Available online at www.sciencedirect.com
ScienceDirect
Genomics 89 (2007) 490–501
GENOMICS
www.elsevier.com/locate/ygeno

A combined computational and experimental analysis of two families of snoRNA genes from *Caenorhabditis elegans*, revealing the expression and evolution pattern of snoRNAs in nematodes

Zhan-Peng Huang [1], Chong-Jian Chen [1], Hui Zhou, Bei-Bei Li, Liang-Hu Qu *

Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory for Biocontrol, Zhongshan University, Guangzhou, 510275, China

Received 13 June 2006; accepted 2 December 2006
Available online 11 January 2007

Abstract

Small nucleolar RNAs (snoRNAs) are an abundant group of noncoding RNAs mainly involved in the posttranscriptional modifications of rRNAs in eukaryotes. Prior to this study, only 28 snoRNA genes had been identified from *Caenorhabditis elegans*, indicating that most snoRNA genes are hidden in the worm genome, which represents a simple multicellular metazoan. In this study, a genome-wide analysis of the two major families of snoRNA genes in *C. elegans* was performed using the snoscan and snoGPS programs incorporating comparative genome analyses. Seventy gene variants, including 36 box C/D and 34 box H/ACA snoRNA genes, were identified, of which 50 are novel. Two families of snoRNAs showed a characteristic genomic organization. Notably, 6 box C/D snoRNA genes were located in the antisense orientation of introns. In contrast to insect and mammal, the distances between many intronic snoRNAs and 3′ splice sites of introns were less than 50 nt in the worm, an unexpected finding as intron-encoded snoRNAs in *C. elegans* are supposed to be expressed in a splicing-dependent pathway. Interestingly, a canonical H/ACA snoRNA, ΨCeU5-48, was revealed to be partially homologous to small Cajal body-specific RNA (scaRNA) U85 and U89 in fly and human, indicating a possible evolutionary relationship between snoRNAs and scaRNAs.
© 2006 Elsevier Inc. All rights reserved.

Keywords: snoRNA; ncRNA; Intron; RNA modification; *Caenorhabditis elegans*

Introduction

The small nucleolar RNAs (snoRNAs) define one of the largest families of small noncoding RNAs (ncRNAs) known in eukaryotes. With the exception of RNase MRP, all the snoRNAs fall into two major families, box C/D and box H/ACA snoRNAs, on the basis of common sequence motifs and structural features. A large number of snoRNAs characterized to date are box C/D snoRNAs that share two conserved motifs, the 5′-end box C and the 3′-end box D, whereas the box H/ACA snoRNAs exhibit a common hairpin–hinge–hairpin–tail secondary structure with the box H and ACA [1]. Several snoRNAs, such as U3, snR30, and RNase MRP, are required for specific cleavage of pre-rRNAs. However, the majority of box C/D snoRNAs function as guides for site-specific 2′-*O*-ribose methylation with most box H/ACA snoRNAs functioning as guides for pseudouridylation in the posttranscriptional processing of rRNAs [2]. Studies have shown that some snoRNAs and scaRNAs participate in the modifications of snRNAs [3,4]. Some modifications in Archaea tRNAs are introduced by box C/D small RNAs, which are the homologues of snoRNAs in eukaryotes [5]. Moreover, an increasing number of orphan snoRNAs with unknown function have been identified from different eukaryotes, suggesting that they play additional roles in cellular processes.

There has been a significant increase in research on snoRNAs in the past five years. Hundreds of snoRNAs have been identified from different eukaryotes using the approach of experimental RNomics [6]. Recently, improved techniques have been developed to increase the specificity of snoRNA gene identification from a cDNA library [7,8]. However, such

* Corresponding author.
E-mail address: lsbrc04@zsu.edu.cn (L.-H. Qu).
[1] These authors contributed equally to this work.

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.12.002

experimental approaches tend to favor the discovery of the most abundant RNAs so that species with low abundance may not be detected. For this reason, computational approaches based on pattern recognition scan algorithms have been developed, such as Snoscan for box C/D snoRNA [9], SnoGPS [10], and the MFE method [11] for box H/ACA snoRNA. Using the computer scanning programs, several studies have provided an unbiased genome-wide search for snoRNA genes in a genome of given sequence [9,10,12]. However, along with experimentally confirmed snoRNA genes, these results also included hundreds of false-positive candidates that need to be excluded from the search.

Recently, genome-wide analyses of snoRNAs have been carried out in several model organisms whose genome sequences were available. With the large amount of data available, different snoRNA gene organizations and expression patterns have been revealed [13–15]. Prior to this study, there has been limited research on snoRNAs in *Caenorhabditis elegans*, with the identification of only 20 types of C/D and 8 types of H/ACA snoRNAs [16–19]. Hence further studies are required on the snoRNA genes and rRNA posttranscriptional modification processes in this model organism, which represents a simple multicellular metazoan. In this study, a genome-wide computational analysis of two classes of snoRNA genes in *C. elegans* is presented for the first time. To avoid a large amount of false-positive candidates, we used the snoscan and snoGPS programs in conjunction with conserved noncoding sequence analysis to scan the genome of *C. elegans* for snoRNA genes. Incorporating our experimental results, this study identified a total of 50 novel snoRNA genes and delineated a characteristic pattern of snoRNA expression and evolution in a nematode.

Results

Computational identification of 50 novel snoRNA genes from C. elegans by snoscan and snoGPS in conjunction with conservation filters

To test the applicability of the snoscan [9] and snoGPS [10] programs for identifying *C. elegans* snoRNAs, we applied the search programs to the 28 known snoRNA sequences. Only two snoRNAs, rpl-7a and rpl-7aps, failed to pass the test due to their atypical H/ACA snoRNA structure. These results therefore support the applicability of the programs to detect snoRNA genes within the *C. elegans* genome.

In the computational search of the conserved noncoding genome sequences between *C. elegans* and *C. briggsae*, more than 30,000 snoRNA candidate hits were found. We then applied the CD_filter.pl and ACA_filter.pl programs, which focus on the conservation of box elements (box C, D or box H, ACA) and functional elements (antisense sequences to rRNAs or snRNAs) in alignments between *C. elegans* and *C. briggsae*, to identify perfectly conserved candidates. In detail, the parameters used in the study were as follows: (1) conserved "ANANNN" sequence present in box H alignments; (2) conserved "A[CTA]A" sequence present in box ACA alignments; (3) more than four of the seven nucleotides in box C being conserved; (4) all four nucleotides in box D being conserved; (5) each separately conserved antisense prediction of H/ACA snoRNA being not less than 3 nt and not less than 8 nt in total; and (6) the conserved antisense prediction of C/D snoRNA being not less than 8 nt. As a result, less than 3000 hits including approximately 300 C/D and 2600 H/ACA hits were extracted. An improved mfold program, which was integrated with a preliminary coarse mountain plots filter [11], was applied to further select approximately 400 of the 2600 H/ACA hits according to the typical hairpin–hinge–hairpin–tail secondary structure. These 300 C/D and 400 H/ACA hits were then scanned for snoRNA genes for the final candidates. In this step, we examined whether the predicted functional antisenses were symmetrically present in the pocket of the hairpin of H/ACA candidates and whether the functional antisenses were viable (forming at least eight continuous base parings with target RNA without a bulge) in the C/D candidates. In total, 70 snoRNA genes, including 36 C/D and 34 H/ACA genes, were identified from both *C. elegans* and *C. briggsae* (Tables 1 and 2 and Supplementary Table S1). In addition to the 20 known snoRNA genes reported previously [17,18], 50 novel nematode snoRNA genes were identified in this study. Eight known snoRNA genes were missed in this search due to their location in the opposite strand of exons, nonconservation in the genome alignment or atypical snoRNA structure (see Discussion).

The sequences of the major snoRNA genes were highly conserved between the two nematodes and the functional antisenses were nearly identical (Tables 1 and 2). The variations in conserved box H/ACA snoRNA were primarily located in the loop and stem region of the hairpin. Notably, the majority of the variation in the stem region appeared to be covariation, which guarantees the stable hairpin–hinge–hairpin–tail secondary structure.

According to the base-pairing interactions between the identified snoRNAs and target RNAs (Supplemetary Figs. S1 and S2), the snoRNAs identified in this study were predicted to guide 73 methylations and pseudouridylations on rRNAs and snRNAs (34 2′-*O*-methylations and 39 pseudouridylations). Most of these RNA modifications are phylogenetically conserved in yeast, plants, and/or metazoa (Tables 1 and 2). All the box C/D snoRNAs from *C. elegans* in this study possess only one antisense sequence, which is predicted to target a single methylation site in substrate RNA. In contrast, one-third of the box H/ACA snoRNAs exhibit two functional elements. Interestingly, some hairpins show the potential of directing two nonadjacent pseudouridylations by a single guide sequence, which was recently reported in *Saccharomyces cerevisiae* [10] (Fig 1).

Experimental evaluation of the computational results

Experimental confirmation of those computationally identified snoRNA genes that were not cloned from the cDNA libraries (see below) was performed by Northern blotting analysis. All the snoRNAs were positively detected by Northern blotting with the labeled probes (Fig. 2). In most cases, unique

Table 1
Box C/D snoRNA genes in *C. elegans*

SnoRNA name	Len (nt)	Seq (%)	Modification	Antisense element	Homology			Location
					Yeast	Plants	Metazoans	
MeCeSSU-A90▲	67	92.5	SSU-A90	20nt (5′); 100%	snR51		U57	IR
MeCeSSU-C400▲	83	90.4	SSU-C400	13nt (3′); 100%	U14		Me18S-C419/U14	Y37E3.11
MeCeSSU-A422*▲	74	79.7	SSU-A422	13nt (5′); 100%		snoR15	U16	F54C9.1
MeCeSSU-A601*▲	73	91.8	SSU-A601	15nt (3′); 100%	snR47	U36	U36	Intron antisense
MeCeSSU-C980▲	70	95.7	SSU-C980	12nt (5′); 100%		snoR20	Me18S-C1096	C08D8.1
MeCeSSU-C1180●▲	78	87.2	SSU-C1180	13nt (5′); 100%		Z277	Z67	H06I04.4a
MeCeSSU-C1196●	80	93.8	SSU-C1196	14nt (5′); 100%		snoR14	Z61	IR
MeCeSSU-U1342●	72	91.7	SSU-U1342	12nt (3′); 100%		U61	U61	IR
MeCeLSU-G668▲	66	93.9	LSU-G668	12nt (5′); 100%			Me28S-G764/U21	Y71D11A.3a
MeCeLSU-A678▲	66	98.5	LSU-A678	11nt (5′); 100%	U18	U18	Me28S-A774/U18	IR
MeCeLSU-G860▲	74	83.8	LSU-G860	16nt (5′); 93.8%	snR39b	snoR39BY	Me28S-G980/snR39b	IR
MeCeLSU-A1185*▲	69	92.8	LSU-A1185	11nt (5′); 100%	snR61	U38	Me28S-A1322/U38	B0334.2
MeCeLSU-C1502*▲	86	96.5	LSU-C1502	13nt (5′); 100%	U24	U24	U24	IR
MeCeLSU-U1566*	71	93.0	LSU-U1566	13nt (5′); 100%				F38E11.12
MeCeLSU-C2300●▲	72	94.4	LSU-C2300	12nt (5′); 100%	snR76		Z6/Z80	R12E2.3
MeCeLSU-A2317●▲	76	76.3	LSU-A2317	17nt (5′); 88.2%		U37	U37	M106.1
MeCeLSU-G2343▲	79	86.1	LSU-G2343	14nt (5′); 100%			U87	Intron antisense
MeCeLSU-A2384a*▲	66	95.5	LSU-A2384	12nt (3′); 100%	snR13		Me28S-A2589/U15	IR
MeCeLSU-A2384b▲	66	98.5	LSU-A2384	12nt (3′); 100%	snR13		Me28S-A2589/U15	Intron antisense
MeCeLSU-A2384c●▲	110	85.5	LSU-A2384	11nt (3′); 100%	snR13		Me28S-A2589/U15	F56C9.1
MeCeLSU-C2407●	78	91.0	LSU-C2407	14nt (5′); 100%				Intron antisense
MeCeLSU-U2417▲	123	91.1	LSU-U2417	10nt (5′); 100%			Z64	IR
MeCeLSU-A2429▲	76	90.8	LSU-A2429	11nt (5′); 100%		snoR44	Me28S-A2634/U79	IR
MeCeLSU-C2440●▲	72	84.7	LSU-C2440	10nt (5′); 100%	snR64	snoR44	Me28S-C2645/U74	IR
MeCeLSU-U2762*	82	96.3	LSU-U2762	12nt (5′); 100%		snoR10		IR
MeCeLSU-U2841●▲	73	94.5	LSU-U2841	11nt (3′); 100%	snR51		U41	IR
MeCeLSU-G2903▲	92	96.7	LSU-G2903	15nt (5′); 100%	snR48	snoR1	Me28S-G3253	F55A11.6
MeCeLSU-A3023▲	74	87.8	LSU-A3023	11nt (5′); 100%		snoR31		Intron antisense
MeCeLSU-C3060●	66	97.0	LSU-C3060	11nt (5′); 100%	snR69	snoR69Y		W04D2.3
MeCeLSU-C3071●▲	77	93.5	LSU-C3071	14nt (5′); 100%	snR73	U35	Me28S-C3420/U35	C51F7.1
MeCeLSU-A3159*▲	76	88.2	LSU-A3159	14nt (5′); 100%				T06D8.3
MeCeU2-G12▲	106	90.6	U2-G12	11nt (5′); 100%			MBII-382	D1054.3
MeCeU2-C42●▲	79	74.7	U2-C42	12nt (3′); 100%			MBII-19	C07H6.5
MeCeU2-U49●	146	72.6	U2-U49	9nt (3′); 100%				Y62F5A.1
MeCeU6-A48*	85	80	U6-A48	11nt (5′); 100%			MgU6-53	Intron antisense
MeCeU6-C55●▲	89	77.5	U6-C55	12nt (3′); 100%			MBII-166	Y116A8C.35

Note. All the identified snoRNA genes were named after their functions. The genes marked with asterisks, circles, and triangles indicate that the genes are experimentally detected in the cDNA libraries, by Northern blotting analysis, and identified recently by Deng et al. [20], respectively. "Len": length of the snoRNA gene; "Seq": sequence conservation of snoRNA genes in *C. elegans* and *C. briggsae*. Column "Antisense element" gives the functional antisenses and their conservation between the two nematodes. In the column "Location," the protein-coding host genes are denoted by their names. "IR" represents intergenic region. "Intron antisense" represents the opposite strand of the intron.

and obvious bands were revealed under stringent conditions of hybridization and the sizes of these snoRNAs were consistent with their predicted sizes. In particular, two strong bands were revealed in the Northern blotting analysis of MeCeU6-C55, which possesses a long terminal stem. The two bands probably represent different sizes of the mature snoRNA. Similar observations have been reported in other species [6,25]. On the other hand, the different intensities of the Northern blotting bands, which were obtained under the same hybridization conditions, indicated a broad range in the expression levels of different snoRNAs in *vivo*. In general, some low-abundance snoRNAs such as ΨCeSSU-499 and MeCeU2-U49 are difficult to detect in a cDNA library. These snoRNAs were not identified in our cDNA library and another recently reported cDNA library [20], although over 100 snoRNAs were identified from the two cDNA libraries (see below). However, they can still be identified by our computational scan. The experimental analyses of Northern blotting and cDNA cloning (see below) demonstrated that the 70 snoRNA genes, identified by computational analysis in this study, were authentic genes.

Under the stringent filtering processes, many snoRNA genes might have been missed in the computational scan. To evaluate the coverage percentage of our computational results in the complete set of snoRNA genes in *C. elegans*, two cDNA libraries were constructed and screened for snoRNAs from two families More than 100 clones were randomly selected and sequenced from each cDNA library. Along with cDNAs of rRNAs, tRNAs and snRNAs, 62 clones representing 33 snoRNAs were found in the two cDNA libraries (Table 3). Although some box C, D elements were atypical, these C/D snoRNAs were also recognized by perfect flanking inverted sequences. As a result, only 14 snoRNAs were covered by the

Table 2
Box H/ACA snoRNA genes in *C. elegans*

SnoRNA name	Len (nt)	Seq (%)	Modification	Antisense element	Homology			Location
					Yeast	Plants	Metazoans	
ΨCeSSU-499●▲	128	77.3	SSU-499	5+5 nt (5′); 90%			Ψ18S-525	E01A2.6
			SSU-1082	8+4 nt (3′); 100%		SnoR72		
ΨCeSSU-513a▲	124	79.8	SSU-513	7+3 nt (3′); 100%				B0412.4
			LSU-2977	6+6 nt (3′); 100%	SnR46		Ψ28S-3327/ACA16	
ΨCeSSU-513b*▲	127	74.8	SSU-513	6+4 nt (3′); 90%				W08G11.3
			LSU-2977	6+5 nt (3′); 91.7%	SnR46		Ψ28S-3327/ACA16	
ΨCeSSU-970●▲	126	91.3	SSU-970	6+4 nt (3′); 100%	SnR31	SnoR72	Ψ18S-1086/ACA8	C05D2.6
			LSU-2847	4+6 nt (3′); 100%	SnR189		ACA2	
ΨCeSSU-971▲	123	89.4	SSU-971	6+3 nt (3′); 100%				ZK546.13
ΨCeSSU-1146●▲	137	77.4	SSU-1146	6+4 nt (5′); 100%	SnR85		ACA5	Y43F8C.7
			LSU-940	6+9 nt (3′); 93.3%			Ψ28S-1060	
ΨCeSSU-1152▲	136	66.9	SSU-1152	6+5 nt (5′); 100%	SnR36		Ψ18S-1275/ACA36	Y71F9AM.4a
ΨCeSSU-1266a●	126	85.7	SSU-1266	7+6 nt (3′); 92.3%		SnoR88		K07C5.4
ΨCeSSU-1266b*▲	127	91.3	SSU-1266	9+6 nt (3′); 93.3%		SnoR88		K07C5.4
ΨCeSSU-1523a*▲	129	79.1	SSU-1523	4+6 nt (5′); 90.9%			ACA5	R151.3
			SSU-1156	5+5 nt (3′); 100%	SnR35		Ψ18S-1377/ACA13	
ΨCeSSU-1523b▲	130	84.6	SSU-1523	5+6 nt (5′); 100%			ACA5	R151.3
			SSU-1156	4+5 nt (3′); 100%	SnR35		Ψ18S-1377/ACA13	
ΨCeLSU-1058*	134	78.4	LSU-1058	6+7 nt (5′); 100%	SnR5	SnoR81	Ψ28S-1192/ACA52	K07C5.4
ΨCeLSU-1176▲	128	90.6	LSU-1176	6+4 nt (3′); 100%	SnR5	SnoR80	Ψ18S-1347	D1046.1
ΨCeLSU-1573●▲	137	92.7	LSU-1573	6+5 nt (5′); 100%				T24H7.2
			LSU-2040	6+3 nt (5′); 100%				
			LSU-2454	5+8 nt (3′); 92.3%	SnR82			
ΨCeLSU-1996●▲	128	83.6	LSU-1996	6+5 nt (5′); 100%			Ψ28S-2149	T08B2.9a
ΨCeLSU-2294a●▲	128	89.8	LSU-2294	7+4 nt (5′); 100%	SnR32			K11H12.2
			LSU-2417	5+6 nt (3′); 100%		SnoR83	Ψ28S-2622/ACA48	
ΨCeLSU-2294b●	131	91.6	LSU-2294	7+4 nt (5′); 100%	SnR32			K11H12.2
			LSU-2417	5+5 nt (3′); 100%		SnoR83	Ψ28S-2622/ACA48	
ΨCeLSU-2361●▲	124	87.1	LSU-2361	5+5 nt (5′); 100%	SnR191	SnoR79	Ψ28S-2566/U19	F17C11.9a
ΨCeLSU-2483▲	131	77.9	LSU-2483	7+4 nt (3′); 100%				F28C6.7a
ΨCeLSU-2519a●	128	87.5	LSU-2519	6+3 nt (5′); 100%	SnR11		ACA3	F28D1.7
			LSU-2558	6+6 nt (3′); 100%			ACA3	
ΨCeLSU-2519b▲	131	83.2	LSU-2519	6+3 nt (5′); 100%	SnR11		ACA3	R151.3
			LSU-2558	6+6 nt (3′); 100%			ACA3	
ΨCeLSU-2836▲	135	85.9	LSU-2836	6+4 nt (5′); 100%			Ψ28S-3186	IR
ΨCeLSU-2966*▲	133	94.0	LSU-2966	6+4 nt (5′); 90%			Ψ28S-3316/ACA21	F53G12.10
			LSU-3035	6+4 nt (3′); 100%	SnR10	SnoR74	Ψ28S-3385/ACA21	
ΨCeLSU-2992●▲	129	92.2	LSU-2992	8+9 nt (5′); 100%	SnR34	U65	Ψ28S-3342/U65	C10C6.6
ΨCeLSU-3220▲	130	90.8	LSU-3220	8+8 nt (5′); 100%			Ψ28S-3571	C14C10.3
			LSU-2828	8+5 nt (3′); 100%		SnoR97	ACA2	
			LSU-1270	5+4 nt (3′); 100%				
ΨCeU2-16●	133	64.7	U2-16	3+7 nt (5′); 100%				C50C3.6
			U5-36	7+3 nt (3′); 100%				
ΨCeU2-41●	141	80.3	U2-41	5+4 nt (3′); 100%			ACA26	C50C3.6
ΨCeU4-62●	146	85.6	U4-62	6+4 nt (5′); 100%				JC8.13
ΨCeU5-45▲	125	68.0	U5-45	6+7 nt (3′); 100%			ACA57	K12D12.1
ΨCeU5-48●▲	139	59.0	U5-48	7+7 nt (5′); 100%			U85/U89	B0393.3
ΨCeU6-26●	134	85.1	U6-26	6+5 nt (5′); 100%				F22D6.5
CeACAOrph1a●▲	137	93.4	Unknown	Unknown				Y37E3.8a
CeACAOrph1b▲	138	92.0	Unknown	Unknown				Y37E3.8a
CeACAOrph2▲	124	71.0	Unknown	Unknown				F25H5.3a

Note. All the identified snoRNA genes were named after their functions. The genes marked with asterisks, circles, and triangles indicate that the genes are experimentally detected in the cDNA libraries, by Northern blotting analysis and identified recently by Deng et al. [20]. "Len": length of the snoRNA gene; "Seq": sequence conservation of snoRNA genes in *C. elegans* and *C. briggsae*. Column "Antisense element" gives the functional antisenses and their conservation between the two nematodes. In the column "Location", the protein-coding host genes are denoted by their names. "IR" represents intergenic region.

computational results. Most of the snoRNAs missing in the computational analysis, which were found in the cDNA libraries, were not conserved in the genome-wide alignment of the two nematode genomes. In addition, a few snoRNAs were also missed due to their mutative elements, antisenses between two nematodes, or atypical box elements and structures. Furthermore, 3 additional novel snoRNAs, which did not overlap with our computational results or the recently reported results (20), were identified (Table 3). In summary, using both bioinformatics and cloning techniques, we have

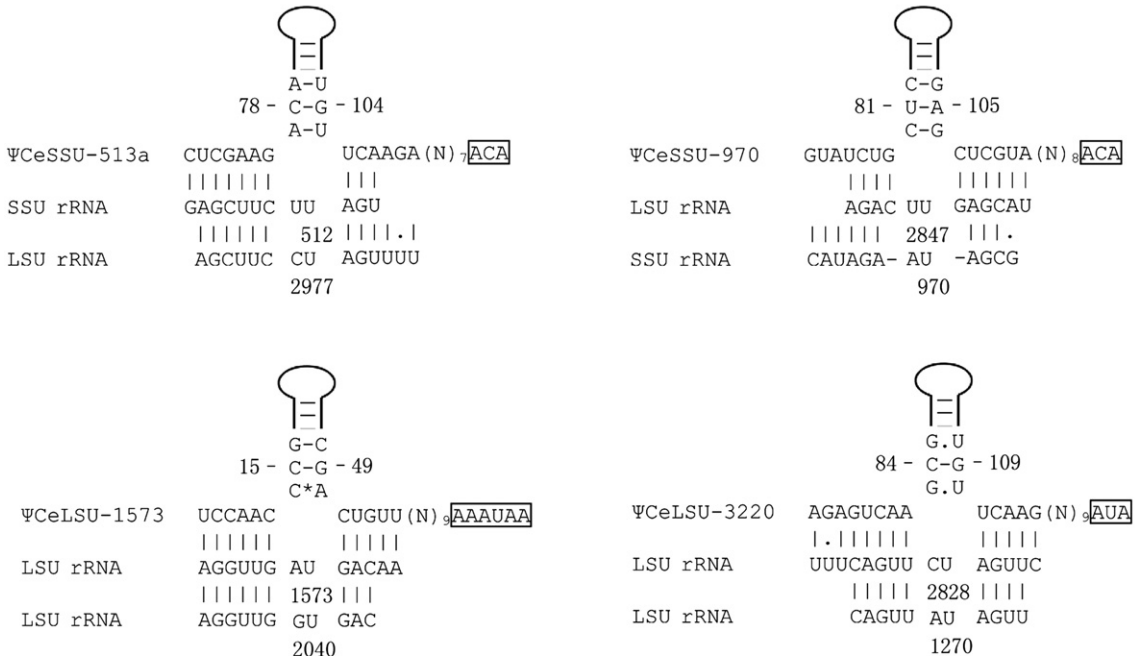

Fig. 1. Functional prediction of box H/ACA snoRNAs that guide two different pseudouridylations by a single guide sequence. Pseudouridylation guide duplexes between snoRNAs and rRNAs are predicted. The snoRNA sequences in a 5′ to 3′ orientation are shown in upper strands, whereas rRNA sequences in a 3′ to 5′ orientation are shown in lower strands. The two sequence motifs are boxed and the upper parts of the hairpins are represented by solid lines. The positions of pseudouridines are indicated by numbers. Only one isoform is shown if the snoRNA has more isoforms.

found 17 novel snoRNA genes in *C. elegans*, including 14 novel species and 3 novel isoforms, which differ from the 28 initial snoRNA genes and escaped detection in a recent study [20].

During the course of this work, another study reported the large-scale detection of small noncoding RNAs from *C. elegans* [20]. Numerous snoRNAs were identified through a cDNA library without any functional analysis. In total, 53 (76%) full-length or partial cDNA sequences matched the computationally identified snoRNAs presented in this study (Tables 1 and 2). The two individual studies, which employ different strategies, demonstrate the importance of both the bioinformatics approach and cDNA cloning in the large-scale identification of snoRNAs.

From the cDNA libraries constructed by two laboratories [20] (the laboratory of Deng and our own; Fig 3), we concluded that approximately one-half [42% (14/33) from our laboratory; 53% (53/100) from Deng´s laboratory] of the snoRNA genes in *C. elegans* were identified in our computational scan. According to this evaluation, the 2′-*O*-methylation and pseudouridylation guide RNA machinery in nematode comprises an approximate total of 150 snoRNA genes. On the other hand, although over 100 snoRNA genes were found in *C. elegans* through the traditional cloning technique, the bioinformatics programs still successfully identified 13 snoRNA genes that were missed in those cDNA libraries.

Particular positioning of intronic snoRNA genes within the host introns in C. elegans

With only one exception, all box H/ACA snoRNAs were intron-encoded in *C. elegans*. In contrast, the genomic organization of the C/D snoRNAs fell into two different categories, either intron-encoded or transcribed independently from intergenic regions. Notably, six guided snoRNA genes were located in the antisense orientation of introns. This has been seldom reported in other organisms. It is worth noting that cotranscription is an important way of regulation required for ribosome and spliceosome assembly. In this study, half of the H/ACA genes predicted to guide pseudouridylations of rRNA (12/25) in *C. elegans* were located in the introns of host genes of ribosomal proteins and proteins associated with snoRNP. Additionally, three of six genes predicted to guide pseudouridylations of snRNA localized in the introns of host genes of snRNA-associated proteins (prp-4 and prp-8). However, most of the C/D snoRNA host genes were characteristic of catalytic proteins, with none encoding ribosomal protein. The correlation between the C/D snoRNA genes and their host genes for ribosome assembly was not evident. We further scanned all the introns of ribosomal protein in *C. elegans* [21] for both box C/D and H/ACA snoRNA genes. Consistent with our point, approximately one-half of the large introns (>160 nt), which are considered large enough to accommodate an H/ACA snoRNA gene, appeared to have H/ACA snoRNA genes (with some nonperfectly conserved candidates; data not shown). As expected, no convincing C/D candidates were found in this analysis.

The distance from the intronic snoRNAs to the 3′ splice sites, which averages between 60 and 80 nt in mammals [22] and insects [14], has been proven to be important for the effective processing of the snoRNAs from their host mRNA precursor. Unexpectedly, more than half of the distances (27/51) from the intronic snoRNAs to the 3′ splice sites were less than 50 nt in *C.*

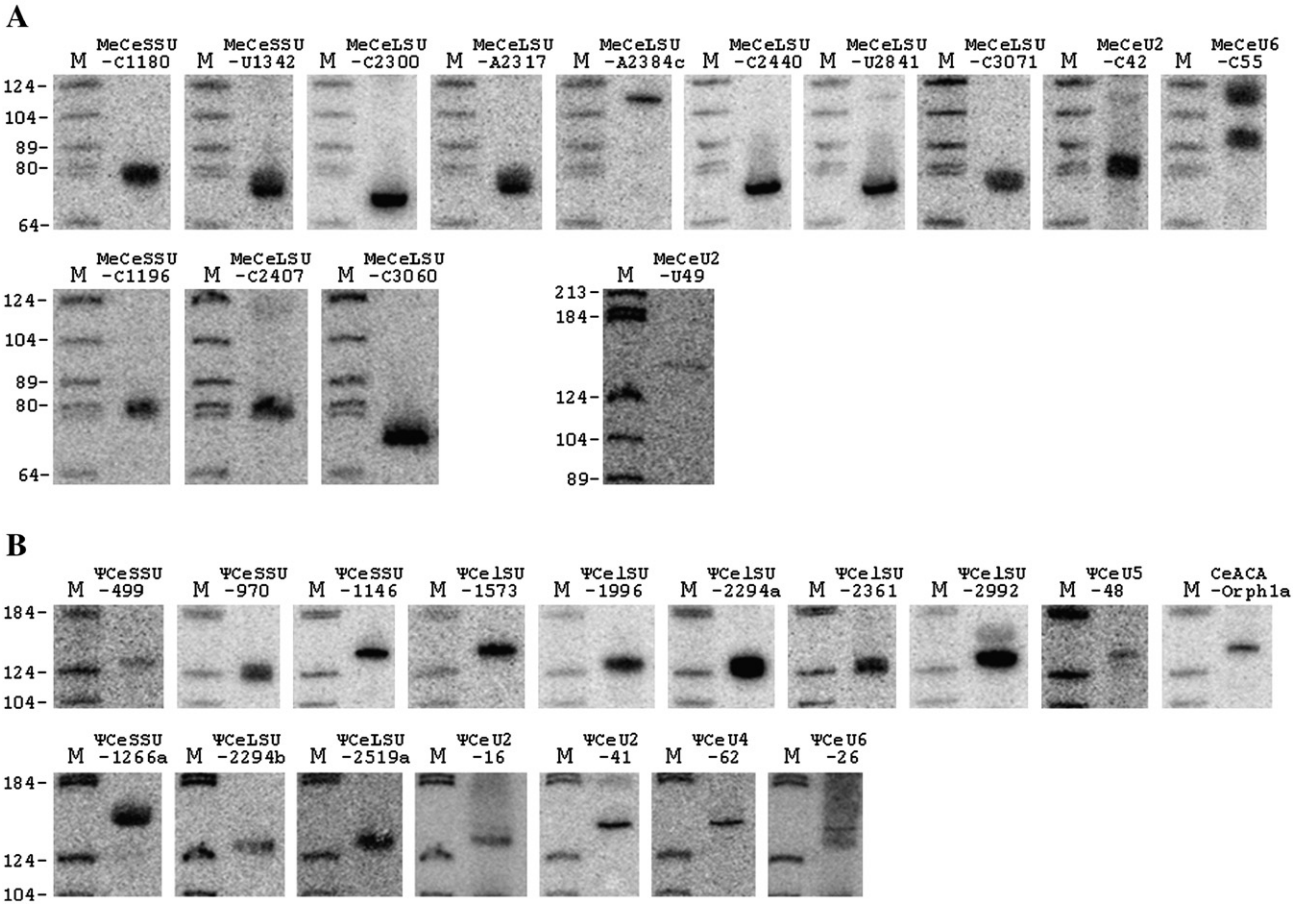

Fig. 2. Northern blot analyses. Aliquots of 40 μg total cellular RNA were separated on a denaturing 8% polyacrylamide gel and hybridized with the labeled oligonucleotide probes described in the experimental procedure section. Lane M, molecular weight markers (pBR322 digested with HaeIII and 5′-end-labeled with [γ-^{32}P]ATP). (A) Northern blot analyses of computationally identified box C/D snoRNAs. (B) Northern blot analyses of computationally identified box H/ACA snoRNAs.

elegans (Fig. 4). The distribution of the distances appeared to peak at 41–50 nt in *C. elegans*, whereas the distances from the snoRNAs to the 5′ splice sites were similar to those in insects and mammals, with a peak at 31–40 nt.

Structural and functional evolution of box H/ACA snoRNAs

Comparative functional analyses have revealed that many rRNA guide snoRNAs are evolutionarily conserved in distant eukaryotes, such as *Saccharomyces cerevisiae*, *C. elegans*, *Drosophila melanogaster*, and *Homo sapiens*. In box H/ACA snoRNAs, for example, most of the functional sequences that are complementary to rRNAs are located in the corresponding hairpins of the counterparts in different organisms (Supplementary Fig. S3A). Our study also demonstrates the extensive separation and recombination of functional hairpins occurring during the evolution of H/ACA snoRNAs. The *C. elegans* ΨCeLSU-2966 possesses two broad-conserved guiding functions. Its counterpart in human is ACA21, which possesses two homologous functions in corresponding hairpins. In contrast, two snoRNAs, Ψ28S-3316 and Ψ28S-3385, possess the two functions in *D. melanogaster*, respectively (Supplementary Fig S3A). This suggests that snoRNA ΨCeLSU-2966 evolves in two pathways, with one evolving as a dual functional snoRNA and the other separating the guiding functions and evolving as two independent snoRNAs with a single functional element. Another example is that ΨCeSSU-970 and ΨCeLSU-3220 possess a conserved guiding function for LSU-2847 and LSU-2828, respectively. In contrast, the *H. sapiens* ACA2 takes the two guiding functions for these two close nucleotide modifications (Supplementary Fig S3B).

A canonical box H/ACA snoRNA, ΨCeU5-48, was positively detected and predicted to guide pseudouridylations of U5-48. Interestingly, its homologues in *Drosophila* and humans were identified as scaRNA U85 and U89 (3), which contain both box C/D and H/ACA domains. All the functional antisenses of the homologues in the four organisms, *H. sapiens*, *D. melanogaster*, *C. elegans*, and *C. briggsae*, are present in the 5′ hairpin of the sno/scaRNA. A sequence alignment of the functional hairpin shows that only the antisenses are conserved (Fig. 5). This indicates that a canonical box H/ACA snoRNA, ΨCeU5-48-like snoRNA, may be the progenitor of U85 and U89 in the fly and human, indicating a possible evolutionary mechanism from the snoRNA to the scaRNAs. We suggest that the formation of U85 and U89 scaRNAs in the fly and human is probably achieved by a cDNA insertion from reverse

Table 3
Experimental detection of snoRNAs from cDNA libraries in *C. elegans*

Clone No.	Gene	Sequence
Box C/D snoRNA		
CD1-1-13	MeCeSSU-A422	agtgcg**atgaatg**acttggataagtttcggctgaaacttggtgatgccaacttttaaaactg**ctga**gcac
CD1-2-58	MeCeSSU-A601	aagcaa**atgacga**aatcgcacctcggcccgactccaaccctgggggcgaaatgagctttttaact**caga**tgct
CD1-1-72	MeCeLSU-A1185	gaatcg**gtgatgt**gatatccagttctgctactgagttattgtgaagattaactttccccgt**ctga**gatt
CD1-2-35	MeCeLSU-C1502	tgtccg**ttgatga**caacatacatacaccattacgatctctgaagacttcgtgctgatcatgtatccatgcaacaccaa**ctga**ggac
CD1-1-76	MeCeLSU-U1566	gggtcg**atgatta**gattattctactgttcacgtaccgactgttcagtggtggataaaccttct**atga**gact
CD1-1-68	MeCeLSU-A2384a	ccgcat**atgaaga**ccactaaatgacgaatcctaataacccaatgggtttcattgcggatatgaggcatttgt**ctga**gcgg
CD1-2-69	MeCeLSU-U2762	agagtt**gtgatga**cttacttaggacacctttggagggcctcggtctgctgaaaagaaaaatagtacgccttagt**ctga**gctc
CD1-1-53	MeCeLSU-A3159	aagccg**atgatta**ccaaaaataaccaaatgtttgagtgattgtttgtgatcgaattttgtcactatcg**ctga**ggct
CD1-1-96	MeCeU6-A48	cctgcg**atgatga**gaatcatgctaatcatgaagagactcgtcagtctccaacctccatgttgtaaaacaatcgatag**ctga**gcag
CD1-1-11	New CD1	ctggca**atgatcg**aattatcattgagccaatccttttctgaattctgtgaggatgtaaatgataggt**ctga**gcca
CD1-1-21	New CD2	acgaca**atgatag**gataacctagagtctctgaaccatttcgtgtttgcaaaaacaatgctccttgt**ctga**gtca
CD1-1-75	New CD3▲	ttgccg**ttgatga**tcaattgagcatatcactgactgttggtgaggtgatttcattag**ctga**ggct
CD1-2-49	New CD4	cgtctg**gtgagga**tgaaaacaggacaggtttcgctaaaatattaccgaatgccaatatgtcgagacaccttggctgt**ctga**ggac
CD1-2-68	New CD5	aggccc**gtgaaga**cacgaattaccgtctgataactaatgacgctaccatggctgtaaac**ccaga**ggcc
CD1-2-28	New CD6	gttcgg**gcgatga**tcttctgtgattacatcgcacggcgaggtgggaacgcaatacccgcctgccagcccgatt**ctga**acca
CD1-2-33	New CD7▲	cagtcg**aggagga**aagtttccatgcaccactctgaaggatagtgtggattatggttcacaatttа**ctga**ggct
ACA1-2-58	New CD8▲	ctgtca**atgaaga**attgtgcaggggtcctgaaatcggccgagatgtgtgccgatgggctttcgggaatctggatatcggtttctggatcgaagattggttcagaaagatagcaccca**ccga**gaca
Box H/ACA snoRNA		
ACA1-1-7	ΨCeSSU-513b	atgcaccttgcactctacgcctttccttcaatgggttggtatgatttaataaggatgca**agagaa**tagtacgggaagttcgaagacttgcggatttgcttccacggctccgcgagttcataactgtcacc**aca**att
ACA1-1-34	ΨCeSSU-1266b	taacctctcttataagccgggggactagcattttgttaagttcactagtaaaatgagaggt**aaagca**tagagacaaccagacaccgagaatgttttgatgttttcggtcactttgtgtgtccc**aca**aat
ACA1-3-16	ΨCeSSU-1523a	cagcatcgaaaatggacggacttttccgatggatcgtttcgtataatttttggtgc**aaaata**gtagagagacgcagtgttactcttccttacgttacctgtatctggagagtgtcaagcgtttccc**aca**ttt
ACA1-3-28	ΨCeLSU-1058	tggcatgtatttaatgcgggtggatgattttagttcagccgtttagcttgcatgca**aaatta**ttttctattgaagattttattcgctgcactcggtcatattcgagtctatgatttttctgaaattggt**aca**ttt
ACA1-4-19	ΨCeLSU-2966	atgctcttcaaaagcactggttttaggatccactattatccaagccagccgtcaaaactgagct**ataaga**attatcttgttttgggtgaggtgtattcaattcagaatgcgtctcaataacacgatg**aca**att
ACA1-2-63	CeR-8	acgctcttcaaaagcactggttatcggactcagacttgtccatgccagccgtcaaaatgagca**atatga**aatatcctgttttgggtgaggtgtaactgtatttagattagatctcaataacacgatg**aca**gtt
ACA1-1-30	New ACA1	attccttcttatcctgcactcagtttgttcattgactgtggagcctggaaatggaggag**aaagta**aatttgatgattactaatacttcgctgtctttagaggacgcggagattgtgagacttgaa**aca**ttt
ACA1-1-35	New ACA2	tagcatgctgttagagcttgtaaggtatatgtgattttacgagtgttgaagtattgca**aaagca**aaggacgggcacaattgccatgtgttggtattattgcttcaagttatttgaagctgtaatatcaataagcatgtctcgtgtgaagtccg**aca**att
ACA1-2-1	New ACA3	ttgcacgactgcaacctgctttcggaactttaccgtctgcggtacttgtagtgct**atatga**gcgtgggccctccgtgtgcgtgtaatttaattgaactagcactccgttggtcaccgg**ata**att
ACA1-2-11	New ACA4	acgcacgcttatttcgcgccgcaagttttgcaatgacgatgtggctaaagtgtagtgca**atatga**gctcgctcacggcgttgcggaccatagtctgaacacggttcgcgtttatgtgagtgaa**aca**att
ACA1-2-65	New ACA5a	atcgctctccagctccatgcagacgtgaaaaagtcggtgtggcgttttcatgagcgg**aaatta**acattgttccaaaaaacaattgctagtctcctgtgagctaatgatcacctgatggttcaa**aca**ctt
ACA1-3-21	New ACA5b	atcgctctccagctccatgccgatgtaaaaaagtcagtgtggcgttttcatgagcgg**aaatta**tcactgttccaaaaaacaattgctagtctcctgtgagctaatgatcacctgatggttcag**aca**ctt
ACA1-4-93	New ACA6	tgcaactattcaagagatttgcctcccgtggggcacgccttttttcgttgcg**aaataa**aacatgtcctttattcccgctcctggattgtgtttcttgcgcgatgatatggctatg**aca**att
ACA1-3-75	New ACA7	tccacttatttttcaacgtcgccggttctagaatcgatgtgaccccaataagctggt**agataa**gttgttactctgccacacgtgtactggtatcatgctgtatgtgttccaaagattttacaa**aca**att
ACA1-3-89	New ACA8	cgcgtcattttcatccaattggcaacgtgattctaatgttggcgattcagcgtatttctgacgca**aaattg**ataacttctccattgacgtctagtcagactaaactggctcggatacaattagggagttt**aca**ctt
ACA1-3-8	New ACA9	tgccggatttaaacacttctgagttgctcacatgctcagaagaacaaggttcgga**aaatta**gtgatcatttgatgtgctgaactccaaagagtcaaaactgagttggtgatcgt**aca**ttt

Note. Structure elements of snoRNAs are shown in boldface. The sequence of one clone is shown if multiple copies were found in the cDNA libraries. The genes marked with triangles indicate the additional novel snoRNAs that do not overlap with our computational results or the recently reported results [20].

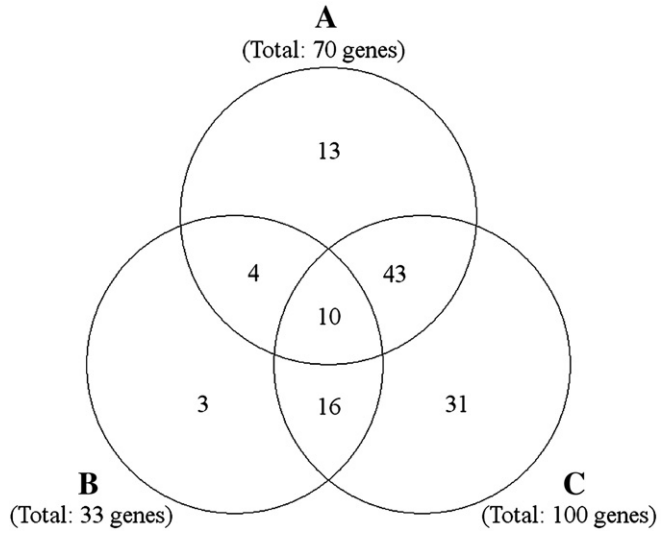

Fig. 3. Venn diagram of the relationship of different snoRNA gene collections involved in this study. "A" represents the computationally identified snoRNA genes using the bioinformatics approach in this study; "B" represents the snoRNAs identified with cDNA cloning in this study. "C" represents the snoRNA genes identified from Deng's cDNA library [20]. The number of each part of the snoRNA genes is shown.

transcription of a canonical H/ACA snoRNA into the middle region of C/D snoRNA genes.

Discussion

In the field of snoRNA research, computational scanning and cDNA cloning are the two main strategies of mining data, both of which have been successfully applied for snoRNA research in the human [23,24], fruitfly [11,25,26], and yeast [9,10,27]. In the previous study by Deng *et al*. (20), cDNA cloning of specific sized fragments was highlighted as a powerful and large-scale technique to identify nematode snoRNAs. However, a genome-wide analysis of nematode snoRNA genes using a bioinformatics approach can contribute a more comprehensive analysis. In this study, we present a genome-wide search of snoRNAs with the snoscan and snoGPS programs in a nematode. More importantly, an additional 13 snoRNAs, which were not identified in this and another cDNA library [20], were computationally detected, including some low-abundance snoRNAs. With our stringent criteria, only 70 candidates were selected. Our computational search also provided a set of potential candidates under less stringent criteria (data not shown), which is useful to the field of nematode snoRNA research. When comparing the computational identifications and experimental results of cDNA libraries, species-specific parameters can be found and applied in the computational programs for better annotation of snoRNA genes in the nematode genomes.

Under selection pressure, functional RNAs, both protein-coding RNAs (ORF) and noncoding RNAs, exhibit highly conserved sequences between related species. Genomic sequence comparisons by virtue of their evolutionary conservation have been extensively used in systematic discovery of functional genomic elements [28]. Recent studies have demonstrated that genomic sequence comparisons have been extremely successful in identifying RNA genes with characteristic structures that perform general biological functions and hence were evolutionarily conserved between related species, such as miRNA [29] and even structural noncoding RNAs [30]. In our study, we have similarly incorporated this strategy into the computational genome-wide search for the repertoire of snoRNAs, a large family of noncoding RNAs, and have likewise successfully detected a large amount of these RNA genes in *C. elegans*. These RNA genes even include some low-abundance snoRNAs, which are difficult to detect within a cDNA library. The filter programs accurately and reliably enabled the snoscan and snoGPS programs to distinguish the authentic snoRNAs from the numerous false-positive candidates (70 snoRNA genes from approximately 30,000 hits) when stringent conservation criteria, sequences, box elements, and functional antisenses, were all considered simultaneously.

Our results covered most of the known *C. elegans* snoRNAs and 50 novel conserved snoRNA genes. However, 8 known *C. elegans* snoRNAs were missed in our search due to a stringent

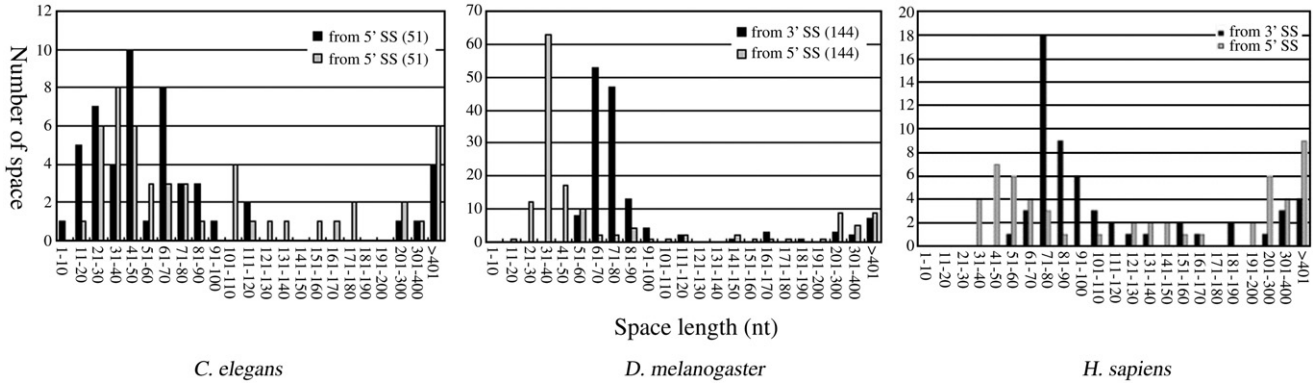

Fig. 4. Comparison of length distributions of space sequences for snoRNA genes/gene clusters in metazoa. The horizontal coordinates of the graphs show the space lengths (nt) from the splice sites to the snoRNA coding regions. The vertical coordinates show the amount of given lengths of spaces. The gray and black bars represent space lengths from the 5′ and 3′ splice sites (ss), respectively, to the snoRNA coding regions. The number in parentheses (if present) indicates the total spaces involved in the investigation. Distribution in *D. melanogaster* and *H. sapiens* were cited from our previous report [14] and Hirose and Steitz's report [22], respectively.

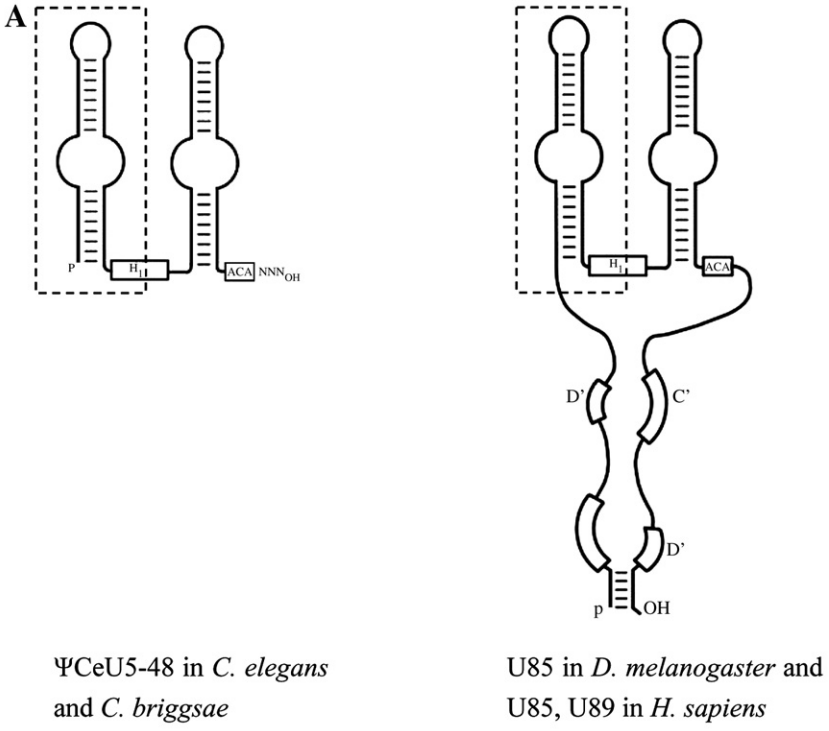

Fig. 5. Sno/scaRNAs guide pseudouridylation of U5-U46 (in fly and human) or U48 (in nematodes). (A) Schematic representation of the homologous sno/scaRNAs in metazoa. The box elements are present and the homologous guide hairpins are boxed. (B) Functional prediction of these sno/scaRNAs. Pseudouridylation guide duplexes between sno/scaRNAs and snRNA are shown. The sno/scaRNAs sequences in a 5′ to 3′ orientation is shown in upper strands, whereas snRNA sequence in a 3′ to 5′ orientation is shown in the lower strands. The sequence motifs are boxed and the upper parts of the hairpins are represented by solid lines.

selection procedure. For example, six sequences of snoRNA genes, sn2524, sn2991, sn3087, mgU6-47, CeR-6, and CeR-8, were excluded from our sequence source due to either their location in the opposite strand of exons or nonconservation in the genome alignment. Two snoRNAs, rpl-7a and rpl-7aps [19], not exhibiting typical box H/ACA structure, also failed to be detected in our search. In addition, many snoRNAs with nonconserved sequences or varied motifs might also have been missed in this study under the strict analyzing program. It is indicated that the computational results showed only part of the conserved nematode snoRNAs. It is worth noting that, similar to a recent study [20], some C/D snoRNAs with atypical box C, D were identified from *C. elegans* cDNA libraries. These snoRNA genes would also be missed in our computational search. In fact, only 14 of the 33 experimentally detected snoRNAs from cDNA libraries were covered in the computational results.

The genomic organization of snoRNA genes is very diverse in different eukaryotes [31]. Most snoRNAs in yeast are encoded by independent singletons and five clusters [9]. SnoRNA gene clusters transcribed from independent promoters have also been shown in plants [32,33], algae [34], and protozoa [15]. In contrast, the majority of snoRNAs in metazoa are intron-encoded. In mammals, almost all the guided snoRNA are nested within the introns of the host genes, with two exceptions reported recently [35]. Our previous study of the *Drosophila* genome showed only 6 of 212 snoRNA genes to be located in the intergenic regions and probably transcribed independently [14]. *C. elegans*, a 959-celled worm, is a model organism that represents the simple multicellular metazoan. Our study addresses the apparent gap of genome-wide studies on snoRNAs between protozoa and higher metazoa. In this study, a high percentage (50%) of the *C. elegans* box C/D snoRNAs

are transcribed independently, showing a different expression strategy from box H/ACA snoRNA genes in the same organism. These data indicate an interesting tendency of an increasing number of snoRNA genes that have "jumped" into the introns of protein-coding or non-protein-coding genes and preserved their location during evolution. Unexpectedly, we did not find any UHG (U snoRNA host gene)-like host genes in *C. elegans*, whereas a number of UHG and dUHG (*Drosophila* U snoRNA host gene) genes with a powerful capability for coding box C/D snoRNAs in their introns have been found in mammals and the fruit fly genomes, respectively [14,36]. This suggests that a different strategy may have developed to utilize noncoding host genes, such as UHGs and dUHGs, to regulate snoRNA expression in higher metazoa.

The snoRNA gene position within the host intron is crucial for assembly of splicing-dependent snoRNP [37]. Studies have revealed that the distance from the intronic snoRNA to the 3′ splice site, especially the distance to the branch point, of the spliceosomal intron is important for the effective processing of the snoRNAs from their host mRNA precursor. Most mammalian intronic snoRNAs are located 71–80 nt upstream of the 3′ splice sites [22] with a similar length of 61–80 nt in *Drosophila* [14] with no exception of less than 50 nt. However, in *C. elegans*, half of the intron-encoded snoRNAs are located less than 50 nt from the 3′ splice sites. This appears not to agree with the model of a splicing-dependent pathway for snoRNA processing in *C. elegans*. However, the critical nucleotide distance from the snoRNA to the branch point of the spliceosomal intron for intronic snoRNA synthesis in most metazoa is supposed to accommodate the binding of the splicing factor $U2AF^{65}$ to the polypyrimidine tract of the intron [38]. Interestingly, without an obvious polypyrimidine tract or convincing branch site consensus in the *C. elegans* introns [39], $U2AF^{65}$ may bind directly to the 3′ splice site consensus sequence [40]. Therefore, the essential distances from intronic snoRNAs to 3′ splice sites (AG) evidently appear shorter than those in other metazoa, which suggests that most of the intron-encoded snoRNAs are expressed in the splicing-dependent pathway in *C. elegans*, despite their short distances to 3′ splice sites.

In conclusion, our study presents a combined computational and experimental search of snoRNA genes in nematodes. The major advance of the scanning is applying the snoscan and snoGPS programs in conjunction with the filter programs CD_filter.pl and ACA_filter.pl to identify a large number (up to 70) of snoRNA genes without one false-positive candidate. Also, in contrast to other metazoa, we demonstrate a complex genome organization and location of snoRNA genes in nematodes that reveals their expression and evolution patterns.

Materials and methods

Extraction of conserved nongenic Caenorhabditis sequences

Genome-wide alignment of *C. elegans* (WormBase, version WS100) and *C. briggsae* (WormBase, version cb25.agp8) was carried out using the LAGAN alignment algorithm [41] and visualized with the VISTA program [42]. We obtained the sequences with the Seq_obtain.pl program and approximately 36,000 candidate conserved nongenic sequences were obtained.

Search for box H/ACA snoRNAs

The snoGPS program [10] was used to screen out potential box H/ACA snoRNAs hidden in the conserved nongenic sequences with the parameters "gapmin" and "gapmax" setting to 15 and 40, respectively. Only *C. elegans* candidate sequences, which contained more than 90 nt of conserved sequence, were used to scan box H/ACA snoRNAs. Since the *Caenorhabditis* rRNA/snRNA pseudouridine sites had not been experimentally determined, the putative pseudouridine sites were mapped according to the conserved pseudouridine sites in *S. cerevisiae* (yeast snoRNA database), *H. sapiens* (snoRNA-LBME-db), and *D. melanogaster* [14]. The candidate snoRNAs with a final program score higher than 15.00 were tested with the ACA_filter.pl program for conservation between two nematodes. An improved mfold program, which was integrated with a preliminary coarse mountain plots filter [11], with default program parameters was then applied to test the secondary structure of the selected perfectly conserved candidates. The mountain plots of candidate snoRNAs were accepted only if they satisfied the following criteria: (1) both of the widths of two mountains exceeded 40 bases; (2) width difference between the two mountains was at most 20 bases; (3) the maximum length of the hinge was 20 bases; and (4) the length of left zero region was at most 5 bases.

Program running with all *Caenorhabditis* rRNAs/snRNAs U residues was also performed to search for probable *Caenorhabditis*-specific box H/ACA snoRNAs with a final cutoff score of 29.00, which was defined from the final score statistic of the results from the above procedures. All the candidates were processed with the above procedures.

Search for box C/D snoRNAs

Snoscan [9], a greedy search algorithm, was performed with default parameters to identify putative box C/D snoRNAs. Candidate sequences with box C/D, rRNA/snRNA complementary region, box D′ when the complementary region was not directly adjacent to box D, and terminal stem were picked out from the 36,000 *C. elegans* conserved nongenic sequences and scored. Methylated sites prepared for the program included 38 experimentally confirmed methylated nucleotides in 26S rRNA [17] and conserved methylated nucleotides of *S. cerevisiae* (yeast snoRNA database), *H. sapiens* (snoRNA-LBME-db), and *D. melanogaster* [14]. The final cutoff score was set to a low value of 10.00 to obtain all potential box C/D snoRNAs. All the obtained candidates were tested with the CD_filter.pl program for conservation between two nematodes.

Description of programs

Program Seq_obtain.pl, ACA_filter.pl and CD_filter.pl were written with PERL. The source codes of the programs are available in Supplementary Materials.

(1) Seq_obtain.pl

In detail, we calculated conserved regions using a 60-bp window and a conservation cutoff score of 69% identity. Conserved nongenic (intronic and intergenic) sequences were extended by 50 nt both upstream and downstream to ensure that conserved snoRNA candidates could be completely identified.

(2) SnoRNAs conservation filter programs: CD_filter.pl and ACA_filter.pl

To eliminate false-positives from the candidate snoRNAs, we applied the CD_filter.pl and ACA_filter.pl programs for picking out the perfectly conserved candidates during the processing of box C/D and box H/ACA candidates. Candidate snoRNAs found in the *C. elegans* genome were examined for corresponding *C. briggsae* conserved structural boxes and, in most cases, rRNA/snRNA complementary sequences in the initial nongenic conserved sequence alignments. Only candidates satisfying these conservation rules were considered reliable and accepted as snoRNA candidates.

Northern blotting analyses and construction and screening of cDNA libraries

Fresh *C. elegans* stain N2 was cultured and mixed-stage worms were collected for RNA extraction. Total cellular RNA was isolated and purified according to the method of guanidine thiocyanate–phenol–chloroform [43].

An aliquot of 40 μg total RNA was analyzed by electrophoresis on 8% acrylamide/7 M urea gels. Electrotransfer onto nylon membrane (Hybond-N^+; Amersham) was followed by UV irradiation for 5 min. Hybridization with 5′-labeled probes was performed as previously described [39].

An aliquot of 50 μg total cellular RNA was polydenylated using poly(A) polymerase (Takara). Synthesis of the first strand of cDNA was performed with 25 μg of poly(A)$^+$-tailed RNA in a 20-μl reaction mix containing 200 U of MMLV reverse transcriptase (Promega) and 0.5 μg of primer oligo(dT) (23-mer) for 45 min at 42°C. The reaction mixture was separated on a denaturing 8% polyacrylamide gel (8 M urea, 1× TBE buffer). cDNAs with sizes ranging from 60 to 120 nt for the box C/D cDNA library and ranging from 120 to 180 nt for the box H/ACA cDNA library were excised and eluted from the gel. cDNAs were tailed with poly(dG) at the 3′-end by using terminal deoxynucleotidyl transferase (Takara), amplified by PCR with primers *Hin*dIII$(T)_{16}$ and *Bam*HI$(C)_{16}$, and cloned into plasmid pTZ18 as described previously [44]. Two cDNA libraries were screened by PCR with the P47 and P48 universal primer pair. Only the recombinant plasmids carrying fragments of the expected size were selected for sequencing, which was performed with an automatic DNA sequencer (Applied Biosystems, 377) using the Big Dye Deoxy Terminator cycle-sequencing kit (Applied Biosystems).

Oligodeoxynucleotides

Oligonucleotides were synthesized and purified by Sangon Co. (Shanghai, China). The sequences of oligonucleotide probes used for Northern blotting and oligonucleotide primers used for cDNA libraries construction and screening are shown in Supplementary Table S2. The probes used in Northern blotting were 5′-end-labeled with [γ-^{32}P]ATP (Yahui Co.) and submitted to purification according to standard laboratory protocols as previously described [45].

Database accession codes

All snoRNA gene sequences identified in this study have been deposited with the EMBL database. Accession numbers are shown in Supplementary Table S1.

Acknowledgments

We thank Xiao-Hong Chen for her technical assistance. We are grateful to Professor M. Ghadessy and Dr R.S. Ghadessy for improving the text. We are also grateful to Professor Todd M. Lowe for kindly releasing the source code of SnoGPS and Snoscan. This research is supported by the National Natural Science Foundation of China (Key Project 30230200, 30370328), funds from the Ministry of Education of China (No. 20030558034), and the National Basic Research Program (No. 2005CB724600).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2006.12.002.

References

[1] A.G. Balakin, L. Smith, M.J. Fournier, The RNA world of the nucleolus: Two major families of small nucleolar RNAs defined by different box elements with related functions, Cell 86 (1996) 823–834.

[2] J.P. Bachellerie, J. Cavaille, L.H. Qu, Nucleotide modifications of eukaryotic rRNAs: The world of small nucleolar RNA guides revisited, in: R.A. Garrett, S. Douthwaite, A. Liljas, A. Matheson, P.B. Moore, H. Noller (Eds.), The Ribosome: Structure, Function, Antibiotics and Cellular Interactions, ASM Press, Washington, DC, 2000, pp. 191–203.

[3] X. Darzacq, B.E. Jady, C. Verheggen, A.M. Kiss, E. Bertrand, T. Kiss, Cajal body-specific small nuclear RNAs: A novel class of 2′-*O*-methylation and pseudouridylation guide RNAs, EMBO J. 21 (2002) 2746–2756.

[4] Z.P. Huang, H. Zhou, L.H. Qu, Maintaining a conserved methylation in plant and insect U2 snRNA through compensatory mutation by nucleotide insertion, IUBMB Life 57 (2005) 693–699.

[5] B. Clouet d'Orval, M.L. Bortolin, C. Gaspin, M.L. Bachellerie, X Box C/D RNA guides for the ribose methylation of archaeal tRNAs: The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp, Nucleic Acids Res. 29 (2001) 4518–4529.

[6] A. Huttenhofer, M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach, J.P. Bachellerie, J. Brosius, RNomics: An experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse, EMBO J. 20 (2001) 2943–2953.

[7] A. Huttenhofer, J. Cavaille, J.P. Bachellerie, Experimental RNomics: A global approach to identifying small nuclear RNAs and their targets in different model organisms, Methods Mol. Biol. 265 (2004) 409–428.

[8] A.D. Gu, H. Zhou, C.H. Yu, L.H. Qu, A novel experimental approach for systematic identification of box H/ACA snoRNAs from eukaryotes, Nucleic Acids Res. 33 (2005) e194, 1–6.

[9] T.M. Lowe, S.R. Eddy, A computational screen for methylation guide snoRNAs in yeast, Science 283 (1999) 1168–1171.

[10] P. Schattner, W.A. Decatur, C.A. Davis, M. Ares Jr., M.J. Fournier, T.M. Lowe, Genome-wide searching for pseudouridylation guide snoRNAs: Analysis of the *Saccharomyces cerevisiae* genome, Nucleic Acids Res. 32 (2004) 4281–4296.

[11] S. Edvardsson, P.P. Gardner, A.M. Poole, M.D. Hendy, D. Penny, V. Moulton, A search for H/ACA snoRNAs in yeast using MFE secondary structure prediction, Bioinformatics 19 (2003) 865–873.

[12] M.C. Accardo, E. Giordano, S. Riccardo, F.A. Digilio, G. Iazzetti, R.A. Calogero, M. Furia, A computational search for box C/D snoRNA genes in the *Drosophila melanogaster* genome, Bioinformatics 20 (2004) 3293–3301.

[13] Z.P. Huang, H. Zhou, D. Liang, L.H. Qu, Different expression strategy: Multiple intronic gene clusters of box H/ACA snoRNA in *Drosophila melanogaster*, J. Mol. Biol. 341 (2004) 669–683.

[14] Z.P. Huang, H. Zhou, H.L. He, C.L. Chen, D. Liang, L.H. Qu, Genome-wide analyses of two families of snoRNA genes from *Drosophila melanogaster*, demonstrating the extensive utilization of introns for coding of snoRNAs, RNA 11 (2005) 1303–1316.

[15] X.H. Liang, S. Uliel, A. Hury, S. Barth, T. Doniger, R. Unger, S. Michaeli, A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification, RNA 11 (2005) 619–645.

[16] K.T. Tycowski, Z.H. You, P.J. Graham, J.A. Steitz, Modification of U6 spliceosomal RNA is guided by other small RNAs, Mol. Cell 2 (1998) 629–638.

[17] S. Higa, N. Maeda, N. Kenmochi, T. Tanaka, Location of 2(′)-*O*-methyl nucleotides in 26S rRNA and methylation guide snoRNAs in *Caenorhabditis elegans*, Biochem Biophys Res Commun. 297 (2002) 1344–1349.

[18] M. Wachi, T. Ogawa, K. Yokoyama, Y. Hokii, M. Shimoyama, A. Muto, C. Ushida, Isolation of eight novel *Caenorhabditis elegans* small RNAs, Gene 335 (2004) 47–56.

[19] Q.M. Mitrovich, P. Anderson, mRNA surveillance of expressed pseudogenes in *C. elegans*, Curr. Biol. 15 (2005) 963–967.

[20] W. Deng, X. Zhu, G. Skogerbo, Y. Zhao, Z. Fu, Y. Wang, H. He, L. Cai, H. Sun, C. Liu, B. Li, B. Bai, J. Wang, D. Jia, S. Sun, H. He, Y. Cui, Y. Wang, D. Bu, R. Chen, Organization of the *Caenorhabditis elegans* small non-coding transcriptome: Genomic features, biogenesis, and expression, Genome Res. 16 (2006) 20–29.

[21] A. Nakao, M. Yoshihama, N. Kenmochi, RPG: The Ribosomal Protein Gene database, Nucleic Acids Res. 32 (2004) D168–D170.

[22] 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. USA 98 (2001) 12914–12919.
[23] A.M. Kiss, B.E. Jady, E. Bertrand, T. Kiss, Human box H/ACA pseudouridylation guide RNA machinery, Mol. Cell Biol. 24 (2004) 5797–5807.
[24] P. Schattner, S. Barberan-Soler, T.M. Lowe, A computational screen for mammalian pseudouridylation guide H/ACA RNAs, RNA 12 (2006) 15–25.
[25] G. Yuan, C. Klambt, J.P. Bachellerie, J. Brosius, A. Huttenhofer, RNomics in *Drosophila melanogaster*: Identification of 66 candidates for novel non-messenger RNAs, Nucleic Acids Res. 31 (2003) 2495–2507.
[26] M.C. Accardo, E. Giordano, S. Riccardo, F.A. Digilio, G. Iazzetti, R.A. Calogero, M. Furia, A computational search for box C/D snoRNA genes in the *Drosophila melanogaster* genome, Bioinformatics 20 (2004) 3293–3301.
[27] C. Torchet, G. Badis, F. Devaux, G. Costanzo, M. Werner, A. Jacquier, The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in *Saccharomyces cerevisiae*, RNA 11 (2005) 928–938.
[28] E.T. Dermitzakis, A. Reymond, R. Lyle, N. Scamuffa, C. Ucla, S. Deutsch, B.J. Stevenson, V. Flegel, P. Bucher, C.V. Jongeneel, S.E. Antonarakis, Numerous potentially functional but non-genic conserved sequences on human chromosome 21, Nature 420 (2002) 578–582.
[29] E. Berezikov, V. Guryev, J. van de Belt, E. Wienholds, R.H. Plasterk, E. Cuppen, Phylogenetic shadowing and computational identification of human microRNA genes, Cell 120 (2005) 21–24.
[30] S. Washietl, I.L. Hofacker, M. Lukasser, A. Huttenhofer, P.F. Stadler, Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome, Nat. Biotechnol. 23 (2005) 1383–1390.
[31] 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.
[32] J.W. Brown, G.P. Clark, D.J. Leader, C.G. Simpson, T. Lowe, Multiple snoRNA gene clusters from *Arabidopsis*, RNA 7 (2001) 1817–1832.
[33] L.H. Qu, Q. Meng, H. Zhou, Y.Q. Chen, Identification of 10 novel snoRNA gene clusters from *Arabidopsis thaliana*, Nucleic Acids Res. 29 (2001) 1623–1630.
[34] A.G. Russell, M.N. Schnare, M.W. Gray, Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*, RNA 10 (2004) 1034–1046.
[35] K.T Tycowski, A. Alar, J.A. Steitz, Guide RNAs with 5′ caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa, Curr. Biol. 14 (2004) 1985–1995.
[36] 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.
[37] T. Hirose, M.D. Shu, J.A. Steitz, Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells, Mol. Cell 12 (2003) 113–123.
[38] J. Valcarcel, R.K. Gaur, R. Singh, M.R. Green, Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA, Science 273 (1996) 1706–1709.
[39] T. Blumenthal, K. Steward, in: *C. elegans* II, D. Riddle, T. Blumenthal, B. Meyer, J. Priess (Eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997, pp. 117–145.
[40] D.A. Zorio, T. Blumenthal, Both subunits of U2AF recognize the 3′ splice site in *Caenorhabditis elegans*, Nature 402 (1999) 835–838.
[41] M. Brudno, C.B. Do, G.M. Cooper, M.F. Kim, E. Davydov, E.D. Green, A. Sidow, S. Batzoglou, LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA, Genome Res. 13 (2003) 721–731.
[42] C. Mayor, M. Brudno, J.R. Schwartz, A. Poliakov, E.M. Rubin, K.A. Frazer, L.S. Pachter, I. Dubchak, VISTA: Visualizing global DNA sequence alignments of arbitrary length, Bioinformatics 16 (2000) 1046–1047.
[43] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction, Anal. Biochem. 162 (1987) 732–735.
[44] H. Zhou, Y.Q. Chen, Y.P. Du, L.H. Qu, The *Schizosaccharomyces pombe* mgU6-47 gene is required for 2′-*O*-methylation of U6 snRNA at A41, Nucleic Acids Res. 30 (2002) 894–902.
[45] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.