Genomics 94 (2009) 11–19

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Analysis of C/D box snoRNA genes in vertebrates: The number of copies decreases in placental mammals

Julia A. Makarova, Dmitri A. Kramerov *

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 32 Vavilov St., Moscow 119991, Russia

ARTICLE INFO

Article history:
Received 18 December 2008
Accepted 27 February 2009
Available online 9 March 2009

Keywords:
C/D box snoRNA genes
2′-O-methylation
rRNA
Xenopus tropicalis
Pseudogenes
Ribosome
Vertebrates

ABSTRACT

C/D box small nucleolar RNAs (snoRNAs) guide site-specific 2′-O-methylation of RNAs. Nearly all C/D box snoRNAs with known targets are involved in rRNA modification. In vertebrates, snoRNAs are encoded in introns of various genes and their processing is coupled with splicing of host gene pre-mRNA. Here, the genes encoding C/D box snoRNAs that guide 2′-O-methylation of rRNA were identified and analyzed in vertebrate genomes. The number of copies of most C/D box snoRNA genes proved to be lower in placental mammals compared to other vertebrates. This can be due to smaller oocytes and accordingly lower number of ribosomes in them in eutherians. The targets of snoRNAs encoded by single-copy and multiple-copy genes proved to have different distribution in rRNAs. The causes of this difference are discussed. In some cases, the transcripts of homologous C/D box RNA genes were shown to guide the modification of neighboring nucleotides in rRNA. C/D box snoRNA pseudogenes were found in all vertebrate classes. Three novel C/D box snoRNAs were found in *Xenopus tropicalis* that may guide 2′-O-methylation of *Xenopus*-specific rRNA sites. A list of 922 annotated C/D box snoRNA genes is presented.

Introduction

Ribosomal RNAs contain a large number of modified nucleotides, mainly pseudouridine and 2′-O-methylated ribose: about 100 modifications of each type are known in vertebrates [1]. These modifications are mediated by small nucleolar RNAs (snoRNAs) in complex with proteins. Based on conserved sequence elements snoRNAs are divided into two classes: H/ACA box snoRNAs and C/D box snoRNAs that mostly guide pseudouridylation and 2′-O-methylation, respectively [2,3]. Most new C/D box snoRNAs were described under the name of U# (e.g., U25). The recent universal nomenclature assigns SNORD# names to C/D box snoRNAs (e.g., SNORD25). C/D box snoRNAs are stably associated with four proteins: NOP58, NOP56, 15.5 kDa protein, and fibrillarin which appears to be the methylase [4]. These RNAs contain C (UGAUGA) and D (CUGA) boxes near their 5′ and 3′ ends, respectively, which are brought together by a hairpin formed by the RNA ends. The resulting structure including the C and D boxes and the terminal hairpin is called the C/D motif. This motif is the binding site for snoRNP proteins and is required for the formation, stability, and nucleolar localization of snoRNPs [5]. Most C/D box snoRNAs contain (often degenerated) copies of the C and D boxes called the C′ and D′ boxes in the central part of the molecule [6]. They also contain 9–20 nt long antisense element located upstream of the D and/or D′ box that is complementary to and can interact with the target site in rRNA. A nucleotide located in the fifth position upstream from the box D or D′ in the resulting duplex is subject to modification [3]. Modified nucleotides have uneven distribution in rRNAs: nearly all of them reside in the conserved regions including the peptidyl transferase center and intersubunit bridges [1,7].

2′-O-methylation of rRNA is required for the ribosome function probably affecting proper rRNA folding, maturation and stability [8–12]. Stabilization of RNA structure by 2′-O-methylation has been shown in the studies of modified oligoribonucleosides [13], tRNAs [14,15], and rRNAs [16] as well as by the analysis of the degree of modification of tRNAs and rRNAs in different archaebacteria species [17–19].

The snoRNA genes in vertebrates are commonly encoded in introns of other genes called host genes [20]. SnoRNAs are processed from the debranched introns by exonucleases. A minor pathway exist for some snoRNAs that are processed by endonucleolytic cleavage [4]. Most host genes code for proteins, although several snoRNA genes are in introns of non-protein-coding host genes [21–23]. One of such examples is SNORD87 RNA and its host gene *U87HG* (*SNHG6*) previously described by us [24,25]. Human SNORD87 RNA is encoded by a single-copy gene in the intron of *U87HG*. Here we demonstrate that the number of *SNORD87* genes is lower in placental mammals compared to other vertebrates, and that this trend is universal for most C/D box snoRNAs. The targets of snoRNAs encoded by single-copy and multiple-copy genes show a bias toward different

* Corresponding author. Fax: +7 495 135 1405.
E-mail address: kramerov@eimb.ru (D.A. Kramerov).

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.02.003

localization in rRNAs. Three novel C/D box snoRNAs were identified in clawed frog *Xenopus tropicalis* and some other vertebrate species not belonging to placental mammals. These snoRNAs can guide species-specific modifications as they were mapped in the frog but not found in human.

Results

Number of copies of C/D box snoRNA genes is lower in placental mammals than in other vertebrates

We searched for *SNORD87* and *U87HG* genes in vertebrate genome databases and found that SNORD87 RNA is encoded in homologous loci (Table S1). For all studied species except gray short-tailed opossum, platypus, and green anole lizard these loci were available in the EST databases, which allowed us to reconstruct *U87HG* transcripts (Fig. 1, Table S2). U87HG RNAs contain no long open reading frames and are poorly conserved, as was previously reported for mammalian *U87HG* transcripts [25]. Thus, SNORD87 RNA is encoded in introns of non-protein-coding gene *U87HG* in all vertebrate species studied.

In placental mammals, the gene structure of *SNORD87* and *U87HG* remains invariant: *U87HG* consists of four exons, and the second intron contains the *SNORD87* gene (Fig. 1). On the contrary, the number of *SNORD87* genes was higher in other studied vertebrate species (except lizard). Moreover, the copy number of *SNORD87* genes increased together with the number of *U87HG* exons in all animals where the U87HG RNA sequence was reconstructed (except chicken).

All identified *SNORD87* gene copies were located in *U87HG* introns, and their sequence contained elements typical of C/D box snoRNAs: the antisense element; C, D, D′, and C′ boxes; and terminal repeats (Fig. 1, Table S3). This indicates that all of them are likely to produce functional snoRNAs.

We tested whether the revealed trend towards low copy numbers of *SNORD87* genes in placental mammals is typical of other C/D box snoRNAs by searching their human homologs in the genomes of some vertebrates. Apart from the human genome, the sequences were screened in the genomes of mouse (*Mus musculus*), frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*), and sometimes of dog (*Canis familiaris*), cow (*Bos taurus*), horse (*Equus caballus*), rat (*Rattus norvegicus*), opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), lizard (*Anolis carolinensis*), fugu (*Takifugu rubripes*), and tetraodon (*Tetraodon nigroviridis*).

We have identified five new copies of four human C/D box snoRNA genes (*SNORD19*, two copies; *SNORD52*, *SNORD63*, and *SNORD70*, single copy each; Table S3). In addition, human SNORD123, SNORD125, and SNORD126 described by Yang et al. [26] as snoRNAs with unidentified target RNA have the conserved antisense elements complementary to rRNA fragments in all vertebrate species studied (Fig. S1).These snoRNAs may guide 2′-O-methylation of three previously unmapped sites in rRNA (T-2031 in 28S rRNA, C-1440 in 18S rRNA, and A-1310 in 28S rRNA). Similar to other 2′-O-methylated nucleotides, the presumable targets of these snoRNAs reside in conserved rRNA regions (Figs. S2A,B).

The copy number of snoRNA genes proved to be higher in one or several VEPM (*v*ertebrates *e*xcept *p*lacental *m*ammals) genomes than in placental mammals in 51 out of 93 SNORD RNAs that guide rRNA modifications (Table S4). In 6 cases increased copy number of the genes in VEPM was accompanied by increased copy number of the genes in one (five cases) or two (one case) placental mammals. Overall, the higher numbers of gene copies in VEPM as compared to placental mammals, were observed for 61% snoRNAs. For 26 snoRNAs (28%), the number of gene copies was the same in all studied vertebrate species. In 10 cases (11%), the number of gene copies varied both in placental mammals and VEPM; however, the copy number in at least one VEPM species was higher than in placental mammals in a half of these cases. All snoRNA genes were localized in host gene introns and contained functional elements typical of C/D box snoRNAs (Table S3). The only exceptions are chicken *SNORD100B* and *SNORD102B* genes located in the 3′ UTR of the host gene that probably does not impair their production [27,28] and human *SNORD14E*, mouse *SNORD53A* and frog *SNORD127* genes that contained degenerated box D (Table S3). Thus, there is a trend towards high copy number of C/D box snoRNA genes in VEPM relative to placental mammals (or towards low copy number of C/D box snoRNA genes in placental mammals relative to VEPM).

VEPM demonstrate a significant difference in the copy number of snoRNA genes between species even within the same class (e.g., fishes or non-placental mammals). Conversely, in placental mammals the

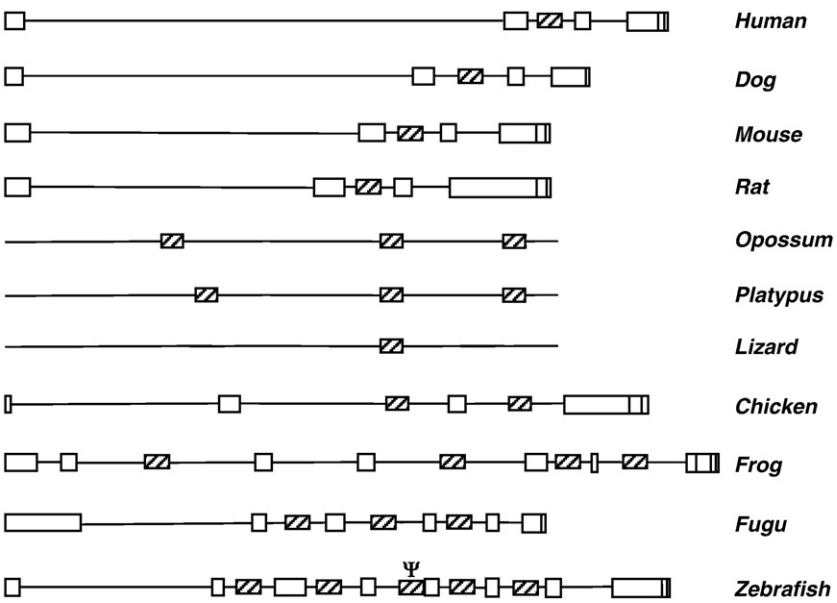

Fig. 1. Structure of *U87HG* gene in different vertebrate species. Exons of *U87HG* are indicated by empty rectangles; *SNORD87* gene, by shaded rectangles. *SNORD87* pseudogene is indicated by Ψ. Polyadenylation signals are marked with vertical lines. ESTs corresponding to *U87HG* transcripts are not yet available for opossum, platypus, and lizard.

number of genes was different only in 12 out of 93 cases, and 6 of these 12 cases belong to rodents (Table S4). Sometimes when rRNA sequences flanking the 2′-O-methylated nucleotide varied among vertebrate species, the complementary substitutions in the snoRNA antisense elements were observed (Table S3).

The copy number of snoRNA genes increased together with the number of exons in non-protein-coding host genes (Fig. 2A, Fig. S3A), which was particularly pronounced for *U87HG* (Fig. 1) and *U50HG* (*SNHG5*) (Fig. S4). It is of interest that the distribution pattern of snoRNA genes in introns varied in VEPM, but not in mammals. No trend to variation in the number of exons was observed in protein-coding host genes; however, the copy number of snoRNA genes and their arrangement in introns could vary between species (Fig. 2B, Figs. S3B,C). In addition, some snoRNA genes could be found in different host genes in different species, which was observed for both protein-coding and non-protein-coding host genes.

We analyzed the regions flanking multiple-copy snoRNA genes in non-protein-coding host genes. Sometimes (Fig. S5), the regions with high sequence similarity extended beyond snoRNA genes; although the similarity region was limited to snoRNA gene sequences in most cases.

Targets of C/D box snoRNAs encoded by multiple-copy genes are concentrated in certain regions of rRNAs

About 30% of C/D box snoRNA genes in placental mammals (30 out of 93 human genes) involved in rRNA modification are represented by 2–3 or more (up to 5) copies. We observed clustering of targets of these snoRNAs in some regions of rRNAs. (Table 1, Fig. S2). For instance, in 18S rRNA all modifications guided by C/D box snoRNAs encoded by multiple-copy genes are localized to the 5′ and central domains. These domains form the bulk of the body and platform of the small ribosomal subunit, respectively [29,30]. Most targets of C/D box snoRNAs encoded by multiple-copy genes and localized in the 5′ domain (7 out of 9, Fig. S2A) seem to be located close in the three-dimensional structure of the ribosome and reside in the bottom of the body which contains less proteins compared to other parts of the small subunit [29–32].

All C/D box snoRNAs, whose targets are localized to 5.8S rRNA and domains I and III of 28S rRNA are encoded by single-copy genes (Table 1). These domains and 5.8S rRNA as well as domain II (excluding several helices) mainly form the back of the large subunit [30,33,34]. Most targets of C/D box snoRNAs that guide 28S rRNA modifications and are encoded by multiple-copy genes reside in domains II, IV and especially in domain V (Table 1). Domains IV and V are adjacent and form the bulk of the surface contacting the small subunit. This surface is largely free from proteins [30,33,34]. Domains IV and V also form the A, P, and E sites of the large subunit; in addition, domain IV forms most of intersubunit contacts, and domain V which is most abundant in multiple-copy gene targets contains the peptidyl transferase center. It is of interest that in domain II 3 out of 5 targets of snoRNAs encoded by multiple-copy genes reside in helices that protrude to the protein-poor subunit interface side of the particle (Fig. S2B, [33]).

In VEPM interpretation of the distribution of snoRNAs targets in rRNAs is more complicated due to increasing the copy number of the most snoRNA genes and incomplete genomic sequences of the majority of species. For instance, snoRNA host genes in VEPM often contain gaps making determination of the exact number of snoRNA gene copies impossible. However, the tendency of clustering of targets of snoRNAs encoded by multiple-copy genes in the 5′ and central domains of 18S rRNA and in the II, IV, and V domains of 28S rRNA remains roughly the same (Table S5).

24 C/D box snoRNAs are encoded by single-copy genes in all studied vertebrate species (Table S4). Interestingly, none of them have targets in domain V of 28S rRNA (Fig. S2C).

Introns of UHG and gas5 genes code for three novel C/D box snoRNAs

Genes of three new C/D box snoRNAs were identified in introns of non-protein-coding host genes *UHG* (*SNHG1*) and *GAS5* in *X. tropicalis* and some other VEPM (Fig. 2A, Fig. S3A). We designated them as 'non-eutherian-specific' (NET1, NET2, and NET3) RNAs. Their size and 5′ end were identified in *X. tropicalis* using PAGE at the nucleotide resolution and primer extension, respectively (Fig. 3). Analysis of their presumptive antisense elements suggests that they guide the 2′-O-methylation of U-252 in 18S rRNA and G-3524 and C-4004 in 28S rRNA, Fig. 4 (numbering according to the GenBank X02995 sequence). Similar to other snoRNA targets, these nucleotides lie in the conserved rRNA regions (Figs. S2A,C). The methylation of the latter two nucleotides was mapped in *Xenopus* but not in human [1]. The methylation of the former nucleotide was not mapped neither in human nor *Xenopus*; however, the *Xenopus* 18S rRNA contains the only nucleotide (U), the methylation of which was demonstrated although it was not mapped [1]. The NET1 snoRNA is likely to guide the methylation of this particular nucleotide.

Homologs of NET1 and NET3 RNAs have been found in other VEPM but neither in human nor in other placental mammals, which could be expected since the modifications guided by these snoRNAs have not been mapped in human. All homologs had boxes C and D and antisense elements, as well as terminal repeats (Fig. 4). No homologs of *X. tropicalis* NET2 RNA were found in other species. This can be attributed to incomplete sequencing of the vertebrate genomes; otherwise, this modification can be restricted to *Xenopus*.

Transcripts of homologous snoRNA genes in vertebrates can modify neighboring sites in rRNA

We have found that antisense elements of some copies of C/D box snoRNA genes vary in length, which allows neighboring rRNA sites to be modified, since the nucleotide at position 5 in the duplex *upstream* of box D or D′ is subject to modification [3]. As the antisense element length changes, a different nucleotide appears at position 5 in rRNA.

For instance, chicken snoRNA host gene encoding ribosomal protein RPL21 contains an extra copy of the *SNORD102* gene, *SNORD102B*, that has a two-nucleotide-longer antisense element, so that the encoded snoRNA can guide 2′-O-methylation of a different nucleotide, C-4018, in 28S rRNA, Fig. 5 (hereafter, the numbering is given according to GenBank sequences U13369 (28S rRNA) and X03205 (18S rRNA)). Interestingly, *SNORD102B* gene is localized in the 3′ UTR of *RPL21* gene, that probably does not impair their processing (see Discussion).

Each of the two copies of a frog gene corresponding to EST *BX757900* has an intron containing snoRNA gene (*SNORD34B* and *SNORD34C*), highly similar to the frog gene *SNORD34A*. However, their antisense elements are shorter by two nucleotides, enabling them to guide 2′-O-methylation of the neighboring nucleotide T-2826 in 28S rRNA (Fig. 5).

Fish SNORD68 RNA is an interesting example of modification potential of a guide RNA. This RNA has two antisense elements and is encoded by two genes in fugu, tetraodon, medaka (*Oryzias latipes*), and stickleback (*Gasterosteus aculeatus*) and by three genes in zebrafish. Due to a different length of the second antisense element in *SNORD68A*, *SNORD68B*, and *SNORD68C* in zebrafish and SNORD68B in other fish species, the transcripts of these genes can guide the modification of A-2388 in 28S rRNA in zebrafish (*SNORD68B*); A-2390 in zebrafish, fugu, and stickleback; A-2391 in tetraodon; and G-2392 in medaka (Fig. 5).

Human SNORD45 RNA contains two antisense elements, and its gene is represented by three copies in the genome. The first antisense element in *SNORD45C* gene is shorter than in other copies, so its transcript can guide 2′-O-methylation of C-174 in 18S rRNA. This modification has been mapped in human; however, the snoRNA

directing it was not known. It seems plausible that SNORD45C RNA guides the modification of C-174 in 18S rRNA (Fig. 5).

C/D box snoRNAs in all vertebrate classes have pseudogenes

Searching for C/D box snoRNA genes in genomes of vertebrates allowed us to identify their numerous pseudogenes (Fig. S6). In human and other mammals, most such pseudogenes demonstrate features typical of retroelements: (1) a long variable A-rich tail at the 3′ end; (2) an A-rich sequence often preceding the pseudogene; and (3) short (6–20 bp) direct repeats called target site duplications (TSDs) flanking many pseudogenes (reviewed in Kramerov and Vassetzky [35]). The identified pseudogenes often contain mutations in the antisense elements and lack terminal stems required for snoRNA formation [5]. They can be truncated at either end and can be located in intergenic regions rather than in introns of host genes, which makes their processing improbable.

Pseudogenes of C/D box snoRNAs have also been found in the genomes of other vertebrates (platypus, chicken, lizard, frog, and zebrafish), (Fig. S6). These pseudogenes commonly have TSDs but lack A-rich tails, which agrees with the absence of LINE1 in their genomes, while retroelements are replicated in these species using other LINEs that require no A-rich tails [36].

Two annotated genes encoding human C/D box snoRNAs likely are nonfunctional retropseudogenes

The *SNORD32* gene is represented by two copies in the human genome, *SNORD32A* [37] and *SNORD32B* [38]. The *SNORD32A* gene is localized in the second intron of ribosomal protein gene *RPL13A*. The sequence of *SNORD32B* proved to be a part of *RPL13A* pseudogene (Fig. 6A). Six spliced exons of *RPL13A* and a short A-rich tail are located downstream of the intron containing the *SNORD32B* gene. This pseudogene is flanked with TSD (Fig. S7A), is not found in introns of annotated genes or ESTs, and is present in primates only. The *SNORD32B* sequence is not flanked with inverted terminal repeats that are required for snoRNA processing [39]. All these features indicate that the *SNORD32B* sequence is a fragment of a primate-specific retropseudogene and cannot give rise to snoRNA.

Two copies of the *SNORD96* gene have been found in the human genome: *SNORD96A* and *SNORD96B* [40]. The *SNORD96B* gene proved to be specific for primates, is flanked with TSD and contains an A-rich sequence at the 3′ end (Fig. 6B, Fig. S7B). It is located 8.5 kb upstream from the 3′ end of the fourth intron of the *AMMERC1* gene, whereas most known C/D box snoRNAs are at a distance of no more than 100 bp from the 3′ end of introns [41]. The proximity to the 3′ end of introns is a prerequisite for successful processing of C/D box snoRNAs [41]. These features suggest that the *SNORD96B* sequence is a primate-specific nonfunctional retropseudogene of the *SNORD96A* gene.

Discussion

A substantial fraction of C/D box snoRNAs (~30% in human and other studied placental mammals) are encoded by genes represented by several copies in the genome. The conservation of functional elements among the homologs of all studied species (Table S3) points to the stabilizing selection that maintains multiple-copy genes encoding snoRNAs. This could be attributed to low expression of the

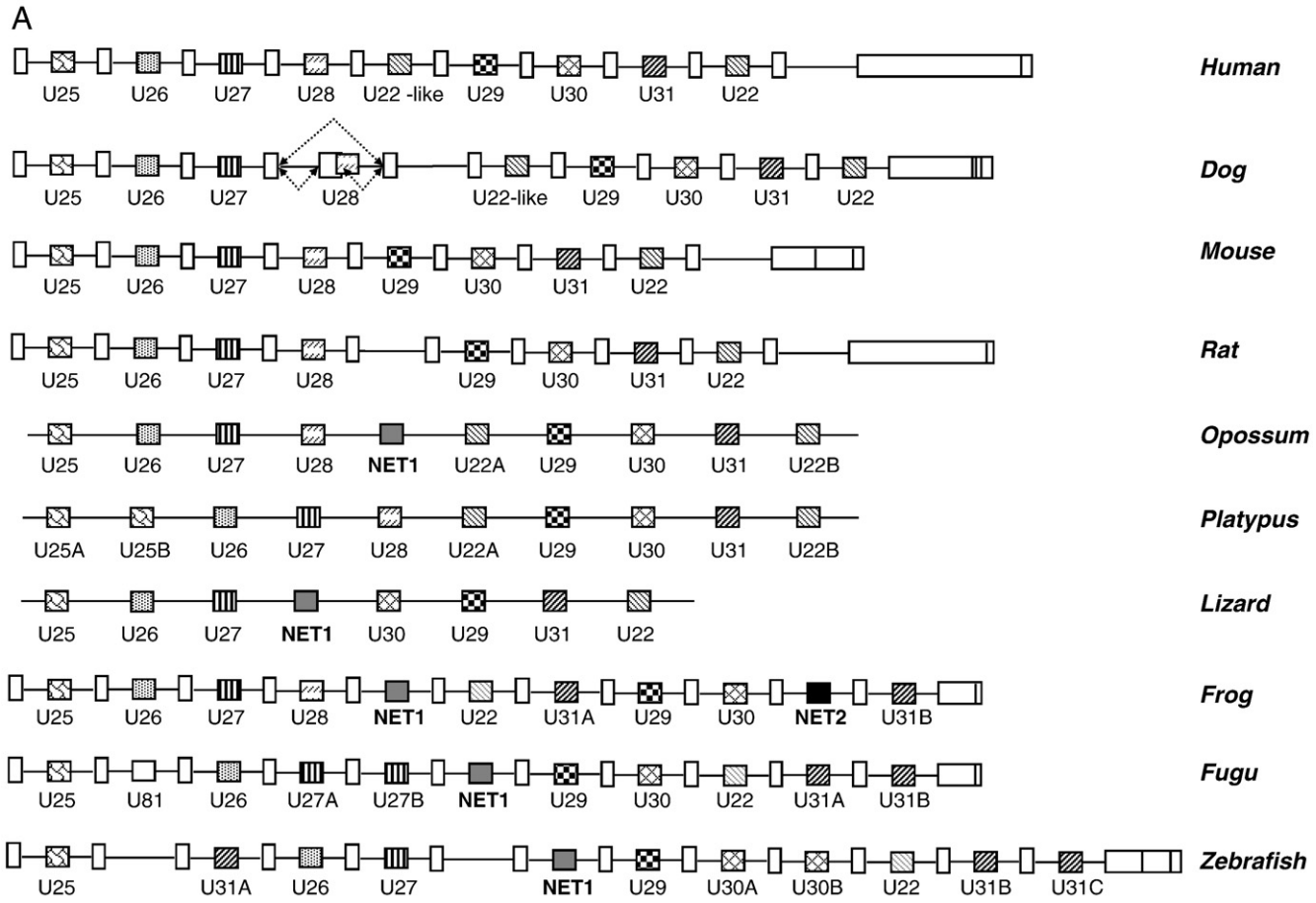

Fig. 2. Structure of *UHG* (*SNHG1*) (A) and *RPS8* (B) genes in different vertebrate species. SnoRNA genes are shown as rectangles with different shading. Arrows point to alternative splicing. RPS8 exons in opossum and lizard were identified by alignment with human RPS8 mRNA. SnoRNA genes are designated according to the old nomenclature for presentation convenience. Frog *U46B* gene resides in a different locus. For other designations, see Fig. 1.

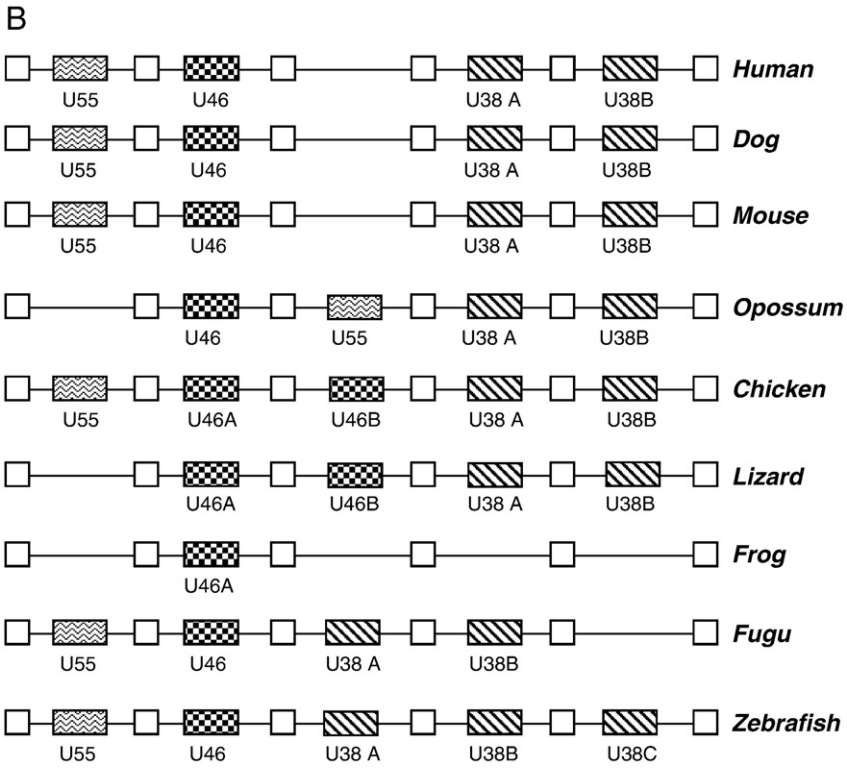

Fig. 2 (*continued*).

host genes; however, both multiple-copy and single-copy genes can be found within the same host gene.

The presence of several snoRNA gene copies can provide for a more complete methylation of the target. Some nucleotides in rRNA are not completely methylated at 2′-OH, i.e., a fraction of rRNA molecules have unmethylated targets of some snoRNAs [1,19]. Different methylation levels were observed for ~20% of modified nucleotides in trypanosomes at different stages of their life cycle [42]. Although this issue was not specifically addressed for human rRNAs, a few partially methylated nucleotides were identified using not very sensitive assays [1]. It is of interest that their modification is guided by snoRNAs encoded by single-copy genes in the human genome. More sensitive assays can shed light on this issue.

The copy number of genes of most C/D box snoRNAs is lower in placental mammals than in VEPM, where ~60% of C/D box snoRNA genes are represented by two or more copies. Moreover, the number of most "multi-copy" human genes is 2–3, while it can reach 4–6 in VEPM (Table S4). This can be credited to the specific features of VEPM reproduction: many ribosomes are needed in large oocytes, which requires the modification of numerous rRNAs. Interestingly, the trend to increased number of snoRNA gene copies is also observed in platypus, which lays eggs and has large oocytes [43] but is less pronounced in opossum (Marsupialia) (Table S4).

2′-O-methylation can stabilize RNA tertiary structure [15,16,18,19]. Accordingly, another possibility is that the high copy number of snoRNA genes in VEPM can mediate a more complete methylation of rRNAs and, hence, the structural stabilization of a higher number of ribosomes at all life cycle stages, which is particularly important for poikilotherms such as most VEPM. Platypus, although not a poikilothermic animal, belongs to monotremes, which are known to have a lower (~32 °C) and more variable body temperature than therian mammals [44]. In this context, it is of interest that plants that depend on environmental changes more than animals have the greatest number of rRNA methylation sites, and most of their snoRNA genes are multiple-copy [45,46].

The targets of C/D box snoRNAs encoded by single-copy and multiple-copy genes have different localization in rRNAs: the latter are concentrated in the 5′ domain of 18S rRNA and in domains IV and V of 28S rRNA as well as in the helices of domain II protruding to the surface contacting the small ribosomal subunit (Fig. S2). This distribution correlates with the overall distribution of 2′-O-methylated nucleotides as well as with the distribution of pseudouridine [32]. As modifications are known to stabilize the tertiary structure and correct folding of rRNA it seems plausible to assume that in these regions such stabilization is particularly important. rRNA fragments enriched with the targets of multi-copy snoRNA genes are largely from the bottom of the body of the small subunit and the surface of the large subunit interacting with a small subunit and containing the peptidyl transferase center. These areas are known to contain fewer proteins than other parts of the ribosome [29–31,33]. One of the main functions of ribosomal proteins is the stabilization of rRNA structure [33]. Accordingly, in the protein-poor context the role of modifications in these domains may be of primary importance and the increased number of snoRNAs performing them would be necessary.

All snoRNA genes found in the present study are intron-encoded. The only exception is chicken *SNORD100B* and *SNORD102B* genes that are located in the 3′-UTR of their host genes. To the best of our knowledge, this is the first example of such snoRNA gene arrangement in vertebrates. In *Drosophila* and yeast some snoRNAs are also encoded by the exons of the host genes and are faithfully processed [27,28]. Their production is apparently independent of the splicing of the mRNA of the host gene and is probably carried out by endonucleases [27]. In vertebrates, the existence of such endonucleolytic pathway was demonstrated for some intron-encoded snoRNAs [47]. This pathway is probably involved in the processing of chicken SNORD100B and SNORD102B exon-encoded snoRNAs.

Although the systematic analysis of C/D box snoRNA pseudogenes in vertebrates was not our goal, we have found many such pseudogenes in mammals and, to a lesser extent, in other vertebrate classes. Previously, H/ACA box snoRNA pseudogenes have been found,

Table 1
Number of modifications in different rRNA domains guided by human C/D box snoRNAs encoded by multiple-copy and single-copy genes.

rRNA domains	Number of sites of 2′-O-methylation			
	Guided by snoRNAs encoded by multiple-copy genes[a]	Guided by snoRNAs encoded by single-copy genes	Unidentified snoRNAs	Total
18S rRNA				
5′ domain	9	11	1	21
Central domain	4	1	1	6
3′ major domain	–	10	2	12
3′ minor domain	–	2	–	2
28S rRNA				
I domain	–	2	–	2
II domain	5	9	1	15
III domain	–	8	1	9
IV domain	5	9	–	14
V domain	10	8	3	21
VI domain	2	3	–	5
5.8S rRNA				
	–	2	–	2

[a] This group includes targets whose methylation is guided by two different snoRNAs (SNORD77 and SNORD80; SNORD32 and SNORD51) as well as targets of snoRNAs containing two antisense elements (SNORD36, SNORD45, and SNORD50).

and a model of new snoRNA gene origin by retroposition has been proposed [40,48,49]. Among C/D box snoRNAs, pseudogenes have been found for U3 and U13 RNAs [50–52]. These RNAs differ from other C/D box snoRNAs in that they are encoded by independent genes rather than reside in introns of host genes and in that they are required to cut pre-rRNA rather than modify it. A number of C/D box snoRNA pseudogenes have been annotated in the human genome but were not discussed (RNA genes Track at UCSC Human Genome Browser; http://genome.ucsc.edu). In this work, the first C/D box snoRNA pseudogenes were described in different vertebrate species and new pseudogenes were found in human. The pseudogenes localize nearly exclusively outside of genes or within a gene, but in the opposite strand. This pattern differs from that described for H/ACA box snoRNA pseudogenes, nearly half of which are localized in introns [48]. Most C/D box snoRNAs pseudogenes have substitutions in the conserved sequence elements, which suggests that they are subject of selection to a lesser extent compared to the functional copies. This coupled with the localization of most pseudogenes outside of introns of host genes and suggests that they are not functional. This distinguishes them from H/ACA box snoRNA retrogenes, some of which can be functional [48,49]. Moreover, *SNORD32B* and *SNORD96B* previously described as putatively functional copies of C/D box snoRNA genes [38,40] appear to be retrogenes, most likely nonfunctional. Thus, we have found no data indicating that copies of C/D box snoRNA genes can be generated by retroposition. In addition, with few exceptions, all identified copies of C/D box snoRNA genes were encoded in introns of the same host gene. In the rare cases when snoRNA gene copies were located in introns of two different host genes (for instance, there are only two such cases in human, Table S3), retrogenes features have never been found for them. The predominant localization of homologous genes of C/D box snoRNAs in introns of the same host gene has been recently reported in platypus [53]. Interestingly, the situation was different in nematodes *C. elegans* and *C. brigssae,* where such pattern was observed for H/ACA but not for C/D box snoRNAs [54].

The NET1–NET3 RNAs identified in *X. tropicalis* represent snoRNAs guiding taxon-specific modifications in rRNAs. The extent to which such modifications have spread among vertebrates have never been studied in detail. This can be partially attributed to insufficient sensitivity and accuracy of assays for 2′-O-methylated nucleotides. Analysis of genome databases is an alternative approach to this problem, and the identification of taxon-specific NET snoRNAs illustrates the potential of this approach.

Materials and methods

Search for nucleotide sequences of snoRNA genes in genome databases

Homologs of C/D box snoRNA genes in the genomes of human and other vertebrates were searched using the WU-BLAST 2.0 algorithm with modified search parameters at Ensembl (http://www.ensembl.org/Multi/blastview). High sensitivity, *W* (word size for seeding alignments) = 3, and *Q* (cost of first gap character) = 1 were set. At the first step, C/D box snoRNA genes were searched in the genomes of vertebrates using the nucleotide sequences of human snoRNA genes from the snoRNA-LBME-db as a query [38]. Then, the identified snoRNA sequences of vertebrates were used as a query to search the genomes of the same species. The found sequences with intact C and D/D′ boxes and the antisense element as well as some sequences with the similarity of at least 70% (to search for possible pseudogenes) were mapped in the genomes of vertebrates using the BLAT algorithm (http://genome.ucsc.edu). In order to find out if the identified sequences lie in introns of host genes, mRNA and EST databases were used or, in the rare cases when no transcripts were available, the loci contained identified sequences were aligned with mRNAs of the putative host gene from other vertebrate species. Sequences containing C and D/D′ boxes and the antisense element, flanked with short inverted repeats, and lying within introns of a host gene were considered as snoRNA genes, and extra copies were searched in the other introns of the same host gene. Several host genes, in particular, non-protein-coding ones are not very conserved, which makes (in the absence of ESTs) the identification of exons and introns based on the alignment with transcripts of other species impossible. In such cases, the found sequences were considered as snoRNA genes based on their localization in a locus homologous to a human locus with a similar set of snoRNA genes.

The homologs of *X. tropicalis NET1–NET3* genes were searched in the genomes of vertebrates using the WU-BLAST 2.0 algorithm as described above and in some cases using FASTA3 [55]. In the human genome, these genes were additionally searched using the fuzznuc program of the EMBOSS package (mismatch = 1 and complement = yes) [56]. The following search patterns were used: NET1 RNA, pattern = N(10)TGATGAN(30,60)TCACCAAAGCNCTGAN(10); NET2 RNA, pattern = N(10)TGATGAN(30,60)CGTCGCTATNCTGAN(10); NET3 RNA, pattern = N(10)TGATGAN(1,8)AA[GT]CAGGTCNCTGAN(20,30)CTGAN(10).

Pairwise and multiple alignments were generated using the ClustalV [57] and ClustalW [58] algorithms.

The nucleotide sequences of *X. tropicalis* C/D box snoRNAs NET1, NET2, and NET3 were deposited to GenBank under accession numbers FJ460491, FJ460492, and FJ460490, respectively.

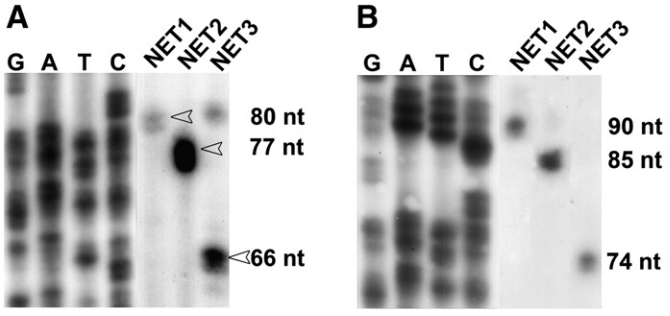

Fig. 3. Primer extension (A) and size determination (B) of NET snoRNAs in *Xenopus tropicalis*. (A) Lanes NET1, NET2, and NET3, products of primer extension with total *X. tropicalis* RNA and NET1rev, NET2rev, and NET3rev primers, respectively. Numbers indicate size of the reaction products marked by arrowheads. The longer product of reaction with NET3rev likely corresponds to NET3 RNA precursor. (B) Northern hybridization of total *X. tropicalis* RNA (10 μg) with NET RNA-specific probes. Numbers indicate the length of NET RNAs. pSL1190 sequencing products were used as a size marker (lanes G, A, T, and C).

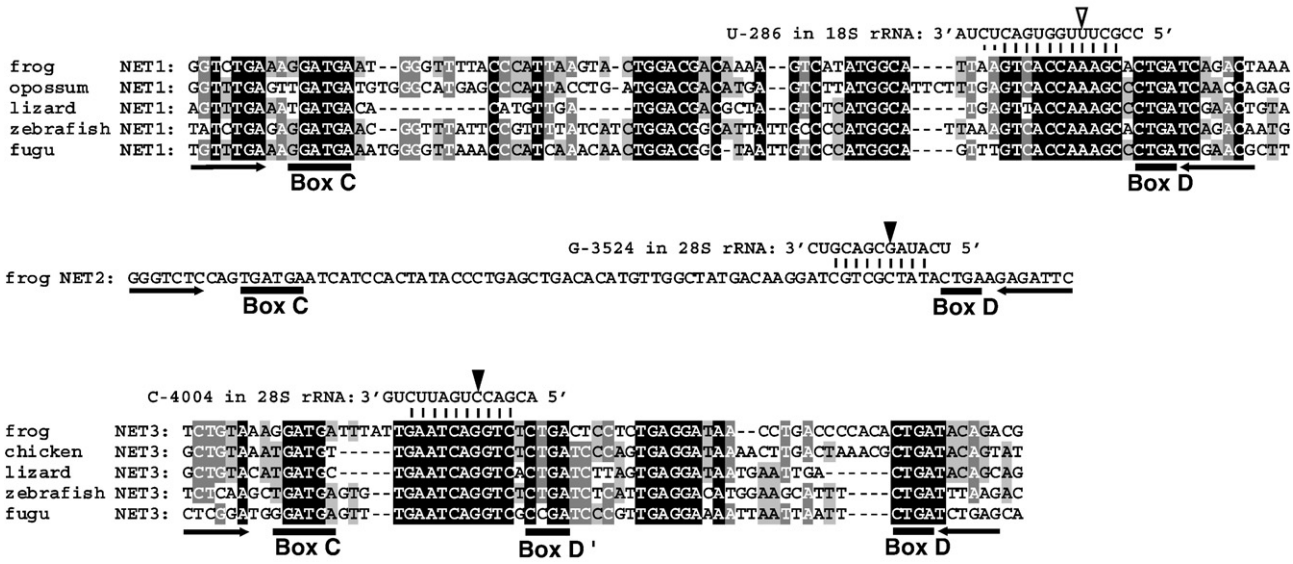

Fig. 4. Novel snoRNAs NET1–NET3 encoded by introns of *UHG* and *gas5* genes. The nucleotide sequences of *NET1–NET3* genes for the indicated species are given and the complementary regions of their antisense elements and rRNAs are indicated. The terminal repeats are indicated by arrows (three 3′-terminal nucleotides of NET1 snoRNA and two 3′-terminal nucleotides of NET3 snoRNA in most species are involved in complementary interactions with the nucleotides upstream of the snoRNA genes; see Table S3). Boxes C, D, and D′ are underlined. The nucleotides in rRNAs where the predicted 2′-O-methylation were and were not mapped are marked with solid and empty triangles, respectively. The numbering corresponds to frog 18S and 28S rRNAs in GenBank sequence X02995.

DNA and RNA isolation

DNA was isolated from the muscle of *X. tropicalis* by incubation with proteinase K followed by phenol/chloroform extraction. Total RNA was isolated from the muscle of *X. tropicalis* using the guanidine isothiocyanate method [59].

Northern analysis

Total RNA (10 μg) was separated by electrophoresis in 6% polyacrylamide gel plates (360 mm long and 1 mm thick) with 7 M urea and transferred to a Hybond-N membrane by semidry electroblotting. Hybridization was performed by incubating the filters

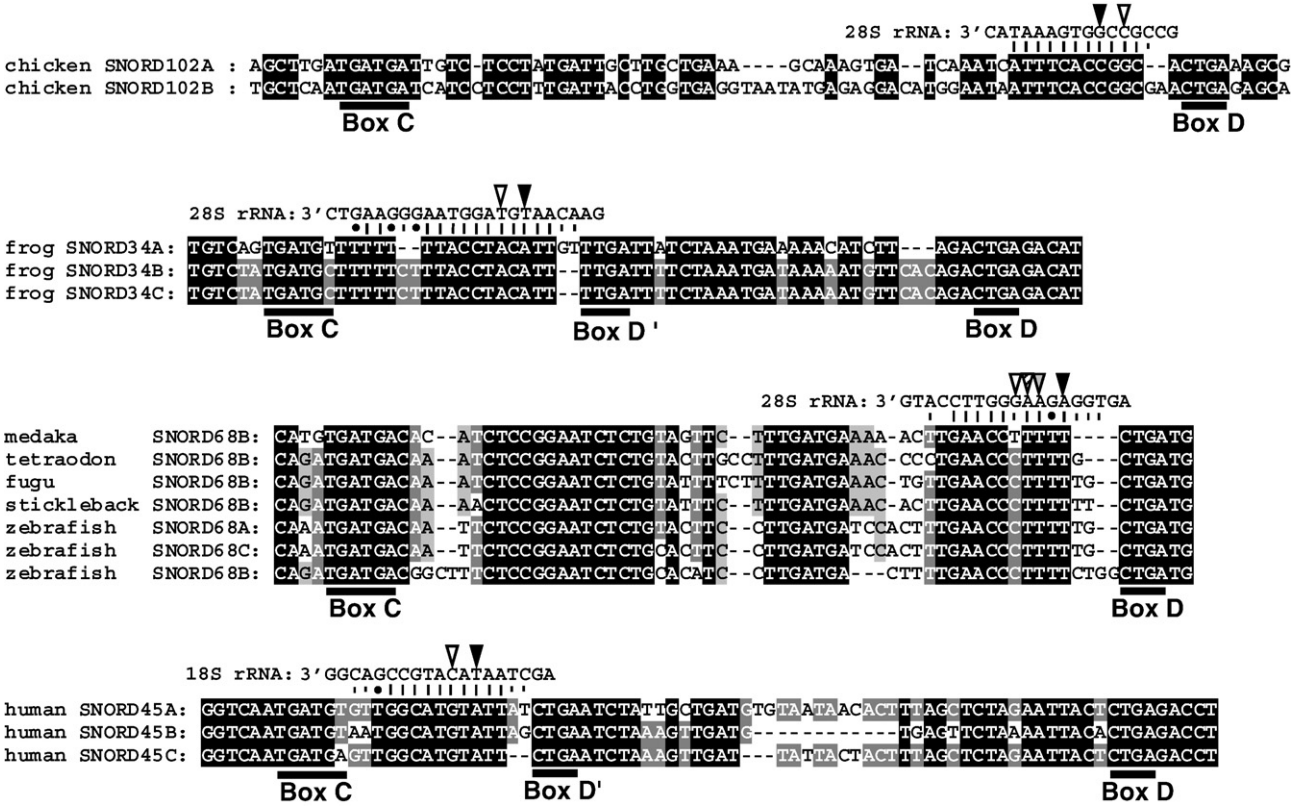

Fig. 5. Alignment of nucleotide sequences of homologous C/D box snoRNA genes. The nucleotides in rRNAs that can be 2′-O-methylated are indicated by triangles; solid triangles point to nucleotides whose modification is guided by transcripts of the homologs described previously, while other triangles correspond to the modifications directed by the homologs described in this work. The G–T complementarity is marked with solid circles. The boundaries of the C/D box RNA genes are given by analogy with the human genes.

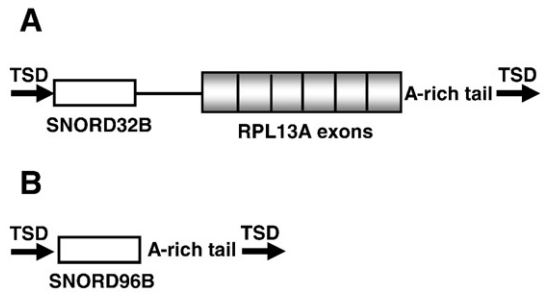

Fig. 6. The structure of *RPL13A* and *SNORD96B* pseudogenes. Target site duplications (TSD) are indicated by arrows.

overnight at 42 °C in 4× SSC, 1% SDS, 50% formamide, 5×Denhardt's solution, and 0.1 mg/ml denaturated herring sperm DNA with ^{32}P-labeled probe. Then filters were washed at 42 °C for 40 min in 0.1× SSC, 0.1% SDS and were exposed to film with an intensifying screen. The probes were generated by PCR amplification of *X. tropicalis* genomic DNA using primers NET1dir (GGTCTGAAAGGATGAATG)/NET1rev (TCAGTGCTTTGGTGACTT), NET2dir (GGTCTCCAGTGATGAATC)/NET2rev (TCAGTATAGCGACGATCC), or NET3dir (GTAAAGGATGATTTATTG)/NET3rev (TCAGTGTGGGGTCAGGTT). The PCR products were purified by electrophoresis in 4% agarose gel (3% NuSieve/1% SeaKem) and ~1% of the sample was labeled by PCR with [α-^{32}P]dATP (25 μCi) using the 'rev' primers.

The products of pSL1190 sequencing with the M13 forward (−40) primer (GTTTTCCCAGTCACGAC) was used as a size marker in RNA PAGE and primer extension analysis.

Primer extension analysis

Primers NET1rev, NET2rev, and NET3rev were 5' end labeled with [γ-^{32}P] ATP (3000 Ci/mmol) and T4 polynucleotide kinase. Primer (0.6 pmol) was mixed with 15 μg of total RNA. Primer was annealed in 25 μl of 10 mM Tris–HCl (pH 7.5), 1 mM EDTA, and 0.3 M NaCl for 1 h at 42 °C. cDNA was synthesized in 70 mM Tris–HCl (pH 8.3), 17 mM $(NH_4)_2SO_4$, and 7.5 mM $MgCl_2$, with 0.5 mM dNTPs and 200 U of M-MLV reverse transcriptase for 1 h at 42 °C. The resulting samples were analyzed by electrophoresis in 6% polyacrylamide gel with 7 M urea.

Acknowledgments

The authors thank Dr. Nikita Vassetzky for help with English version of the manuscript. This research was supported by the Russian Foundation for Basic Research (grant 08-04-00350) and the Molecular and Cellular Biology Program of the Russian Academy of Sciences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.02.003.

References

[1] B.E. Maden, The numerous modified nucleotides in eukaryotic ribosomal RNA, Prog. Nucleic Acid Res. Mol. Biol. 39 (1990) 241–303.
[2] P. Ganot, M.L. Bortolin, T. Kiss, Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs, Cell 89 (1997) 799–809.
[3] Z. Kiss-Laszlo, Y. Henry, J.P. Bachellerie, M. Caizergues-Ferrer, T. Kiss, Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs, Cell 85 (1996) 1077–1088.
[4] W. Filipowicz, V. Pogacic, Biogenesis of small nucleolar ribonucleoproteins, Curr. Opin. Cell. Biol. 14 (2002) 319–327.
[5] D.A. Samarsky, M.J. Fournier, R.H. Singer, E. Bertrand, The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization, EMBO J. 17 (1998) 3747–3757.
[6] Z. Kiss-Laszlo, Y. Henry, T. Kiss, Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA, EMBO J. 17 (1998) 797–807.
[7] W.A. Decatur, M.J. Fournier, rRNA modifications and ribosome function, Trends Biochem. Sci. 27 (2002) 344–351.
[8] J. Esguerra, J. Warringer, A. Blomberg, Functional importance of individual rRNA 2'-O-ribose methylations revealed by high-resolution phenotyping, RNA 14 (2008) 649–656.
[9] J.L. Baxter-Roshek, A.N. Petrov, J.D. Dinman, Optimization of ribosome structure and function by rRNA base modification, PLoS ONE 2 (2007) e174.
[10] S.C. Blanchard, J.D. Puglisi, Solution structure of the A loop of 23S ribosomal RNA, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 3720–3725.
[11] B. Liu, X.H. Liang, D. Piekna-Przybylska, Q. Liu, M.J. Fournier, Mis-targeted methylation in rRNA can severely impair ribosome synthesis and activity, RNA Biol. 5 (2008).
[12] C.S. Chow, T.N. Lamichhane, S.K. Mahto, Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications, ACS Chem. Biol. 2 (2007) 610–619.
[13] L.L. Cummins, et al., Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity, Nucleic Acids Res. 23 (1995) 2019–2024.
[14] G. Kawai, et al., Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group, Biochemistry 31 (1992) 1040–1046.
[15] G. Kawai, et al., Relation between functions and conformational characteristics of modified nucleosides found in tRNAs, Nucleic Acids Symp. Ser. (1991) 49–50.
[16] X. Song, R.N. Nazar, Modification of rRNA as a 'quality control mechanism' in ribosome biogenesis, FEBS Lett. 523 (2002) 182–186.
[17] I. Kumagai, K. Watanabe, T. Oshima, Thermally induced biosynthesis of 2′-O-methylguanosine in tRNA from an extreme thermophile, *Thermus thermophilus* HB27, Proc. Natl. Acad. Sci. U. S. A. 77 (1980) 1922–1926.
[18] P.P. Dennis, A. Omer, T. Lowe, A guided tour: small RNA function in Archaea, Mol. Microbiol. 40 (2001) 509–519.
[19] K.R. Noon, E. Bruenger, J.A. McCloskey, Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*, J. Bacteriol. 180 (1998) 2883–2888.
[20] E.S. Maxwell, M.J. Fournier, The small nucleolar RNAs, Annu. Rev. Biochem. 64 (1995) 897–934.
[21] K.T. Tycowski, M.D. Shu, J.A. Steitz, A mammalian gene with introns instead of exons generating stable RNA products, Nature 379 (1996) 464–466.
[22] C.M. Smith, J.A. Steitz, Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes, Mol. Cell. Biol. 18 (1998) 6897–6909.
[23] R. Tanaka, et al., Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human B-cell lymphoma, Genes Cells 5 (2000) 277–287.
[24] I.K. Gogolevskaya, et al., U87 RNA, a novel C/D box small nucleolar RNA from mammalian cells, Gene 292 (2002) 199–204.
[25] J.A. Makarova, D.A. Kramerov, Noncoding RNA of U87 host gene is associated with ribosomes and is relatively resistant to nonsense-mediated decay, Gene 363 (2005) 51–60.
[26] J.H. Yang, et al., snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome, Nucleic Acids Res. 34 (2006) 5112–5123.
[27] D. Filippini, F. Renzi, I. Bozzoni, E. Caffarelli, U86, a novel snoRNA with an unprecedented gene organization in yeast, Biochem. Biophys. Res. Commun. 288 (2001) 16–21.
[28] M.C. Accardo, et al., A computational search for box C/D snoRNA genes in the *Drosophila melanogaster* genome, Bioinformatics 20 (2004) 3293–3301.
[29] B.T. Wimberly, et al., Structure of the 30S ribosomal subunit, Nature 407 (2000) 327–339.
[30] P. Chandramouli, et al., Structure of the mammalian 80S ribosome at 8.7 A resolution, Structure 16 (2008) 535–548.
[31] F. Schluenzen, et al., Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution, Cell 102 (2000) 615–623.
[32] D. Piekna-Przybylska, W.A. Decatur, M.J. Fournier, The 3D rRNA modification maps database: with interactive tools for ribosome analysis, Nucleic Acids Res. 36 (2008) D178–D183.
[33] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, Science 289 (2000) 905–920.
[34] M.M. Yusupov, et al., Crystal structure of the ribosome at 5.5 A resolution, Science 292 (2001) 883–896.
[35] D.A. Kramerov, N.S. Vassetzky, Short retroposons in eukaryotic genomes, Int. Rev. Cytol. 247 (2005) 165–221.
[36] M. Kajikawa, N. Okada, LINEs mobilize SINEs in the eel through a shared 3′ sequence, Cell 111 (2002) 433–444.
[37] M. Nicoloso, L.H. Qu, B. Michot, J.P. Bachellerie, Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2′-O-ribose methylation of rRNAs, J. Mol. Biol. 260 (1996) 178–195.
[38] L. Lestrade, M.J. Weber, snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs, Nucleic Acids Res. 34 (2006) D158–D162.
[39] J. Cavaille, J.P. Bachellerie, Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure, Biochimie 78 (1996) 443–456.

[40] P. Vitali, et al., Identification of 13 novel human modification guide RNAs, Nucleic Acids Res. 31 (2003) 6543–6551.
[41] T. Hirose, J.A. Steitz, Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 12914–12919.
[42] S. Barth, et al., Elucidating the role of C/D snoRNA in rRNA processing and modification in *Trypanosoma brucei*, Eukaryotic Cell 7 (2008) 86–101.
[43] R.R. Behringer, G.S. Eakin, M.B. Renfree, Mammalian diversity: gametes, embryos and reproduction, Reprod. Fertil. Dev. 18 (2006) 99–107.
[44] G. Grigg, L. Baeard, T. Grant, M. Augee, Body temperature and diurnal activity patterns in the platypus (*Ornithorhynchus anatinus*) during winter, Aust. J. Zool. 40 (1992) 135–142.
[45] J.W. Brown, M. Echeverria, L.H. Qu, Plant snoRNAs: functional evolution and new modes of gene expression, Trends Plant. Sci. 8 (2003) 42–49.
[46] C.L. Chen, et al., The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*, Nucleic Acids Res. 31 (2003) 2601–2613.
[47] E. Caffarelli, M. Arese, B. Santoro, P. Fragapane, I. Bozzoni, In vitro study of processing of the intron-encoded U16 small nucleolar RNA in *Xenopus laevis*, Mol. Cell. Biol. 14 (1994) 2966–2974.
[48] Y. Luo, S. Li, Genome-wide analyses of retrogenes derived from the human box H/ACA snoRNAs, Nucleic Acids Res. 35 (2007) 559–571.
[49] M.J. Weber, Mammalian small nucleolar RNAs are mobile genetic elements, PLoS Genet. 2 (2006) e205.
[50] R.A. Denison, S.W. Van Arsdell, L.B. Bernstein, A.M. Weiner, Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome, Proc. Natl. Acad. Sci. U. S. A. 78 (1981) 810–814.
[51] R. Reddy, et al., Isolation and characterization of three rat U3 RNA pseudogenes colinear with U3 RNA, J. Biol. Chem. 260 (1985) 5715–5719.
[52] S.J. Baserga, X.D. Yang, J.A. Steitz, Three pseudogenes for human U13 snRNA belong to class III, Gene 107 (1991) 347–348.
[53] J. Schmitz, et al., Retroposed SNOfall—a mammalian-wide comparison of platypus snoRNAs, Genome Res. 18 (2008) 1005–1010.
[54] A. Zemann, A. op de Bekke, M. Kiefmann, J. Brosius, J. Schmitz, Evolution of small nucleolar RNAs in nematodes, Nucleic Acids Res. 34 (2006) 2676–2685.
[55] W.R. Pearson, D.J. Lipman, Improved tools for biological sequence comparison, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 2444–2448.
[56] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open Software Suite, Trends Genet. 16 (2000) 276–277.
[57] D.G. Higgins, A.J. Bleasby, R. Fuchs, CLUSTAL V: improved software for multiple sequence alignment, Comput. Appl. Biosci. 8 (1992) 189–191.
[58] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[59] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction, Anal. Biochem. 162 (1987) 156–159.
[60] J.J. Cannone, et al., The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, BMC Bioinformatics 3 (2002) 2.
[61] R. Tanaka-Fujita, Y. Soeno, H. Satoh, Y. Nakamura, S. Mori, Human and mouse protein-noncoding snoRNA host genes with dissimilar nucleotide sequences show chromosomal synteny, RNA 13 (2007) 811–816.