Copyright © 2008 by the Genetics Society of America
DOI: 10.1534/genetics.107.086025

Genomewide Analysis of Box C/D and Box H/ACA snoRNAs in *Chlamydomonas reinhardtii* Reveals an Extensive Organization Into Intronic Gene Clusters

Chun-Long Chen,*,†,‡,[1] Chong-Jian Chen,*,†,§,[1] Olivier Vallon, Zhan-Peng Huang,* Hui Zhou* and Liang-Hu Qu*,[2]**

**Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory for Biocontrol, Zhongshan University, Guangzhou 510275, People's Republic of China, †Université Paris-Sud 11, Orsay 91405, France, ‡Centre National de la Recherche Scientifique, UMR 8080, Orsay 91405, France, §Centre National de la Recherche Scientifique, UMR 8621, Orsay 91405, France and **Université Paris 6, Centre National de la Recherche Scientifique, UMR 7141, Institut de Biologie Physico-Chimique, Paris 75005, France*

Manuscript received December 15, 2007
Accepted for publication January 29, 2008

ABSTRACT

Chlamydomonas reinhardtii is a unicellular green alga, the lineage of which diverged from that of land plants >1 billion years ago. Using the powerful small nucleolar RNA (snoRNA) mining platform to screen the *C. reinhardtii* genome, we identified 322 snoRNA genes grouped into 118 families. The 74 box C/D families can potentially guide methylation at 96 sites of ribosomal RNAs (rRNAs) and snRNAs, and the 44 box H/ACA families can potentially guide pseudouridylation at 62 sites. Remarkably, 242 of the snoRNA genes are arranged into 76 clusters, of which 77% consist of homologous genes produced by small local tandem duplications. At least 70 snoRNA gene clusters are found within introns of protein-coding genes. Although not exhaustive, this analysis reveals that *C. reinhardtii* has the highest number of intronic snoRNA gene clusters among eukaryotes. The prevalence of intronic snoRNA gene clusters in *C. reinhardtii* is similar to that of rice but in contrast with the one-snoRNA-per-intron organization of vertebrates and fungi and with that of *Arabidopsis thaliana* in which only a few intronic snoRNA gene clusters were identified. This analysis of *C. reinhardtii* snoRNA gene organization shows the functional importance of introns in a single-celled organism and provides evolutionary insight into the origin of intron-encoded RNAs in the plant lineage.

SMALL nucleolar RNAs (snoRNA) are one of the largest classes of noncoding RNAs in eukaryotes. They play an essential role in ribosomal RNAs (rRNA) biosynthesis (Maxwell and Fournier 1995). A small fraction of snoRNAs such as U3, U8, U14, U22, U17, and RNase MRP RNA are involved in the cleavage of pre-rRNAs (Venema and Tollervey 1999). However, most of them guide the 2′-*O*-ribose methylation and pseudouridylation of rRNAs (Smith and Steitz 1997). On the basis of common sequence motifs and structural features, all snoRNAs except RNase MRP fall into two families: box C/D snoRNAs and box H/ACA snoRNAs, which guide site-specific 2′-*O*-ribose methylations and pseudouridylations of rRNAs, respectively, via base complementarity (Balakin *et al.* 1996; Bachellerie *et al.* 2000; Kiss 2001). The box C/D snoRNAs display two conserved motifs, the 5′-end C box (5′-RUGAUGA-3′) and the 3′-end D box (5′-CUGA-3′), usually flanked by short inverted repeats. In addition to the H box (ANANNA) in the hinge region and an ACA motif 3 nucleotides upstream of the 3′-end of the molecule, box H/ACA snoRNAs are characterized by a common hairpin–hinge–hairpin–tail secondary structure (Ganot *et al.* 1997; Ni *et al.* 1997). The spectrum of snoRNA targets is continuously growing. They are now known to guide post-transcriptional modifications of snRNAs (Tycowski *et al.* 1998; Zhou *et al.* 2002) and tRNAs (Clouet d'Orval *et al.* 2001; Zemann *et al.* 2006), as well as the alternative splicing of a pre-mRNA (Cavaille *et al.* 2000; Kishore and Stamm 2006). Furthermore, some "orphan" snoRNAs with no obvious target of rRNAs and snRNAs have also been reported (Huttenhofer *et al.* 2001; Chen *et al.* 2003).

The genomic organization of snoRNA genes displays a great diversity in various organisms. In human, almost all of the snoRNAs are encoded by single genes nested within introns and maturated by a splicing-dependent processing (Tycowski *et al.* 1993; Kiss and Filipowicz 1995; Bachellerie *et al.* 2000). A few of them are transcribed independently with their own promoter. In the intron-poor genome of *Saccharomyces cerevisiae*, only a few of the snoRNA genes are intronic. The majority of

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU410622–EU410943. Sequence data are also available as annotation tracks displayed on the JGI *C. reinhardtii* genome browser.

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Biotechnology Research Center, Zhongshan University, Guangzhou 510275, People's Republic of China.
E-mail: lssqlh@mail.sysu.edu.cn

the snoRNAs are encoded by single genes, but 17 (20%) snoRNAs are encoded in five gene clusters each driven by an upstream promoter (Lowe and Eddy 1999; Qu *et al.* 1999). In contrast, snoRNA gene clusters dominate in the plant lineage. More than 80% of the snoRNA genes of *Arabidopsis thaliana* and *Oryza sativa* are organized into gene clusters (Leader *et al.* 1997; Brown *et al.* 2003a; Chen *et al.* 2003). Four intronic snoRNA gene clusters were found in *A. thaliana*, and one-half of the *O. sativa* clusters have been found within introns of protein-coding genes.

Chlamydomonas reinhardtii is a unicellular green alga belonging to Chlorophyta, a phylum that diverged from land plants over one billion years ago, following the eukaryotic radiation that gave rise to the animal, fungal, and plant kingdoms (Merchant *et al.* 2007). Comparison of the *C. reinhardtii* proteome to *Homo sapiens* and *A. thaliana* showed that *C. reinhardtii* proteins are overall slightly more similar to *A. thaliana* than to human proteins. However, *C. reinhardtii* shares many genes with animals, in particular those associated with flagellar functions. These proteins have been inherited from the common ancestor of plants and animals, but were lost in land plants (Li *et al.* 2004). The availability of the *C. reinhardtii* genome provides an intriguing opportunity to investigate the ancestor snoRNA gene constitution and organization in the plant lineage. Here we present a genomewide analysis of two major families of snoRNAs, *i.e.*, box C/D and box H/ACA snoRNAs in the *C. reinhardtii* genome and compare them with those of *Volvox carteri*, a multicellular green alga diverged from *C. reinhardtii* ~50 million years ago (Coleman 1999) and of other eukaryotes including plants, animals, and fungi.

MATERIALS AND METHODS

The draft sequence of the *C. reinhardtii* genome assembly v3.1, the *Volvox carteri* f. *nagariensis* genome assembly v1.0, and gene annotations provided in the GeneCatalog track were downloaded from genome browsers in the DOE Joint Genome Institute (JGI) (http://genome.jgi-psf.org). SnoRNA genes of *C. reinhardtii* were identified using the snoRNA mining platform (snoRMP) (Chen *et al.* 2003; Huang *et al.* 2007), which is based on the SnoScan (Lowe and Eddy 1999) and SnoGPS (Schattner *et al.* 2006) algorithms. Further characterization was based on secondary structure prediction, gene organization, and comparative genomic analysis.

The box C/D snoRNA search program identified segments <150 bp, with box C (RUGAUGA), box D (CUGA), at least 10 nucleotides complementary (Watson–Crick and G:U base pairs) to a rRNA (5.8S, 18S, or 26S rRNA) or snRNA (U1, U2, U4, U5, or U6 snRNA) sequence and terminal short inverted repeats. Among these parameters, we considered primarily the quality of the potential snoRNA-target duplexes (the duplex length, the number and position of the duplex mismatches, and those of the GU pairings). The box H/ACA snoRNA search program identified segments <200 bp exhibiting a typical hairpin–hinge–hairpin–tail secondary structure, with box H (ANANNA) in the hinge region and box ACA in the tail, as well as two sequences of the pocket of the hairpin with complementarity to a rRNA or snRNA sequence (each >3 nucleotides and the sum >9 nucleotides). Sequences flanking the snoRNA candidates were examined for other possible snoRNAs, and BLAST searches (Altschul *et al.* 1997) were performed within the *C. reinhardtii* genome to identify paralogs or within the *V. carteri* genome to identify orthologs. The Vista track (Mayor *et al.* 2000) constructed with the *C. reinhardtii* and *V. carteri* genomes with a window of 100 bp and a minimum percentage of conservation identity of 60% was used to recover orphan snoRNA genes in the regions flanking the newly discovered snoRNAs. SnoRNA secondary structures were predicted with the Mfold program (Mathews *et al.* 1999). Sequence alignments and phylogenetic trees were performed with ClustalW (Thompson *et al.* 1994) and T-coffee (Notredame *et al.* 2000). The interval sequences of snoRNAs were scanned to check for AGNN stem-loop structures by using the program designed to identify Rnt1p cleavage signals with a cutoff of final score = 0.8 (Ghazal *et al.* 2005).

RESULTS

Identification of 322 snoRNA genes in the *C. reinhardtii* genome: Using the snoRMP to screen the *C. reinhardtii* genome (see materials and methods), 316 putative snoRNA genes have been identified, with 183 of the box C/D type and 133 of the box H/ACA type (Tables 1 and 2). On the basis of sequence similarity (Blast E value <0.01) and antisense element conservation, the genes defined 72 box C/D and 42 box H/ACA snoRNA families (see Figure 1 for an alignment of CD11 snoRNAs). The orthologs of 93 snoRNA families could be detected in the *V. carteri* genome by sequence homology search. Furthermore, comparative genomic analysis indicated that one-third of the *C. reinhardtii* snoRNA genes reside in regions that are highly conserved between *C. reinhardtii* and *V. carteri*: 71 genes from 48 box C/D snoRNA families and 36 genes from 21 box H/ACA snoRNA families could be viewed as peaks in the Vista track. Manual inspection of the *C. reinhardtii–V. carteri* conserved regions flanking the newly discovered snoRNA genes allowed further identification of 6 additional putative snoRNAs, representing 4 families that appeared as orphan snoRNAs with no obvious rRNA or snRNA target: 2 single-copy box C/D snoRNAs, 1 single-copy box H/ACA snoRNA, and 1 box H/ACA snoRNA family with 3 members. Overall, 55 of the 118 *C. reinhardtii* snoRNA families have only 1 member, and 63 contain multiple members ranging from 2 to 14.

Properties of *C. reinhardtii* box C/D and box H/ACA snoRNA functions: The 74 box C/D snoRNA families of the *C. reinhardtii* genome were predicted to guide methylation at 96 sites of rRNAs, including 1, 33, and 62 sites of the 5.8S, 18S, and 26S rRNAs, respectively, and 3 sites of the U6 snRNAs (Table 1). More than 80% of the *C. reinhardtii* rRNA putative methylated sites have analogs in plants (*A. thaliana* and *O. sativa*), fungi (*S. cerevisiae and Schizosaccharomyces pombe*), or animals (*Homo sapiens* and *Mus musculus*) (Table 3A). These include 27 sites well conserved among all three groups, 7 shared with plants and fungi, 11 with plants and animals, and 1 with fungi and animals; 24 have analogs solely in plants, 2 in fungi,

TABLE 1

Box C/D snoRNA genes in *C. reinhardtii*

snoRNA	Iso	Target	Antisense (nt)	Homology			Volvox iso
				Plants	Fungi	Animals	
CrCD01	5	Cr26S-G2754	18 (5′)	snoR1	SnR48	—	1
CrCD02	2	Cr26S-A1819	12 (5′)	—	—	BII420	—
CrCD03	6	Cr26S-A2603	12 (5′)	snoR68Y	SnR68	—	—
CrCD04	2	Cr26S-U1892	13 (5′)	—	—	U50	1
CrCD05	2	CrU6-C55	11 (5′)	—	—	U94	1
CrCD06	5	Cr18S-G1408	12 (5′)	snoR69	—	—	3
CrCD07	2	Cr26S-C2799	13 (3′)	snoR24	—	—	1
CrCD09	4	Cr26S-A795	13 (5′)	U51	SnR39	U51	3
		Cr18S-U1539	15 (3′)	—	—	—	
CrCD10	7	Cr26S-U1043	16 (5′)	snoR41Y	—	—	—
CrCD11	10	Cr18S-A419	13 (5′)	snoR120	SnR52	U83	3
		Cr26S-C1533	11 (3′)	—	—	—	
		Cr26S-C2645	14 (3′)	snoR162	—	—	
CrCD13	5	Cr18S-A539	14 (3′)	snoR41Y	SnR41	U62	2
CrCD14	1	Cr18S-G1567	10 (5′)	—	SnR57	—	1
CrCD15	6	Cr26S-C2443	13 (5′)	—	—	—	3
		Cr26S-A2432	11 (3′)	—	—	—	
CrCD16	6	Cr18S-U609	13 (5′)	snoR13	—	—	6
		Cr26S-G1425	11 (3′)	—	—	—	
CrCD17	3	Cr26S-U1868	15 (5′)	U34	SnR62	U34	1
CrCD18	3	Cr26S-C1454	13 (3′)	Z270	—	—	2
CrCD19	1	Cr18S-G593	11 (5′)	U54	—	U54	2
CrCD20	5	Cr26S-G1831	12 (3′)	U59	—	—	—
CrCD21	2	Cr26S-C1492	15 (5′)	U49	—	—	—
CrCD22	4	Cr26S-G2088	11 (5′)	snoR60	—	—	2
CrCD23	5	Cr18S-A1199	15 (5′)	—	—	—	2
		Cr18S-C1211	13 (3′)	snoR130	—	BII142	
CrCD24	3	Cr18S-A1087	15 (3′)	—	—	—	1
CrCD25	2	Cr18S-G632	12 (5′)	J27	—	BII108	2
		Cr18S-A617	14 (3′)	U36	SnR47	U36	
CrCD26	2	Cr18S-C413	21 (3′)	U14	U14	U14	5
CrCD27	1	Cr26S-C2922	15 (5′)	U35	SnR73	U35	2
CrCD28	1	Cr26S-G2582	11 (5′)	U35	SnR73	U35	2
CrCD29	1	Cr26S-A2909	13 (5′)	U29	SnR71	U29	2
CrCD30	1	Cr26S-A639	15 (5′)	U18	U18	U18	2
CrCD31	1	Cr26S-U2385	13 (5′)	snoR37	SnR78	U52	—
		Cr26S-C2329	13 (3′)	snoR37	—	U53	
CrCD32	4	Cr18S-U122	12 (5′)	—	—	Z17	3
		Cr18S-G1122	12 (3′)	snoR41YII	SnR41	—	
CrCD33	1	Cr18S-C38	13 (5′)	snoR66	—	—	1
CrCD34	1	Cr26S-A2874	14 (5′)	snoR31	—	—	1
		Cr5.8S-A43	13 (3′)	snoR9	—	—	
CrCD35	1	Cr26S-U2613	14 (5′)	snoR10	—	U58	1
CrCD36	1	Cr18S-A970	12 (5′)	U59	SnR54	U59	2
CrCD37	1	Cr26S-A2285	13 (3′)	U30	—	U30	1
CrCD38	1	Cr18S-U598	13 (5′)	Z267	—	—	1
CrCD39	1	Cr26S-U2077	13 (5′)	—	—	—	1
CrCD40	1	Cr26S-C2301	13 (5′)	snoR44	SnoR64	Z18	1
		Cr26S-A2290	13 (3′)	snoR44	—	Z22	
CrCD41	6	Cr26S-G1907	13 (5′)	—	—	U50	3
CrCD43	1	Cr18S-U1227	11 (5′)	snoR14	SnR82	BII55	1
		Cr18S-U1374	14 (3′)	U61	—	U61	
CrCD44	1	Cr18S-U576	12 (5′)	snoR77Y	SnR77	—	1
CrCD45	1	Cr18S-U1438	15 (3′)	snoR19	—	—	1
CrCD46	2	Cr26S-C1836	12 (5′)	snoR15	—	U39	—
CrCD47	3	Cr18S-G560	14 (5′)	—	snR80	—	—
		Cr18S-C585	11 (3′)	—	—	—	

(*continued*)

TABLE 1

(Continued)

snoRNA	Iso	Target	Antisense (nt)	Homology			Volvox iso
				Plants	Fungi	Animals	
CrCD48	7	Cr26S-G2200	14 (5′)	U36	—	—	4
		Cr26S-A2184	17 (3′)	U36	snR47	U36	
CrCD49	1	Cr26S-U2698	13 (3′)	snoR68	—	—	1
CrCD50	6	Cr26S-G2252	11 (5′)	U15	snR75	U15	4
		Cr26S-A2245	11 (3′)	U15	—	U15	
CrCD51	1	Cr26S-A864	13 (5′)	snoR72Y	snR72	—	1
		Cr18S-C147	13 (3′)	—	—	—	—
CrCD52	1	Cr26S-C1983	12 (5′)	—	—	—	—
CrCD53	1	Cr26S-G896	15 (3′)	U80	snR60	Z15	2
CrCD54	1	Cr26S-A805	12 (5′)	U80	snR60	Z15	—
CrCD55	1	Cr18S-A28	11 (5′)	U27	snR74	U27	1
CrCD56	1	Cr18S-A795	13 (5′)	snoR53Y	snR53	—	1
		Cr26S-G1419	13 (3′)	—	—	—	
CrCD57	1	Cr26S-A1118	13 (5′)	U38	snR61	U38	1
CrCD58	1	Cr26S-C2268	15 (5′)	—	—	—	1
		CrU6-C70	14 (5′)	—	gU6-77	—	
		CrU6-A40	12 (3′)	—	gU6-47	—	
CrCD59	1	Cr26S-A1434	11 (5′)	U24	U24	U76	1
		Cr26S-C1422	10 (3′)	U24	U24	U24	
CrCD60	2	Cr26S-U2560	12 (5′)	—	—	—	2
CrCD61	1	Cr18S-G1424	12 (5′)	snoR19	snR56	U25	1
CrCD62	4	Cr26S-U784	10 (5′)	—	—	—	4
		Cr26S-G793	11 (3′)	snoR39BY	snR39b	snR39B	
CrCD63	2	Cr18S-C1634	11 (5′)	U43	snR70	U43	2
CrCD64	1	Cr26S-A2220	11 (3′)	U40	snR63	U40	1
CrCD65	3	Cr26S-A905	12 (5′)	snoR133	snR84	—	3
		Cr26S-C849	13 (3′)	—	—	—	
CrCD66	7	Cr26S-A2641	15 (5′)	—	—	—	2
CrCD67	2	Cr26S-A1352	12 (3′)	snoR7	—	—	2
CrCD68	1	Cr26S-G2778	12 (5′)	snoR38Y	snR38	snR38	2
		Cr18S-A161	10 (3′)	snoR18	—	U44	
CrCD69	1	Cr26S-G2373	12 (5′)	—	—	—	1
		Cr26S-G2355	11 (3′)	snoR29	—	—	
CrCD70	1	Cr26S-A2090	12 (5′)	snoR12	—	—	1
CrCD71	1	Cr26S-G2756	14 (3′)	—	snR48	U60	—
CrCD72	1	Orphan		—	—	—	4
CrCD73	6	Cr18S-A1322	10 (5′)	snoR32	—	—	3
CrCD74	1	Orphan		—	—	—	1
CrCD75	1	Cr26S-A1875	13 (3′)	—	—	—	1
CrCD76	3	Cr26S-U1253	11 (5′)	snoR22	—	—	2
CrCD77	2	Cr18S-G559	12 (3′)	—	—	—	—

SnoRNA family names are listed with the numbers of isoforms in the *C. reinhardtii* (Iso) *and V. carteri* (Volvox iso) genomes and the rRNA and snRNA targets. Antisense sequence lengths are indicated and the modification sites are compared to those of plants (*A. thaliana* and *O. sativa*) (Brown *et al.* 2003b; Chen *et al.* 2003), fungi (*S. cerevisiae* and *S. pombe*) (Li *et al.* 2005; Piekna-Przybylska *et al.* 2007), and animals (*H. sapiens* and *M. musculus*) (Huttenhofer *et al.* 2001; Lestrade and Weber 2006). SnoRNAs that lack a target site are indicated as orphan. —, no corresponding snoRNA identified.

and 5 in animals. Remarkably, 19 rRNA putative methylated sites that did not have any analogs are probably *C. reinhardtii* specific. In addition, among the 3 U6 snRNA putative methylated sites, 2 are conserved with fungi, and 1 is *C. reinhardtii* specific.

The 44 Box H/ACA snoRNA families of the *C. reinhardtii* genome were predicted to guide pseudouridylation at 60 sites of rRNAs, including 26 and 34 sites of the 18S and 26S rRNAs, respectively, and 2 sites of the U6 snRNAs (Table 2). Among the 60 *C. reinhardtii* rRNA putative pseudouridylation sites, 37 have analogs in one of the six species of higher plants, fungi, and animals (Table 3B), including 6 sites common to the three kingdoms; 3 have analogs within both plants and fungi, 7 within both plants and animals, and 3 within both fungi and animals; 9 have analogs solely in plants, 3 in fungi,

TABLE 2

Box H/ACA snoRNA genes in *C. reinhardtii*

snoRNA	Iso	Target	Antisense (nt)	Homology			Volvox iso
				Plants	Fungi	Vertebrates	
CrACA01	8	Cr26S-U2278	7 + 4 (5′)	snoR83	snR86	ACA48	2
		Cr26S-U2315	5 + 8 (3′)	snR82	—	—	
CrACA02	1	Cr26S-U887	5 + 4 (5′)	—	—	—	1
CrACA03	1	Cr26S-U46	5 + 6 (5′)	—	—	—	1
CrACA04	7	Cr18S-U558	7 + 4 (3′)	—	—	ACA24	3
CrACA05	3	Cr26S-U2175	7 + 5 (3′)	snoR99	snR90	ACA27	2
CrACA06	2	Cr26S-U873	5 + 6 (5′)	Osaca052[a]	—	—	—
		CrU6-U88	5 + 6 (3′)	—	—	—	
CrACA07	1	Cr18S-U831	6 + 5 (5′)	—	—	—	—
		Cr26S-U1994	5 + 4 (3′)	—	—	—	
CrACA08	2	Cr26S-U857	4 + 5 (5′)	—	—	—	1
		Cr26S-U923	4 + 5 (5′)	—	—	—	
CrACA09	2	Cr26S-U2311	6 + 5 (3′)	Osaca003[a]	—	—	1
CrACA10	10	Cr18S-U1301	6 + 5 (5′)	—	—	—	3
		Cr18S-U1707	3 + 6 (5′)	—	—	—	
CrACA13	2	Cr26S-U2789	6 + 3 (3′)	snoR2	snR34	U65	—
CrACA15	1	Cr26S-U3281	6 + 3 (3′)	—	—	ACA22	—
CrACA16	3	Cr18S-U110	6 + 5 (5′)	snoR100	—	ACA42	3
		Cr18S-U206	7 + 4 (3′)	—	snR49	—	
CrACA18	1	Cr18S-U1187	7 + 5 (5′)	—	snR35	ACA13	1
		Cr18S-U708	5 + 4 (3′)	—	—	—	
		Cr18S-U1246	7 + 3 (3′)	—	—	—	
CrACA19	1	Cr26S-U857	6 + 5 (5′)	—	—	—	1
		Cr26S-U1028	7 + 4 (3′)	—	—	—	
CrACA21	1	Cr26S-U2230	6 + 8 (5′)	Osaca019[a]	snR84	—	1
CrACA22	4	Cr26S-U2828	6 + 6 (5′)	—	snR46	ACA16	1
CrACA23	1	Cr26S-U1765	5 + 5 (3′)	—	—	—	1
CrACA24	2	Cr18S-U942	5 + 8 (5′)	—	—	—	1
		Cr26S-U2311	6 + 5 (3′)	Osaca003[a]	—	—	
CrACA26	1	Cr18S-U757	5 + 5 (3′)	snoR91	—	ACA25	—
CrACA28	5	Cr18S-U1131	4 + 5 (5′)	—	—	—	—
		Cr26S-U1820	5 + 4 (3′)	—	—	—	
		Cr18S-U1598	5 + 5 (3′)	—	—	—	
CrACA29	1	Cr26S-U3074	7 + 4 (5′)	Osaca019[a]	—	ACA17	1
		Cr26S-U2938	6 + 9 (3′)	Osaca003[a]	snR42	ACA27	
CrACA30	1	Cr18S-U1428	7 + 3 (5′)	—	—	—	1
		Cr18S-U1183	6 + 5 (3′)	—	—	ACA36	
CrACA31	3	Cr26S-U2100	6 + 5 (5′)	snoR87	—	ACA19	3
CrACA32	6	Cr26S-U1224	7 + 3 (5′)	snoR96	—	—	3
CrACA33	4	Cr18S-U600	6 + 7 (5′)	snoR91	—	ACA20	1
CrACA35	7	Cr26S-U2380	7 + 3 (5′)	—	snR11	ACA3	5
CrACA36	2	Cr18S-U1428	4 + 5 (5′)	—	—	—	—
		Cr26S-U224	4 + 5 (5′)	—	—	—	
		Cr18S-U200	5 + 5 (5′)	—	—	—	
		Cr26S-U2592	6 + 8 (3′)	snoR78	—	—	
CrACA37	1	Cr18S-U357	7 + 3 (5′)	snoR86	—	U71	—
CrACA38	5	Cr18S-U1556	5 + 7 (3′)	Osaca053[a]	—	ACA5	4
CrACA39	14	Cr18S-U102	4 + 7 (3′)	Osaca025[a]	—	—	5
CrACA40	4	Cr26S-U764	6 + 4 (5′)	snoR82	snR80	—	7
		Cr26S-U807	5 + 4 (3′)	snoR77	—	—	
CrACA41	6	Cr18S-U1623	6 + 4 (5′)	—	—	U70	2
CrACA42	6	Cr26S-U2685	4 + 5 (5′)	—	—	ACA34	6
		Cr18S-U301	6 + 6 (5′)	—	snR49	—	
		Cr18S-U378	7 + 5 (3′)	—	—	—	
CrACA43	1	Cr18S-U1210	7 + 5 (3′)	Osaca069[a]	—	—	—
CrACA44	1	Cr26S-U1667	5 + 8 (5′)	—	snR91	—	1

(*continued*)

TABLE 2

(Continued)

snoRNA	Iso	Target	Antisense (nt)	Homology			Volvox iso
				Plants	Fungi	Vertebrates	
CrACA45	2	Cr18S-U985	7 + 3 (5′)	—	—	—	2
CrACA46	3	Cr26S-U2278	5 + 8 (5′)	snoR83	snR86	ACA48	—
CrACA48	1	CrU6-U24	5 + 8 (5′)	—	—	ACA65	—
CrACA50	2	Cr26S-U2224	7 + 9 (5′)	U19	snR191	—	3
CrACA51	1	Cr26S-U2817	6 + 4 (5′)	—	—	ACA21	—
CrACA52	1	Cr26S-U2886	6 + 4 (3′)	snoR74	snR10	ACA21	—
CrACA54	3	Orphan		—	—	—	3
CrACA55	1	Orphan		—	—	—	1

SnoRNA family names are listed with the number of isoforms in the *C. reinhardtii* (Iso) *and V. carteri* (Volvox iso) genomes and the rRNA and snRNA targets. Antisense sequence lengths are indicated and the modification sites are compared to those of plants (*A. thaliana* and *O. sativa*), (Brown *et al.* 2003b; Chen *et al.* 2003), fungi (S. cerevisiae *and S. pombe*) (Li *et al.* 2005; Piekna-Przybylska *et al.* 2007), and animals (*H. sapiens* and *M. musculus*) (Huttenhofer *et al.* 2001; Lestrade and Weber 2006). SnoRNAs that lack a target site are indicated as orphan. —, no corresponding snoRNA identified.

[a] Plant snoRNAs that are from our unpublished data.

and 6 in animals. Twenty-three sites display as *C. reinhardtii* specific. Among the 2 snRNA putative pseudouridylation sites, 1 is conserved with animals, the other is *C. reinhardtii* specific.

The predicted modification pattern of *C. reinhardtii* rRNAs was closely related to that of land plants. More than 72% of the *C. reinhardtii* rRNA putative methylation sites are conserved with land plants compared to only 46 and 39% with animal and fungi, respectively. Pseudouridylation sites are also more conserved with land plants (42%) than with animal (37%) and fungi (25%). The fact that more pseudouridylation sites appear specific to *C. reinhardtii* is probably a consequence of the smaller number of box H/ACA snoRNAs that have been identified in other organisms, especially in higher plants than box C/D snoRNAs. It is worth noting that although we used all the U1, U2, U4, U5, and U6 snRNAs as target to look for the putative modification sites guiding by snoRNAs, only the snoRNAs guiding U6 snRNA modification were found. This correlates with the fact that synthesis and maturation processes are different between the RNA polymerase III-transcribed U6 snRNA and the polymerase II-transcribed snRNAs (U1, U2, U4, and U5). In mammals, the 2′-*O*-methylation and pseudouridylation of the U6 snRNA takes place in the nucleolus, and it is directed by snoRNAs (Tycowski *et al.* 1998; Ganot *et al.* 1999). While the RNA polymerase II-transcribed snRNAs undergo site-specific post-transcriptional modification in Cajal bodies directed by another type of ncRNAs called small Cajal body-specific RNA (scaRNAs) (Darzacq *et al.* 2002; Kiss 2004).

Prevalence of intronic snoRNA gene clusters in the *C. reinhardtii* genome: Gene cluster is the main genomic organization of snoRNA genes in *C. reinhardtii*. Besides 80 singletons, 242 snoRNA genes are arranged into 76 gene clusters, with a <500-bp interval between 2 adjacent snoRNAs (supplemental Table S1). Remarkably, 67 singletons and 61 clusters were found to reside within introns of protein-coding genes presented on the Gene-Catalog track. Examination of the 64 snoRNA genes that were initially predicted to lie between protein-coding genes revealed that only 30 of them were truly intergenic. The 34 others were found to lie within 16 genes that had been mispredicted. In each case, more appropriate gene models were placed in the catalog. Nine of them are supported by EST and/or homology data. Altogether 76 singleton and 70 snoRNA clusters, totaling 292 (>90%) snoRNA genes are encoded in introns of 98 host genes that function mainly in translation-related processes, such as ribosomal protein, tRNA synthetase, tRNA processing protein, and elongation factor. Athough some snoRNA genes may have escaped our analysis, the *C. reinhardtii* genome contains the highest number of known intronic snoRNA clusters among eukaryotes.

The *C. reinhardtii* snoRNA gene clusters are composed of 2–7 snoRNA genes. In contrast with the land plants in which about two-thirds of the snoRNA gene clusters consist of heterologous snoRNA genes (heterocluster) (Brown *et al.* 2003a; Chen *et al.* 2003), >77% of the *C. reinhardtii* snoRNA gene clusters are made up of homologous snoRNA genes (homocluster). This is probably the result of extensive local tandem duplications. All 25 box H/ACA snoRNA gene clusters and 34 of the 42 box C/D snoRNA gene clusters are homoclusters. Only 8 box C/D snoRNA gene clusters are heteroclusters. In addition, 9 snoRNA gene clusters containing both box C/D and H/ACA snoRNAs were also identified.

Duplication and evolution of the *C. reinhardtii* snoRNA genes: The evolution of snoRNAs is thought to have occurred through a repeated series of duplications, accompanied by mutations and selection for their ability to associate into stable snoRNPs and to influence

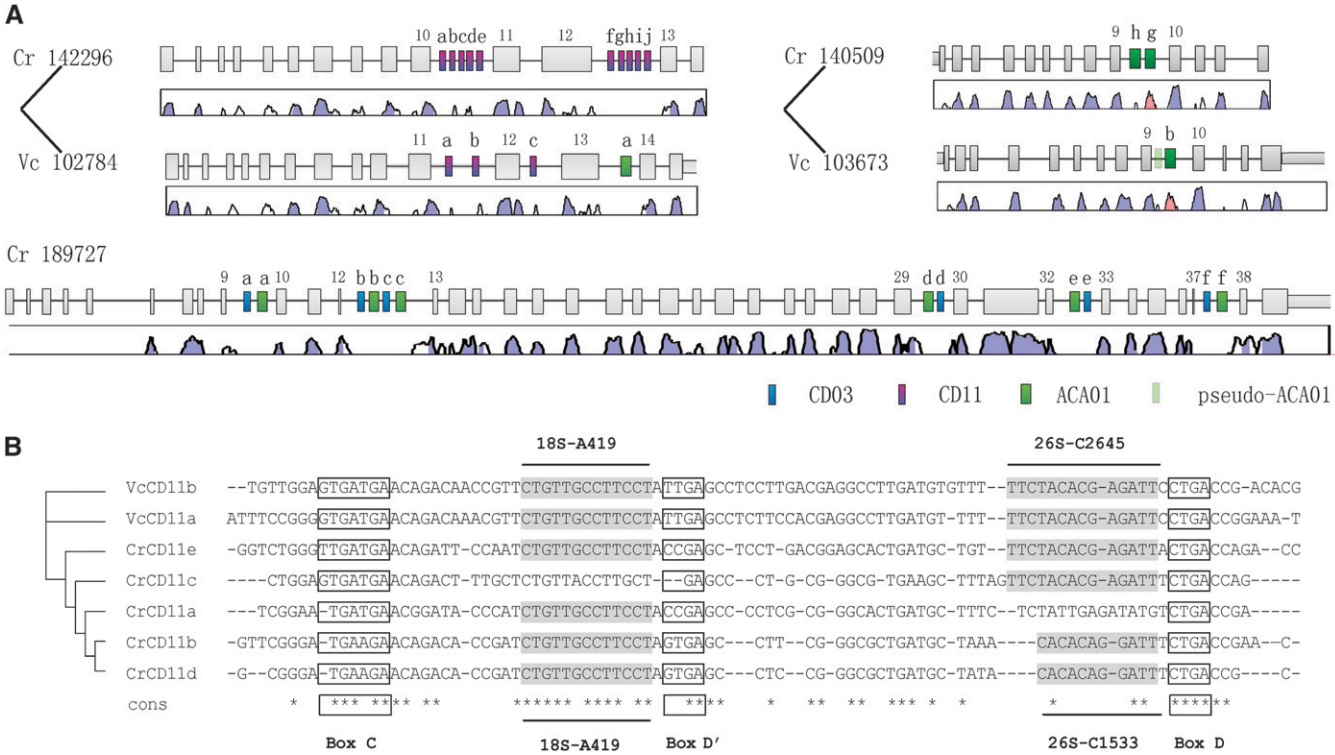

FIGURE 1.—Evolution of *C. reinhardtii* snoRNA genes. (A) Organization of snoRNA genes. Host genes from *C. reinhardtii* (*Cr*) and *V. carteri* (*Vc*) are indicated by their protein identification from the JGI genome browser. SnoRNA families are represented by different colored boxes, and isoforms are denoted by their suffixal characters (a, b, c, . . .). Exons are indicated by thick gray boxes, UTRs by intermediate-sized gray bars, and introns by black lines. The ordinals of exons flanking the snoRNAs are given. For each gene, the Vista track is attached below, with pink for conserved noncoding regions and blue for conserved coding regions. All are drawn to scale. (B) Functional evolution of CD11 snoRNA genes in *C. reinhardtii* and *V. carteri*. A rectangular cladogram tree is placed to the left of the sequence alignment. Boxes C, D′, and D are shown in open boxes, antisense sequences are shaded, and related methylated sites are indicated.

ribosome assembly and function (BROWN *et al.* 2003a). We took advantage of the clustered organization of the *C. reinhardtii* snoRNA genes to investigate their evolution following gene duplication. In most cases, despite extensive sequence divergence among snoRNA paralogs, their functions presumably remain unchanged because nucleotide differences mainly map outside of the guiding sequences. However, the function of duplicates may differ from the original one by subfunctionalization (where the snoRNA duplicates a fraction of the ancestral function) or by neofunctionalization (where the duplicate evolves to gain a novel function). One such example is the bifunctional CrCD11 family. A cluster with five CrCD11 snoRNAs (CrCluster 13) and a cluster with two VcCD11 snoRNAs reside in the corresponding introns of *C. reinhardtii–V. carteri* orthologs, Cr142296 and Vc102784 (Figure 1A). The two VcCD11 snoRNAs have the same antisense sequences as CrCD11e, which in *C. reinhardtii* can guide methylation at A419 of 18S rRNA and at C2645 of 26S rRNA (Figure 1B). This suggests that

TABLE 3

Homology information on modification sites in the *C. reinhardtii* rRNAs and snRNAs

Homologous with	PFA	PA	PF	FA	P	F	A	Non	Total
A. Methylated sites guided by the *C. reinhardtii* box C/D snoRNAs									
rRNA sites	27	11	7	1	24	2	5	19	96
snRNA sites						2		1	3
B. Pseudouridylation sites guided by the *C. reinhardtii* box H/ACA snoRNAs									
rRNA sites	6	7	3	3	9	3	6	23	60
snRNA sites							1	1	2

The modification sites of the *C. reinhardtii* rRNAs and snRNAs were compared to those of plants (P), fungi (F), and animals (A).

these two methylation sites already existed in the last common ancestor. In contrast, CrCD11c and CrCD11a have lost the 5′ and the 3′ antisense element, respectively. CrCD11b and CrCD11d have maintained the 5′ antisense element, but the 3′ antisense elements mutated so as to guide a new methylation site on *C. reinhardtii* 26S rRNA, at C1533.

Another snoRNA gene cluster with a quasi-identical sequence was found in a nearby intron of the same host gene (Figure 1A). A similar situation was encountered in 19 other host genes of *C. reinhardtii*. This suggests that snoRNAs and snoRNA clusters can be duplicated not only in the immediate vicinity, but also several hundred base pairs from their original site, into another intron of the same host gene. Cr189727 presents an interesting case of inversion of gene order (Figure 1A). The snoRNA clusters in the 9th and 37th introns are CD03-ACA01 dimers, while that in the 12th intron is made up of two CD03-ACA01 dimers. The intervening sequences between CD03 and ACA01 are well conserved in these clusters, suggesting that they were duplicated as a block. In contrast, the clusters in the 29th and 32nd introns are ACA01-CD03 dimers, *i.e.*, showing the genes in the reverse order. The intervening sequences between ACA01 and CD03 in these dimers are highly similar to that lying between CrACA01b and CrCD03c. This suggests that the CD03b-ACA01b-CD03c-ACA01c tetramer was first formed by a local tandem duplication of a CD03-ACA01 dimer, and then was duplicated twice in the 29th and 32nd intron, after which loss of the outlying snoRNAs led to the ACA01-CD03 dimers that can now be observed.

DISCUSSION

High frequency of snoRNA gene duplication in the *C. reinhardtii* genome: Owing to the similar sizes of the *V. carteri* and *C. reinhardtii* genomes (~140 Mb and 120 Mb, respectively) and the short time of the divergence between the two species, we expected to detect similar numbers of snoRNA genes in the two organisms. However, among 93 snoRNA gene families that are conserved between *C. reinhardtii* and *V. carteri*, 46 contain more members in *C. reinhardtii* than *V. carteri*, 9 in the extreme case. In contrast, only 12 families contain more members in *V. carteri* than *C. reinhardtii*. Phylogenetic analyses were performed on the 39 families that harbor ≥2 members in both *C. reinhardtii* and *V. carteri*. In 30 families, the paralogous sequences of the same genome grouped with each other, distinct from their orthologs of the other genome. This, together with the large differences in the number of family members, suggests that most of the snoRNA genes in the *C. reinhardtii* and *V. carteri* genomes have been generated by recent duplication after the lineages diverged. The alternative hypothesis, namely that snoRNA duplicates are constantly homogenized within each organism by gene conversion, is unlikely because sequence conservation was uneven, with guiding sequences markedly more conserved that the rest of the snoRNAs.

The enrichment of snoRNAs in *C. reinhardtii* may be linked to its ability to survive in a variety of environmental conditions. Although mutants defective in one or two of most rRNA modifications of *S. cerevisiae* have no detectable phenotype (Qu *et al.* 1999; Piekna-Przybylska *et al.* 2007), loss of three to all five modifications in the helix 69 of *S. cerevisiae* 26S rRNA has been reported to alter its structure in the ribosome and causes the broadest defects including reduction of amino acid incorporation rate, reduction of rRNA level, and increase of stop codon read-through activity (Liang *et al.* 2007). Interestingly, different growth rates among yeast recombinants with deletion of modification guide snoRNAs to rRNAs indicated varied sensitivity of the strains to antibiotics at different temperatures (Li *et al.* 2005; Bi *et al.* 2007; Liang *et al.* 2007). Ribosomal modification has been suggested to incur to rRNA stability and influence ribosome function and may be required in extreme environments (Omer *et al.* 2000; Barneche *et al.* 2001; Chen *et al.* 2003). The abolition of a methylation of U6 snRNA by deleting the yeast mgU6-47 snoRNA may slightly affect the efficiency of mRNA splicing (Zhou *et al.* 2002). It has been reported that 90% of the ribosomal proteins are degraded when *C. reinhardtii* cells differentiate into gametes as a result of nitrogen starvation (Siersma and Chiang 1971; Rochaix 1995). As the gametes dedifferentiate, new ribosomes will assemble. During these transformations, the large number of snoRNAs in *C. reinhardtii* may provide for the rapid buildup of an efficient rRNA modification machinery.

The ancestral organization of intronic snoRNA gene clusters in plants: Intronic snoRNA gene clusters, first reported in our study of rice Hsp70 (Qu *et al.* 1997), have since then been shown to be widespread in this organism (Liang *et al.* 2002; Chen *et al.* 2003). Their prevalence in *C. reinhardtii* suggests that they represent an ancestral characteristic of the plant lineage, distinct from the one-snoRNA-per-intron organization of vertebrates and yeasts. The small number of intronic snoRNA clusters characterized in *A. thaliana* probably relates to the smaller size of *A. thaliana* introns (~170 nt on average) (Arabidopsis Genome Initiative 2000) compared to *C. reinhardtii* (~373 nt) (Merchant *et al.* 2007) and *O. sativa* (~360 nt) (Yu *et al.* 2002). In this respect, it might be interesting to examine the practically intronless Ostreococcus genomes (Palenik *et al.* 2007) or *Physcomitrella patens* and its long introns.

Still, intronic snoRNA clusters are not exclusive to the plant lineage. A total of 17 intronic clusters encoding 77 box H/ACA snoRNAs (70% of the whole gene set) have been identified in the *Drosophila melanogaster* genome (Huang *et al.* 2005). However, the 98 box C/D snoRNAs all obey the one-snoRNA-per-intron rule, which suggests that clustering is not an ancient feature. Intronic snoRNA clusters may have evolved independently in

different lineages. Alternatively, they may have been present in ancestral eukaryotes and lost in some lineages. Further investigation of snoRNA genes in fully sequenced genomes of various taxa will help answer the question.

Whatever its origin, the intronic clustering of snoRNAs represents an economical way of producing these essential molecules. Bypassing the need for a dedicated transcription event, they make use of what is usually a discarded by-product of protein gene expression, the intron. At the same time, it may allow regulated and coordinated production of specific classes of snoRNAs, making use of the regulatory properties of the host gene. Unfortunately, nothing is known of the processing of intronic snoRNA gene clusters. Does it occur during or after intron splicing? Does it rely on endonucleolytic cleavage of the pre-mRNA or of the spliced circularized intron? Does the intron recircularize after a snoRNA is released, to allow processing of the other ones or is the maturation simultaneous for all members of the cluster? The *S. cerevisiae* polycistronic snoRNA that transcribe from a common promoter are processed by class II RNase III, Rnt1p (Qu *et al.* 1999; Ghazal *et al.* 2005). However, the AGNN stem-loop structure, the recognition signal of Rnt1p, could not be characterized in the interval of adjacent snoRNA genes in the *C. reinhardtii* snoRNA clusters. By using homology search of Rnt1p from yeast, a family with three members of RNase III genes, named AtRTL (RNase three-like) has been recently characterized in *A. thaliana* (Comella *et al.* 2007). Within this RNase III gene family, the *AtRTL2* is the closest functional homolog of the *S. cerevisiae Rnt1p* and the only gene ubiquitously expressed in *A. thaliana*. While disruption of the *AtRTL2* gene has no effect in the processing of polycistronic snoRNA in this organism, it seems that plant polycistronic snoRNAs are processed by a mechanism that does not involve RNase III-like activities. An interesting parallel can be made with the spliceosomal snRNA genes. In *C. reinhardtii*, the majority of U1, U2, and U4 snRNA genes also reside within introns of protein-coding genes (Merchant *et al.* 2007). However, their mechanism of expression is different: it has been proposed that they use a Pol II-dependent transcription start internal to the host gene, yielding transcripts that are polyadenylated and spliced and may serve as precursors for the mature snRNAs. Here also, the maturation pathway is largely unknown.

Thirty *C. reinhardtii* snoRNAs do not lie in predicted genes and have no EST nearby that could serve as evidence for a host gene. Twenty-seven of them that are grouped into six clusters were examined in greater detail; one of them could be found in the intron of a hemD pseudo gene, and two could be located in the intron of a hypothetical noncoding precursor. This suggests that they lie within specialized host genes that do not code for a protein, similar to the Drosophila and vertebrate dUhg genes (Tycowski and Steitz 2001; Huang *et al.* 2007). Comparison of the *V. carteri* and *C. reinhardtii* sequences will help elucidate the origin of these genes. One hypothesis is that they arise when a host gene loses its coding capacity while retaining its ability to splice because its introns harbor essential snoRNA genes.

We thank J.-H. Yang and L. Amar for valuable comments. This research is supported by the National Basic Research Program (2005CB724600), the National Natural Science Foundation of China (30470385 and 30570398), and the Ministry of Education of China and Guangdong province (IRT0447 and NSF-05200303). C.-L.C was a grant recipient of the French Embassy Ph.D. program and the Excellent Doctor Thesis program of Sun Yat-sen University. C.-J.C was the grant recipient of the Ph.D. program (China Scholarship Council) from the Chinese government.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Arabidopsis Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

Bachellerie, J. P., J. Cavaille and L. H. Qu, 2000 Nucleotide modifications of eukaryotic rRNAs: the world of small nucleolar RNA guides revisited, pp. 191–203 in *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions*, edited by R. A. Garrett, S. Douthwaite, A. Liljas, A. Matheson, P. B. Moore *et al.* ASM Press, Washington, DC.

Balakin, A. G., L. Smith and M. J. Fournier, 1996 The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. Cell **86:** 823–834.

Barneche, F., C. Gaspin, R. Guyot and M. Echeverria, 2001 Identification of 66 box C/D snoRNAs in Arabidopsis thaliana: extensive gene duplications generated multiple isoforms predicting new ribosomal RNA 2′-O-methylation sites. J. Mol. Biol. **311:** 57–73.

Bi, Y. Z., L. H. Qu and H. Zhou, 2007 Characterization and functional analysis of a novel double-guide C/D box snoRNA in the fission yeast. Biochem. Biophys. Res. Commun. **354:** 302–308.

Brown, J. W., M. Echeverria and L. H. Qu, 2003a Plant snoRNAs: functional evolution and new modes of gene expression. Trends Plant Sci. **8:** 42–49.

Brown, J. W., M. Echeverria, L. H. Qu, T. M. Lowe, J. P. Bachellerie *et al.*, 2003b Plant snoRNA database. Nucleic Acids Res. **31:** 432–435.

Cavaille, J., K. Buiting, M. Kiefmann, M. Lalande, C. I. Brannan *et al.*, 2000 Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. USA **97:** 14311–14316.

Chen, C. L., D. Liang, H. Zhou, M. Zhuo, Y. Q. Chen *et al.*, 2003 The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from Oryza sativa. Nucleic Acids Res. **31:** 2601–2613.

Clouet d'Orval, B., M. L. Bortolin, C. Gaspin and J. P. Bachellerie, 2001 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:** 4518–4529.

Coleman, A. W., 1999 Phylogenetic analysis of "Volvocacae" for comparative genetic studies. Proc. Natl. Acad. Sci. USA **96:** 13892–13897.

Comella, P., F. Pontvianne, S. Lahmy, F. Vignols, N. Barbezier *et al.*, 2007 Characterization of a ribonuclease III-like protein required for cleavage of the pre-rRNA in the 3′ETS in Arabidopsis. Nucleic Acids Res. **36:** 1163–1175.

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

Ganot, P., M. L. Bortolin and T. Kiss, 1997 Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell **89:** 799–809.

Ganot, P., B. E. Jady, M. L. Bortolin, X. Darzacq and T. Kiss, 1999 Nucleolar factors direct the 2′-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. Mol. Cell. Biol. **19:** 6906–6917.

Ghazal, G., D. Ge, J. Gervais-Bird, J. Gagnon and S. Abou Elela, 2005 Genome-wide prediction and analysis of yeast RNase III-dependent snoRNA processing signals. Mol. Cell. Biol. **25:** 2981–2994.

Huang, Z. P., C. J. Chen, H. Zhou, B. B. Li and L. H. Qu, 2007 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. Genomics **89:** 490–501.

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

Huttenhofer, A., M. Kiefmann, S. Meier-Ewert, J. O'Brien, H. Lehrach *et al.*, 2001 RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J. **20:** 2943–2953.

Kishore, S., and S. Stamm, 2006 The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science **311:** 230–232.

Kiss, T., 2001 Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J. **20:** 3617–3622.

Kiss, T., 2004 Biogenesis of small nuclear RNPs. J. Cell Sci. **117:** 5949–5951.

Kiss, T., and W. Filipowicz, 1995 Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. Genes Dev. **9:** 1411–1424.

Leader, D. J., G. P. Clark, J. Watters, A. F. Beven, P. J. Shaw *et al.*, 1997 Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-snoRNAs. EMBO J. **16:** 5742–5751.

Lestrade, L., and M. J. Weber, 2006 snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res. **34:** D158–D162.

Li, J. B., J. M. Gerdes, C. J. Haycraft, Y. Fan, T. M. Teslovich *et al.*, 2004 Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell **117:** 541–552.

Li, S. G., H. Zhou, Y. P. Luo, P. Zhang and L. H. Qu, 2005 Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from Schizosaccharomyces pombe. J. Biol. Chem. **280:** 16446–16455.

Liang, D., H. Zhou, P. Zhang, Y. Q. Chen, X. Chen *et al.*, 2002 A novel gene organization: intronic snoRNA gene clusters from Oryza sativa. Nucleic Acids Res. **30:** 3262–3272.

Liang, X. H., Q. Liu and M. J. Fournier, 2007 rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. Mol. Cell **28:** 965–977.

Lowe, T. M., and S. R. Eddy, 1999 A computational screen for methylation guide snoRNAs in yeast. Science **283:** 1168–1171.

Mathews, D. H., J. Sabina, M. Zuker and D. H. Turner, 1999 Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. **288:** 911–940.

Maxwell, E. S., and M. J. Fournier, 1995 The small nucleolar RNAs. Annu. Rev. Biochem. **64:** 897–934.

Mayor, C., M. Brudno, J. R. Schwartz, A. Poliakov, E. M. Rubin *et al.*, 2000 VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics **16:** 1046–1047.

Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz *et al.*, 2007 The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science **318:** 245–250.

Ni, J., A. L. Tien and M. J. Fournier, 1997 Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell **89:** 565–573.

Notredame, C., D. G. Higgins and J. Heringa, 2000 T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302:** 205–217.

Omer, A. D., T. M. Lowe, A. G. Russell, H. Ebhardt, S. R. Eddy *et al.*, 2000 Homologs of small nucleolar RNAs in Archaea. Science **288:** 517–522.

Palenik, B., J. Grimwood, A. Aerts, P. Rouze, A. Salamov *et al.*, 2007 The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. USA **104:** 7705–7710.

Piekna-Przybylska, D., W. A. Decatur and M. J. Fournier, 2007 New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. RNA **13:** 305–312.

Qu, L. H., A. Henras, Y. J. Lu, H. Zhou, W. X. Zhou *et al.*, 1999 Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. Mol. Cell. Biol. **19:** 1144–1158.

Qu, L. H., L. Zhong, S. H. Shi, Y. J. Lu, R. Fang *et al.*, 1997 Two snoRNAs are encoded in the first intron of the rice hsp70 gene. Prog. Nat. Sci. **7:** 371–377.

Rochaix, J. D., 1995 Chlamydomonas reinhardtii as the photosynthetic yeast. Annu. Rev. Genet. **29:** 209–230.

Schattner, P., S. Barberan-Soler and T. M. Lowe, 2006 A computational screen for mammalian pseudouridylation guide H/ACA RNAs. RNA **12:** 15–25.

Siersma, P. W., and K. S. Chiang, 1971 Conservation and degradation of cytoplasmic and chloroplast ribosomes in Chlamydomonas reinhardtii. J. Mol. Biol. **58:** 167–185.

Smith, C. M., and J. A. Steitz, 1997 Sno storm in the nucleolus: new roles for myriad small RNPs. Cell **89:** 669–672.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 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:** 4673–4680.

Tycowski, K. T., M. D. Shu and J. A. Steitz, 1993 A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. Genes Dev. **7:** 1176–1190.

Tycowski, K. T., Z. H. You, P. J. Graham and J. A. Steitz, 1998 Modification of U6 spliceosomal RNA is guided by other small RNAs. Mol. Cell **2:** 629–638.

Tycowski, K. T., and J. A. Steitz, 2001 Non-coding snoRNA host genes in Drosophila: expression strategies for modification guide snoRNAs. Eur. J. Cell. Biol. **80:** 119–125.

Venema, J., and D. Tollervey, 1999 Ribosome synthesis in Saccharomyces cerevisiae. Annu. Rev. Genet. **33:** 261–311.

Yu, J., S. Hu, J. Wang, G. K. Wong, S. Li *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. *ssp. indica*). Science **296:** 79–92.

Zemann, A., A. op de Bekke, M. Kiefmann, J. Brosius and J. Schmitz, 2006 Evolution of small nucleolar RNAs in nematodes. Nucleic Acids Res. **34:** 2676–2685.

Zhou, H., Y. Q. Chen, Y. P. Du and L. H. Qu, 2002 The *Schizosaccharomyces pombe* mgU6–47 gene is required for 2′-O-methylation of U6 snRNA at A41. Nucleic Acids Res. **30:** 894–902.

Communicating editor: S. Dutcher