2 of pol III promoters. Usually, boxes A and B box are sufficient for active transcription from such promoters *in vitro* and *in vivo* (see Section 1 for references). At the same time there are some genes with type 2 promoter for which the transcription is affected or completely stopped by removing "native" 5′-flanking DNA sequences (Section 1).

Here we tested whether 5′-flanking sequences are significant for 4.5S$_I$ RNA gene transcription *ex vivo* and *in vitro*. To address this question, the constructs of two mouse genes (Mmu1′ and Mmu2) containing the upstream sequences of the different length were obtained. Although the gene without upstream sequences (+1 construct) showed rather effective transcription *in vitro,* the constructs with longer upstream sequences were transcribed much more efficiently. The maximal transcription for Mmu1′ gene was observed for −65 construct; in the case of Mmu2 gene, the maximum was for −87 construct. For these constructs the transcription level was approximately tenfold the basal level. The longest upstream sequence (−400) constructs were also transcribed more effectively (fivefold) than the genes without 5′-adjacent native sequence. Although, in general the upstream sequences substantially increase 4.5S$_I$ RNA gene transcription *in vitro,* the region from −87 to −65 in gene Mmu1′ seems to negatively influence the gene transcription.

Fig. 5. The most likely scenario of the 4.5S$_I$ RNA gene evolution in the *Mus* and *Rattus* lineages. The duplications of the 4.5S$_I$ RNA genes with adjacent sequences in mouse as well as in the common *Mus* and *Rattus* ancestor are marked by the triangles. The horizontal dashed line shows the time (12 Mya) of split between the *Mus* and *Rattus* lineages (Benton et al., 2009). The time scale (Mya) is shown on the left. The time of the duplications was estimated based: (i) on the proportion of nucleotide substitutions in the long (800–1000 bp) sequences upstream of the genes, and (ii) on the rate of nucleotide substitutions (1% per Myr) in mouse and rat sequences subject to no or weak selection (O'HUigin and Li, 1992).

In HeLa cell transfection experiments, the effect of the upstream sequences on 4.5S$_I$ RNA gene transcription is even more significant than in the *in vitro* transcription experiments. In transfected cells, the level of expression of the gene lacking native −400 bp 5′-flanking sequence, was almost abolished, i.e. it was 30–80 fold lower than the expression of the native gene (Figs. 4A, B). An especially strong reduction of the transcription took place upon removing the region from −67 to −44. Interestingly, there are several conserved boxes in the region spanning from the gene transcription start site to −90 position that were found by alignment of the mouse and rat genes (Fig. 2). Among these conserved boxes neither TATA-boxes (which are frequent in yeast and plant tRNA genes (Hamada et al., 2001; Giuliodori et al., 2003)), nor DSE- and PSE-like elements (which are characteristic for upstream sequences of U6 RNA genes (Lobo and Hernandez, 1989)) were present. It is quite possible that the conserved boxes revealed in 5′-flanking sequences of 4.5S$_I$ RNA genes are important for expression of these genes and are subject to natural selection.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2009.11.007.

References

Benton, M.J., Donoghue, P.C.J., Asher, R.J., 2009. Calibrating and constraining molecular clocks. In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford University Press, pp. 35–86.

Bhat, R.A., Metz, B., Thimmappaya, B., 1983. Organization of the noncontiguous promoter components of adenovirus VAI RNA gene is strikingly similar to that of eucaryotic tRNA genes. Mol. Cell. Biol. 3, 1996–2005.

Borodulina, O.R., Kramerov, D.A., 2001. Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mamm. Genome 12, 779–786.

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

Buzdin, A.A., 2004. Retroelements and formation of chimeric retrogenes. Cell. Mol. Life Sci. 61, 2046–2059.

Chesnokov, I., Schmid, C.W., 1996. Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors. J. Mol. Evol. 42, 30–36.

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

Ciliberto, G., Traboni, C., Cortese, R., 1982. Relationship between the two components of the split promoter of eukaryotic tRNA genes. Natl. Acad. Sci. U.S.A. 79, 1921–1925.

Deininger, P.L., Batzer, M.A., 2002. Mammalian retroelements. Genome Res. 12, 1455–1465.

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., Pagano, A., 2007. The expanding RNA polymerase III transcriptome. Trends Genet. 23, 614–622.

Galli, G., Hofstetter, H., Birnstiel, M.L., 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature 294, 626–631.

Giuliodori, S., et al., 2003. A composite upstream sequence motif potentiates tRNA gene transcription in yeast. J. Mol. Biol. 333, 1–20.

Gogolevskaya, I.K., Kramerov, D.A., 2002. Evolutionary history of 4.5SI RNA and indication that it is functional. J. Mol. Evol. 54, 354–364.

Guilfoyle, R., Weinmann, R., 1981. Control region for adenovirus VA RNA transcription. Proc. Natl. Acad. Sci. U.S.A. 78, 3378–3382.

Hamada, M., Huang, Y., Lowe, T.M., Maraia, R.J., 2001. Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. Mol. Cell. Biol. 21, 6870–6881.

Hofstetter, H., Kressman, A., Birnstiel, M.L., 1981. A split promoter for a eucaryotic tRNA gene. Cell 24, 573–585.

Jurka, J., 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. U.S.A. 94, 1872–1877.

Koval, A.P., Kramerov, D.A., 2009. 5′-flanking sequences can dramatically influence 4.5SH RNA gene transcription by RNA polymerase III. Gene 446, 75–80.

Kramerov, D.A., Vassetzky, N.S., 2005. Short retroposons in eukaryotic genomes. Int. Rev. Cytol. 247, 165–221.

Lee, S.Y., Rasheed, S., 1990. A simple procedure for maximum yield of high-quality plasmid DNA. Biotechniques 9, 676–679.

Lobo, S.M., Hernandez, N., 1989. A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell 58, 55–67.

Martignetti, J.A., Brosius, J., 1995. BC1 RNA: transcriptional analysis of a neural cell-specific RNA polymerase III transcript. Mol. Cell. Biol. 15, 1642–1650.

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

Perez-Stable, C., Ayres, T.M., Shen, C.K., 1984. Distinctive sequence organization and functional programming of an Alu repeat promoter. Proc. Natl. Acad. Sci. U. S. A. 81, 5291–5295.

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

Ro-Choi, T.S., 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.

Saba, J.A., 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.

Schmitz, J., Churakov, G., Zischler, H., Brosius, J., 2004. A novel class of mammalian-specific tailless retropseudogenes. Genome Res. 14, 1911–1915.

Schramm, L., Hernandez, N., 2002. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 16, 2593–2620.

Serdobova, I.M., Kramerov, D.A., 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.

Ullu, E., Weiner, A.M., 1985. Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature 318, 371–374.

Weil, P.A., Segall, J., Harris, B., Ng, S.Y., Roeder, R.G., 1979. Faithful transcription of eukaryotic genes by RNA polymerase III in systems reconstituted with purified DNA templates. J. Biol. Chem. 254, 6163–6173.