RNA (2001), *7*:1817–1832. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S1355838201011980

Multiple snoRNA gene clusters from *Arabidopsis*

JOHN W.S. BROWN,[1] GILLIAN P. CLARK,[1] DAVID J. LEADER,[1,3] CRAIG G. SIMPSON,[1] and TODD LOWE[2]

[1]Unit of Gene Expression, Genetics Division, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, United Kingdom
[2]Department of Computer Engineering, University of California, Santa Cruz, California 95064, USA

ABSTRACT

Small nucleolar RNAs (snoRNAs) are involved in precursor ribosomal RNA (pre-rRNA) processing and rRNA base modification (2′-*O*-ribose methylation and pseudouridylation). In all eukaryotes, certain snoRNAs (e.g., *U3*) are transcribed from classical promoters. In vertebrates, the majority are encoded in introns of protein-coding genes, and are released by exonucleolytic cleavage of linearized intron lariats. In contrast, in maize and yeast, nonintronic snoRNA gene clusters are transcribed as polycistronic pre-snoRNA transcripts from which individual snoRNAs are processed. In this article, 43 clusters of snoRNA genes, an intronic snoRNA, and 10 single genes have been identified by cloning and by computer searches, giving a total of 136 snoRNA gene copies of 71 different snoRNA genes. Of these, 31 represent snoRNA genes novel to plants. A cluster of four *U14* snoRNA genes and two clusters containing five different snoRNA genes (*U31*, *snoR4*, *U33*, *U51*, and *snoR5*) from *Arabidopsis* have been isolated and characterized. Of these genes, *snoR4* is a novel box C/D snoRNA that has the potential to base pair with the 3′ end of 5.8S rRNA and *snoR5* is a box H/ACA snoRNA gene. In addition, 42 putative sites of 2′-*O*-ribose methylation in plant 5.8S, 18S, and 25S rRNAs have been mapped by primer extension analysis, including eight sites novel to plant rRNAs. The results clearly show that, in plants, the most common gene organization is polycistronic and that over a third of predicted and mapped methylation sites are novel to plant rRNAs. The variation in this organization among gene clusters highlights mechanisms of snoRNA evolution.

Keywords: gene clustering; pre-rRNA processing; ribose methylation; snoRNA genes

INTRODUCTION

Ribosomal RNA (rRNA) transcription, processing, and ribosome assembly occur in the nucleolus. Cleavage and modification of precursor rRNA (pre-rRNA) are catalyzed by *trans*-acting factors including small nucleolar ribonucleoprotein particles (snoRNPs). Small nucleolar RNAs (snoRNAs) are the RNA component of snoRNPs. Some snoRNAs, including the most abundant, *U3* and *U14*, are required for processing of the pre-rRNA into 18S, 5.8S, and 28S rRNA subunits (Maxwell & Fournier, 1995; Venema & Tollervey, 1999). However, the majority of snoRNAs are involved in 2′-*O*-ribose methylation and pseudouridylation modification of rRNAs. In higher eukaryotes, there are around 100 of each type of modification. All of the snoRNAs characterized to date, with the exception of RNase MRP, fall into two major classes: box C/D and box H/ACA snoRNAs (Maxwell & Fournier, 1995; Balakin et al., 1996; Weinstein & Steitz, 1999; Bachellerie et al., 2000). The box C/D snoRNAs are characterized by well-conserved boxes C (AUGAUGA) and D (CUGA) near their 5′ and 3′ ends, imperfect internal copies of these features (boxes C′ and D′), a short terminal stem (3–8 nt), and most importantly, one or two internal sequences (10–21 nt) with complementarity to rRNAs (Maxwell & Fournier, 1995, Balakin et al., 1996; Bachellerie & Cavaillé, 1997; Weinstein & Steitz, 1999; Lowe & Eddy, 1999; Bachellerie et al., 2000). The regions of complementarity act as guides for site-specific 2′-*O*-ribose methylation of rRNAs (Cavaillé et al., 1996; Kiss-László et al., 1996, 1998; Nicoloso et al., 1996; Cavaillé & Bachellerie, 1998). Box H/ACA snoRNAs are characterized by the presence of stem-loop structures in the 5′ and 3′ halves of the molecules, the conserved box H sequence (ANANNA) between the two stem-loops, and the box ACA (ACANNN) at the 3′ end (Balakin et al., 1996; Ganot et al., 1997a). These snoRNAs are involved in pseudouridylation of rRNAs

Reprint requests to: Dr. John W.S. Brown, Unit of Gene Expression, Genetics Division, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, United Kingdom; e-mail: jbrown@scri.sari.ac.uk, or Dr. Todd Lowe, Department of Computer Engineering, 227 sinsheimer Labs, University of California, Santa Cruz, California 95064, USA; e-mail: lowe@soe.ucsc.edu.
[3]Present address: Syngenta, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom.

(Ganot et al., 1997b; Ni et al., 1997; Bortolin et al., 1999).

The majority of rRNA modification sites have been mapped to specific nucleotide positions for a number of eukaryotes, including human, *Xenopus*, mouse, and the yeast *Saccharomyces* (Maden, 1990; Ofengand & Fournier, 1998). Yeast contain about half as many rRNA modification sites compared to mammals, though the majority of these are conserved at precisely the same positions as mammalian modifications (Maden, 1990; Ofengand & Fournier, 1998). Across phyla, conserved sites are guided by snoRNAs that are usually similar only in their box features and complementary sequence. Virtually all of the box C/D snoRNAs that determine the 55 2′-*O*-ribose methylation sites in yeast rRNAs (Lowe & Eddy, 1999; Samarsky & Fournier, 1999) and >60 human box C/D snoRNAs have been identified (Kiss-László et al., 1996; Nicoloso et al., 1996; Smith & Steitz, 1998). Ribose methylation sites and cognate snoRNAs have also been characterized in trypanosomes (Dunbar et al., 2000) and archae (Gaspin et al., 2000; Omer et al., 2000). The number of conserved modification sites declines with evolutionary distance, with only a small handful of methylation sites conserved between eukaryotes and Archae. However, for all species examined, the mechanism appears to be well conserved: within the snoRNA/target RNA duplex, the target of methylation is invariably paired to the fifth nucleotide upstream of the D or D′ sequences. Information on plant rRNA methylation sites comes from early biochemical analyses comparing the methylation profiles of plant and animal rRNAs by determining the composition of alkali-resistant di- and trinucleotides (Lane, 1965; Lau et al., 1974). Differences were observed, particularly in the higher number of alkali-resistant trinucleotides in plant 25S rRNA. Recently, the first methylation sites to be mapped in plants were reported (Barneche et al., 2000; Qu et al., 2001).

Different modes of expression of snoRNAs are utilized among eukaryotes. Vertebrates, plants, and yeast all contain snoRNAs that are transcribed from their own promoters (e.g., *U3* and RNase MRP). However, the majority of vertebrate snoRNAs and seven yeast snoRNAs are encoded within introns of protein-coding genes. In mammals, there are several examples of "host" genes that harbor snoRNAs in multiple introns; when spliced, the host mRNAs contain multiple stop codons and are rapidly degraded, indicating they do not code for proteins (Tycowski et al., 1996, Bortolin & Kiss, 1998; Pelczar & Filipowicz, 1998; Smith & Steitz, 1998; Tycowski & Steitz, 2001). The third mode of expression of snoRNAs is within polycistronic arrays of genes. The first example of such an organization was the dicistronic *snR190/U14* genes in yeast. The genes are transcribed together and subsequently processed (Chanfreau et al., 1998b; Petfalski et al., 1998). Yeast is now known to contain a total of five polycistronic snoRNA gene arrangements: two dicistrons, two tricistrons, and a heptacistron (Lowe & Eddy, 1999; Qu et al., 1999; Samarsky & Fournier, 1999). Polycistronic snoRNAs have also been found in trypanosomes (Dunbar et al., 2000). Polycistronic organization and transcription in plants was demonstrated initially for *U14* genes in maize, barley, and potato (Leader et al., 1994). Detailed analysis of *U14* gene clusters in maize found additional box C/D and box H/ACA snoRNA genes both upstream and downstream of the *U14* genes, forming gene clusters containing five and seven snoRNAs that are expressed polycistronically from upstream promoters (Leader et al., 1997). In addition to nonintronic polycistrons, a cluster of six snoRNAs have been detected in the intron of a rice heat shock protein gene, *hsp70* (Qu et al., 1997), and two snoRNAs were found within a single intron of fibrillarin genes in *Arabidopsis* (Barneche et al., 2000). More recently, a computer-assisted search of the *Arabidopsis* genome sequence identified 10 gene clusters containing 15 different snoRNA gene types, suggesting that polycistronic organization is the most common in plants (Qu et al., 2001).

The different gene organizations seen in eukaryotes reflect different processing pathways used to produce mature snoRNAs from precursor snoRNA transcripts. In vertebrates and yeast, at most one snoRNA is encoded in any particular intron, consistent with largely splicing-dependent processing by exonucleases from linearized, debranched intron lariats (Kiss & Filipowicz, 1995; Cavaillé & Bachellerie, 1996; Ooi et al., 1998; Petfalski et al., 1998). However, there are examples of snoRNAs that appear to be processed independently of splicing: *U16* and *U18* in *Xenopus* and U18 in yeast (Caffarelli et al., 1996; Villa et al., 1998). The intronic polycistron found in the intron of rice *hsp70* is in contrast to the one-snoRNA-per-intron organization of vertebrates and yeast. However, successful processing of single and multiple U14 snoRNAs from both intronic and nonintronic constructs demonstrated that processing of those plant snoRNAs studied to date is splicing independent (Leader et al., 1997, 1999). Processing of yeast di- and polycistronic pre-snoRNAs involves RNase III, which cuts in duplex RNA formed between the snoRNAs (Chanfreau et al., 1998a, 1998b; Qu et al., 1999). Trimming of the 5′ and 3′ flanking sequences involves the 5′ → 3′ exonucleases, Xrn1p and Rat1p, and the 3′ → 5′ exonuclease, Rrp6p (Petfalski et al., 1998; Qu et al., 1999; Venema & Tollervey, 1999). Although plant RNase III and Xrn1 orthologs have been identified in *Arabidopsis* (Kastenmayer & Green, 2000), their role in pre-snoRNA processing has not yet been defined.

The aims of this research were to determine how widespread snoRNA gene clustering is in plants, to identify novel plant snoRNAs and to examine the major differences in organization between plant snoRNAs genes and those of yeast and vertebrates. To this

end, we have isolated gene clusters from *Arabidopsis* by genomic library screening, and have identified other individual snoRNA genes and gene clusters by computer-based screens. In this article, we describe the molecular characterization of cloned *Arabidopsis* snoRNA gene clusters encoding *U14* snoRNAs and genes related to the rice *hsp70* intronic cluster. In addition, we describe other gene clusters identified by computer searches and relate these to mapping of 2′-*O*-ribose methylation sites. Many of the identified genes are present in more than one copy and gene variants can differ substantially. The evolution of snoRNA gene organization in plants is discussed on the basis of the genomic organization of the *Arabidopsis* snoRNA gene clusters, and differences in organization and location of related genes in different plant species.

RESULTS

Isolation and characterization of *Arabidopsis* U14- and U51-containing genomic clones

Following screening of an *Arabidopsis* genomic library, two *U14* clones and one *U51* clone were plaque purified, the phage DNA characterized, and hybridizing fragments cloned and completely sequenced. Sequencing of λAtU14.1 and λAtU14.3 showed them to be identical and to contain a cluster of four *U14* genes (Fig. 1A). The four genes were contained in approximately 800 bp. All of the *U14* genes contained box C and D sequences, terminal stems of varying sequence and length, and sequences complementary to 18S rRNA, one of which specifies methylation at Cm416 (Fig. 2A).

Two further gene clusters, each containing variants of five different snoRNA genes (*U31*, *snoR4*, *U33*, *U51*, and *snoR5*) were isolated. Each cluster contained four box C/D snoRNA genes and one box H/ACA snoRNA gene (*snoR5*) and are linked, lying around 7.8 kb apart (Fig. 1B). *U33*, *U51*, and *snoR5* are the *Arabidopsis* homologs of the genes *C*, *D*, and *F* first identified in the rice *hsp70* cluster (Fig. 1C; Qu et al., 1997). Three of the box C/D snoRNAs have regions of complementarity to 18S or 25S rRNAs related to vertebrate *U31*, *U33*, and *U51* (Fig. 2A). These snoRNAs would determine ribose methylation at Gm 2610 (25S), Um 1270 (18S), and Am 814 (25S), respectively, in *Arabidopsis*. *SnoR5* is a box H/ACA snoRNA with two pseudouridylation pockets that would determine pseudouridylation at two sites in 18S rRNA (Ψ 1000 and Ψ 1118), both conserved in human and yeast 18S. *SnoR5* is related to yeast snR31 in guiding modification at an equivalent position-Ψ 998 (yeast) and Ψ 1000 (*Arabidopsis*). However, the former is determined by the 3′ loop of snR31 but the latter by the 5′ loop of *snoR5* (Fig. 2B).

SnoR4 is a novel snoRNA with a region complementary to the 3′ end of 5.8S rRNA that potentially could guide methylation at position 155 in 5.85 rRNA (Fig. 3A). However, it is known that this site is unmethylated (Lau et al., 1974). The region of complementarity covers the 3′ end of 5.8S rRNA, and overlaps the region of 5.8S rRNA that base pairs with the 5′ end of 25S rRNA (Fig. 3A,B). The formation of this ITS2-proximal stem is essential for pre-rRNA processing in yeast (Peculis & Greer, 1998; Côté & Peculis, 2001), suggesting that snoR4 may have a different role in pre-rRNA processing.

Expression of cloned snoRNAs in *Arabidopsis*

Although three of the box C/D snoRNAs are functional homologs of vertebrate snoRNAs, *snoR4* is unique to

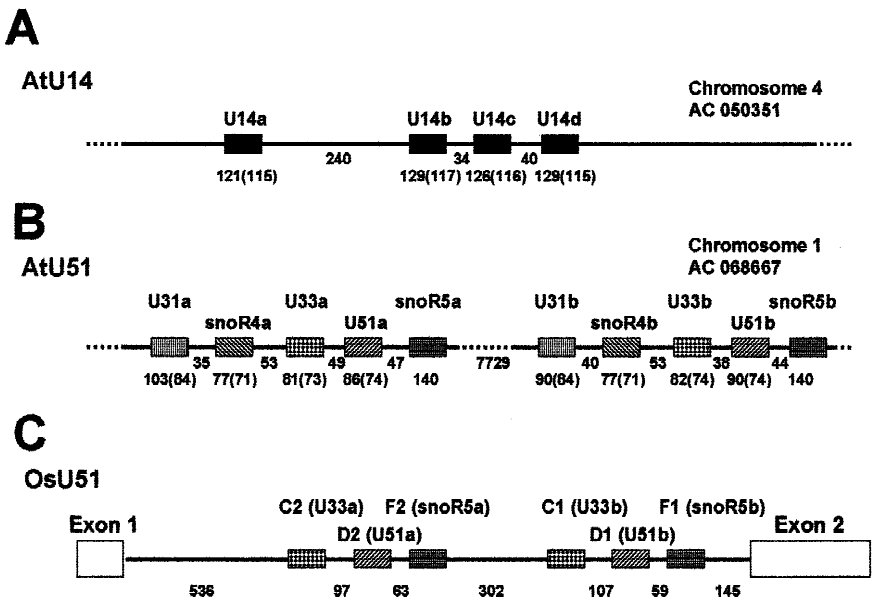

FIGURE 1. Organization of cloned plant snoRNA gene clusters. Schematic representation of the genomic organization of characterized gene clusters. **A**: *Arabidopsis U14* gene clusters. **B**: *Arabidopsis U51*-containing clusters. **C**: Rice intronic *U51*-containing clusters. Genes are shown by variously shaded boxes; exons in OsU51.1 *hsp70* are represented by large open boxes (not to scale); intergenic sequences are given and the lengths of genes are given with the length minus terminal stems in brackets. The chromosomal location and GenBank accession number are given for *Arabidopsis* gene clusters.

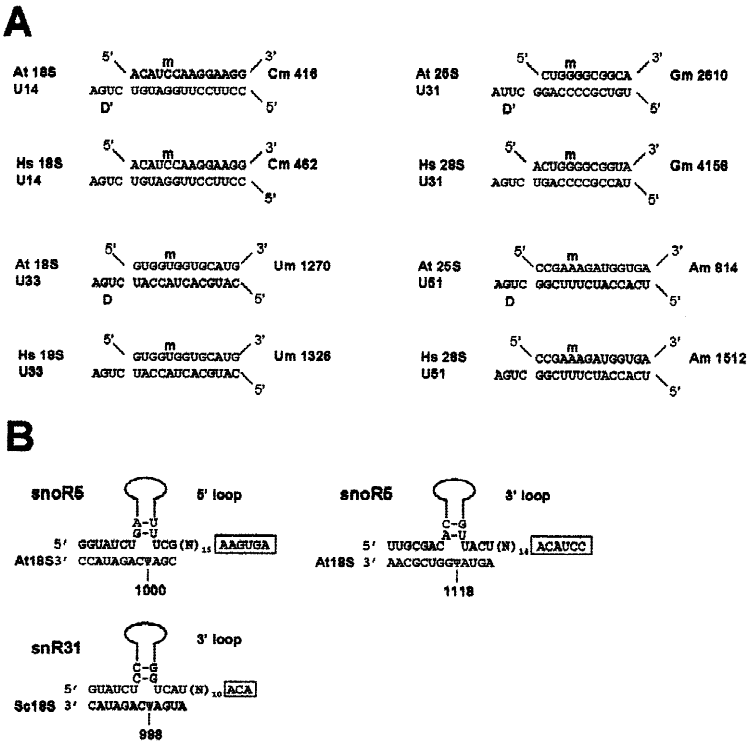

FIGURE 2. Complementary regions of cloned snoRNA genes to plant 18S and 25S rRNAs. **A**: Potential base-pairing interactions between complementary regions of isolated box C/D snoRNAs and *Arabidopsis* (At) 18S and 25S rRNAs are compared to the equivalent interactions between human snoRNAs and rRNAs (Kiss-László et al., 1996, Nicoloso et al., 1996). The putative sites of plant rRNA ribose methylation and known sites of human rRNA methylation (Maden, 1990) are marked with an m, and D or D′ boxes are indicated. **B**: Potential base-pairing interactions of the box H/ACA gene, snoR5, with *Arabidopsis* 18S rRNA are compared to that of yeast snR31. The pseudouridylated bases in the rRNA are shown (Ψ). Numbering of human and yeast rRNAs is from GenBank accession number U13369 and Samarsky and Fournier (1999), respectively.

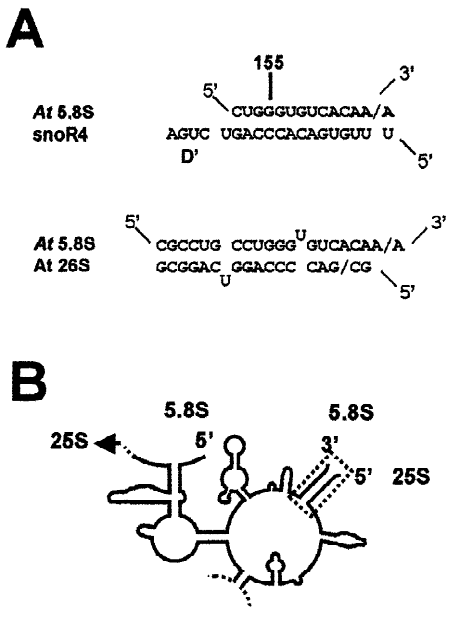

FIGURE 3. Putative base-pairing interaction of snoR4 with 5.8S rRNA. **A**: A 14-bp complementary region of snoR4 can base pair with the 3′ end of 5.8S rRNA, which is known to be unmethylated. The 3′ end of mature 5.8S rRNA is indicated by /. An extended sequence of the 3′ region of 5.8S rRNA can also base pair with the 5′ end region of 25S rRNA to form the ITS2-5.8S proximal stem (Peculis & Greer, 1998). **B**: Partial predicted secondary structure of 25S and 5.8S rRNA of *Arabidopsis* showing the base-pairing interactions found in pre-rRNA processing and ribosomes. The ITS2-5.8S proximal stem is boxed.

plants and does not appear to be involved in methylation. To demonstrate that *snoR4* is an expressed gene, expression of all four box C/D snoRNAs was investigated by RNase A/T_1 protection mapping (Fig. 4). Radiolabeled, gene-specific, antisense probes were produced for each of the genes of the upstream cluster, hybridized to *Arabidopsis* total leaf RNA and treated with RNase A/T_1. All four gene probes protected products of the size expected for full-length protection products: U31: 87–92 nt (Fig. 4, lane 1), snoR4: 77–81 nt (Fig. 4, lane 2), U33: 80–84 nt (Fig. 4, lane 3), and U51: 82–83 nt (Fig. 4, lane 4). Within the limitations of the RNase A/T_1 technique, the full-length products suggest that the gene variants in the clusters are expressed in *Arabidopsis* leaf. To demonstrate that snoR4 was a nucleolar RNA, in situ hybridization was carried out on root tip preparations of *Arabidopsis* seedlings. Antisense probes to *snoR4*, *U31*, *U33*, and *U51* labeled the nucleolus showing that the snoRNAs were located and accumulated in the nucleolus (results not shown).

Mapping of 2′-*O*-ribose methylation sites in plant rRNAs

Higher eukaryotic rRNAs contain around 100 sites of 2′-*O*-ribose methylation. To begin to identify such sites

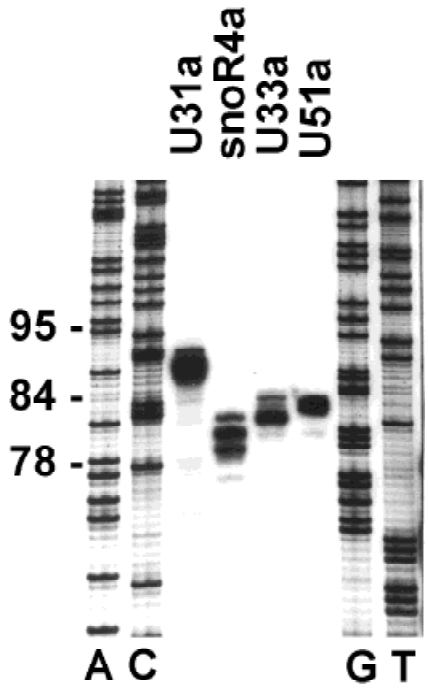

FIGURE 4. Expression of cloned box C/D snoRNAs. RNase A/T_1 protection mapping of total *Arabidopsis* leaf RNA with [^{32}P]-labeled antisense probes to U31a, snoR4a, U33a, and U51a. A, C, G, and T: DNA sequencing size markers.

in plant rRNAs, a range of rRNA primers were used in primer extension analyses with total leaf RNA from both *Arabidopsis* and maize. Mapping was carried out in both species in parallel to give greater confidence to the mapped sites. At lower concentrations of dNTPs, reverse transcription pauses at the nucleotide before the ribose methylation site, giving rise to one or more intense bands (Maden et al., 1995; Kiss-László et al., 1996). Examples of mapping gels are shown in Figure 5 that map the novel 5.8S rRNA site, Am 47 (Fig. 5A), and three 25S sites (Am 2311, Am 2316, and Cm 2327; Fig. 5B). In total, 42 putative 2′-*O*-ribose methylation sites have been mapped (Table 1). Of these sites, 15 correspond to methylation sites conserved in both vertebrates and yeast, 15 to sites found in vertebrates only, 4 to sites in yeast only, and 8 novel to plant rRNAs (Table 1). Although the mapping program was limited and did not cover all regions of the plant rRNAs, it is clear that many methylation sites are conserved with those of vertebrates and yeast, and that plant rRNAs contain a number of novel sites.

Identification of snoRNA gene clusters and genes in the *Arabidopsis* genome sequence

Using a combination of computer-assisted searches, 40 further snoRNA gene clusters of two to four genes were identified (Fig. 6). Of the 43 gene clusters described here, 10 are common to those reported recently by Qu et al. (2001) with minor differences in gene identification. The 136 genes in the 43 clusters presented here represent 71 different snoRNA genes, of which 31 are genes novel to plants. Taken with the findings of Qu et al. (2001) and Barneche et al. (2000), a total of 72 different *Arabidopsis* snoRNA genes have been identified. Of these, *snoR2* and *snoR5* are box H/ACA snoRNAs related to yeast *snR31* and *snR34*. Clearly, snoRNA gene clusters are dispersed across all chromosomes of the *Arabidopsis* genome, and one of the key features is the number of duplicated genes and gene clusters. Duplication of single genes has given rise to homogeneous clusters containing only variants of the same gene (Figs. 1A, 6A,C,H,N,R,U). In 12 cases, there are two or three related heterogeneous clusters at different sites in the genome (Fig. 6B,D,F,G,I,J,L,M,O,Q,U,W). Where heterogeneous gene clusters are present at more than one chromosomal location, there is conservation of gene order and sometimes gene number (Figs. 1B, 6B,I,L,O). However, in other cases, at one location, there has been duplication of one of the genes (Fig. 6D,G,Q), whereas in others, one site contains an additional different snoRNA gene (Fig. 6F,J). In addition, one intronic snoRNA, in an intron of a putative RNA-binding protein gene, and 10 single snoRNA genes were identified (Fig. 6X and Fig. 6V,Y, respectively).

Due to the number of *Arabidopsis* snoRNA genes with similarity to vertebrate and yeast genes, the *Arabidopsis* snoRNAs were named to maximize relationships using the following criteria. When the com-

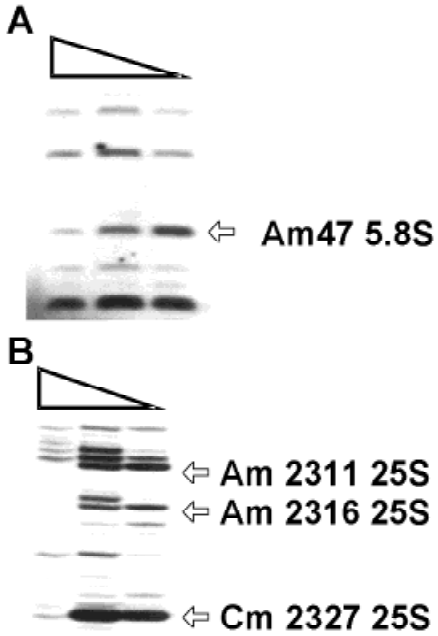

FIGURE 5. Mapping of 2′-*O*-ribose methylation sites. Primer extension analysis of *Arabidopsis* leaf RNA with end-labeled primers to regions of 5.8S (**A**) and 25S rRNAs (**B**) in decreasing concentrations of dNTPs. The methylated sites in rRNA corresponding to the bands with increasing intensity are indicated.

TABLE 1. Predicted and mapped methylation sites in *Arabidopsis* rRNAs: Plant, human and yeast snoRNA orthologs.

Predicted sites	Mapped sites	Human sites	Yeast sites	Plant snoRNA	Human snoRNA	Yeast snoRNA	Conservation of sites
5.8S							
Am47	Am47	—	—	AtsnoR9(Z41a/b)	—	—	P
Gm79	Gm79	Gm75	—	AtsnoR39BY(Z37)	Nd	—	PH
18S							
Am28		Am27	Am28	AtU27 (Z4)	U27	SnR74	PHY
Am162		Am166	—	AtsnoR18	U44	—	PH
Gm390		Gm436	—	AtsnoR30	Nd	—	PH
Cm416		Cm462	Cm414	AtU14	U14	U14	PHY
Am438	Am438	Am484	—	AtU16/AtsnoR15	U16	—	PH
Am466	Am466	Am512	—	AtsnoR17	Nd	—	PH
Cm471	Cm471	Cm517	—	AtU56	U56	—	PH
Am543		Am590	Am541	AtsnoR41Y	Nd	SnR41	PHY
Um580		Um627	Um578	AtsnoR77Y	Nd	SnR77	PHY
Gm597		Gm644	—	AtU54	U54	—	PH
Um613		—	—	AtsnoR13	—	—	P
Am621		Am668	Am619	AtU36a	U36a	SnR47	PHY
Cm795		—	—	AtsnoR25	—	—	P
Am799	Am799	—	Am796	AtsnoR53Y	—	SnR53	PY
Am975	Am975	Am1031	Am975	AtU59	U59	SnR54	PHY
Um1010		—	—	AtsnoR20.1(Z45)	—	—	P
Cm1011		—	—	AtsnoR20.2 (Z46)	—	—	P
Um1104		—	—	AtsnoR35	—	—	P
Um1232		Um1288	—	AtsnoR14	Nd	—	PH
Um1261		—	—	AtsnoR8	—	—	P
Um1263		—	—	AtsnoR32 (Z43)	—	—	P
Um1270	Um1270	Um1326	Um1267	AtU33/AtsnoR34	U33	SnR55	PHY
Gm1272	Gm1272	Gm1328	Gm1269	AtsnoR21	U32	SnR40	PHY
Um1381		Um1442	—	AtU61	U61	—	PH
Gm1431		Gm1490	Gm1427	AtsnoR19 (U25a/b)	U25	SnR56	PHY
Um1445		—	—	AtsnoR19 (U25a/b)	—	—	P
Am1575		—	—	AtsnR8	—	—	P
Cm1641		Cm1705	Cm1638	AtU43	U43	SnR70	PHY
Am1662		—	—	AtsnoR11	—	—	P
Am1754		—	—	AtsnoR23	—	—	P
25S							
Um36		—	—	AtsnoR16.2	—	—	P
Um48		—	—	AtsnoR16	—	—	P
Am660		Am1313	Am647	AtU18	U18	U18	PHY
Cm674		Cm1327	Cm661	AtsnoR58Y	Nd	SnR58	PHY
Gm812	Gm812	Gm1509	Gm803	AtsnoR39BY (Z37)	Nd	SnR39B	PHY
Am814	Am814	Am1511	Am805	AtU51	U51/U32a	SnR39/59	PHY
Am824		Am1521	Am815	AtU80 (Z15)	U80/U77	SnR60	PHY
Am883		—	Am874	AtsnoR72Y (Z2a-d)	—	SnR72	PY
Um915		Gm1612	Gm906	AtU80 (Z15)	U80	SnR60	PHY
Am943		—	—	AtsnoR12	—	—	P
Um1064		—	—	AtsnoR41Y	—	—	P
Am1140		Am1858	Am1131	AtU38	U38ab	SnR61	PHY
Am1260		—	—	AtsnoR22	—	—	P
Um1275		—	—	AtsnoR22	—	—	P
Am1372		—	—	AtsnoR7	—	—	P
Cm1439	Cm1439	Cm2338	Cm1435	AtU24	U24	U24	PHY
Um1445		—	—	AtsnoR19	—	—	P
Am1451	Am1451	Am2350	Am1447	AtU24	U76	U24	PHY
	Gm1452	Gm2351	Gm1448	AtU24	Nd	U24	PY
	Cm1510	Cm2409	—	AtU49	U49	—	PH
	Um1634	—	—	Nd	—	—	P
	Cm1840	—	—	Z42	—	—	P
Gm1845	Gm1845	—	—	AtU59	—	—	P
Cm1850	Cm1850	Cm2791	—	AtU55/AtsnoR15	U55/U39	—	PH
Am1861	Am1861	Am2802	—	AtsnoR33	Nd	—	PH
Um1882	Um1882	Um2824	Um1886	AtU34	U34	SnR62	PHY
	Um1918	—	—	Nd	—	—	P

(*continued*)

TABLE 1. *Continued.*

Predicted sites	Mapped sites	Human sites	Yeast sites	Plant snoRNA	Human snoRNA	Yeast snoRNA	Conservation of sites
25S (continued)							
Am2116		—	—	AtsnoR12 (Z44a/b)	—	—	P
Am2204	Am2204	Am3697	—	AtU37	U37	—	PH
Am2210	Am2210	Am3703	Am2218	AtU36a	U36	SnR47	PHY
Gm2226		—	—	AtU36a	—	—	P
Am2271	Am2271	Am3764	Am2279	AtU15	U15	SnR13	PHY
Gm2278	Gm2278	—	Gm2286	AtU15	—	SnR75	PY
Am2311	Am2311	Am3804	—	AtU30	U30	—	PH
Am2316	Am2316	Am3809	—	AtU79	U79	—	PH
Cm2327	Cm2327	Cm3820	Cm2335	AtU79	U74	SnR64	PHY
Am2355		Am3848	—	AtU53	U53	—	PH
Gm2404		—	—	AtsnoR29			P
Um2411		Um3904	Um2414	AtU52	U52	SnR78	PHY
Um2445		—	—	AtsnoR16.1	—	—	P
Gm2610	Gm2610	Gm4166	Gm2616	AtU31/AtsnoR35	U31	SnR67	PHY
Am2631	Am2631	—	Am2637	AtsnoR68Y	—	SnR68	PY
Um2641	Um2641	Um4197	—	AtsnoR10 (Z27a/b)	Nd	—	PH
	Gm2642	Gm4198	—	Nd	U58	—	PH
Gm2781	Gm2781	—	Gm2788	AtsnoR1	—	SnR48	PY
	Gm2783	Gm4340	Gm2790	Nd	U60	SnR48	PHY
Gm2805		—	Gm2812	AtsnoR38Y (snR38)	—	SnR38	PY
Cm2826	Cm2826	—	—	AtsnoR24	—	—	P
	Cm2847	—	—	Nd	—	—	P
	Cm2856	Cm4413	—	Nd	Nd	—	PH
Cm2869	Cm2869	Cm4426	—	AtU49/ZmU49	U49	—	PH
Am2901		—	—	AtsnoR31	—	—	P
Um2906		—	—	AtsnoR31	—	—	P
Gm2907		Gm4464	—	AtsnoR34	Nd	—	PH
Am2924		—	—	AtsnoR18	—	—	P
Am2936		Am4493	Am2943	AtU29	U29	SnR71	PHY
Cm2938		—	Cm2945	AtsnoR69Y	—	SnR69	PY
	Gm2942	—	—	Nd	—	—	P
	Um2943	Um4500	—	Nd	Nd	—	PH
Cm2949	Cm2949	Cm4506	Cm2956	AtU35	U35	SnR73	PHY
Um3289		—	—	AtsnoR13	—	—	P

Nd: methylation site but no snoRNA identified; —: no human/yeast site and no known human/yeast snoRNA.

Numbering of *Arabidopsis,* human and yeast rRNAs are based on Genbank accessions X16077 and X52320, U13369 (see Kiss-László et al., 1996), and the snoRNA database for *S. cerevisiae* methyl guide snoRNAs (http://ma.wustl.edu/eddy/snoRNAdb/Sc/Sc-snos-bysno.html). Zm: *Zea mays* (maize). P: plant, H: human, Y: yeast. Names of genes identified by Qu et al. (2001) are given in parentheses.

plementary region(s) of an identified gene corresponded to that of a vertebrate or vertebrate/yeast snoRNA, the plant snoRNA was given the vertebrate name (e.g., *U14*, *U34*, etc.). If the complementary sequence corresponded only to that of a yeast snoRNA, then it is given the yeast number followed by "Y" (e.g., *snoR77Y*). Novel plant genes were named following the nomenclature for the maize snoRNA genes, *snoR1*, *snoR2*, and *snoR3* (Leader et al., 1997), such that the novel genes presented here are called *snoR4–snoR35*. The majority of cases fell into these categories. For a small number of genes, when a plant gene contained two guide sequences found separated in different single genes in human or yeast, the gene was considered novel. Similarly, if a plant gene contained a single guide sequence where the corresponding vertebrate gene contained two, the plant gene was given a plant snoRNA name (e.g., *snoR21*). When a plant gene contained two guide sequences, one of which corresponded to a vertebrate/yeast guide sequence and the other was unique to plants, the plant gene was given the vertebrate/yeast name (e.g., *U36a*, *U59*). When more than one gene variant was found at the same locus, the genes were given the suffix a, b, c, and so forth, and when at different loci, they were given the suffix .1, .2, and so forth, or combinations thereof (Fig. 6).

The most extensive sequence variation (base substitutions, deletions, and insertions) among gene variants was found internally within the genes, as evidenced by the variation in gene lengths (Figs. 1 and 6). However, some sequence differences among variants were found in the highly conserved complementary sequences and box C, D, C′, and D′ sequences (Fig. 6A–D,G,J,L,O) and in the less conserved terminal stems.

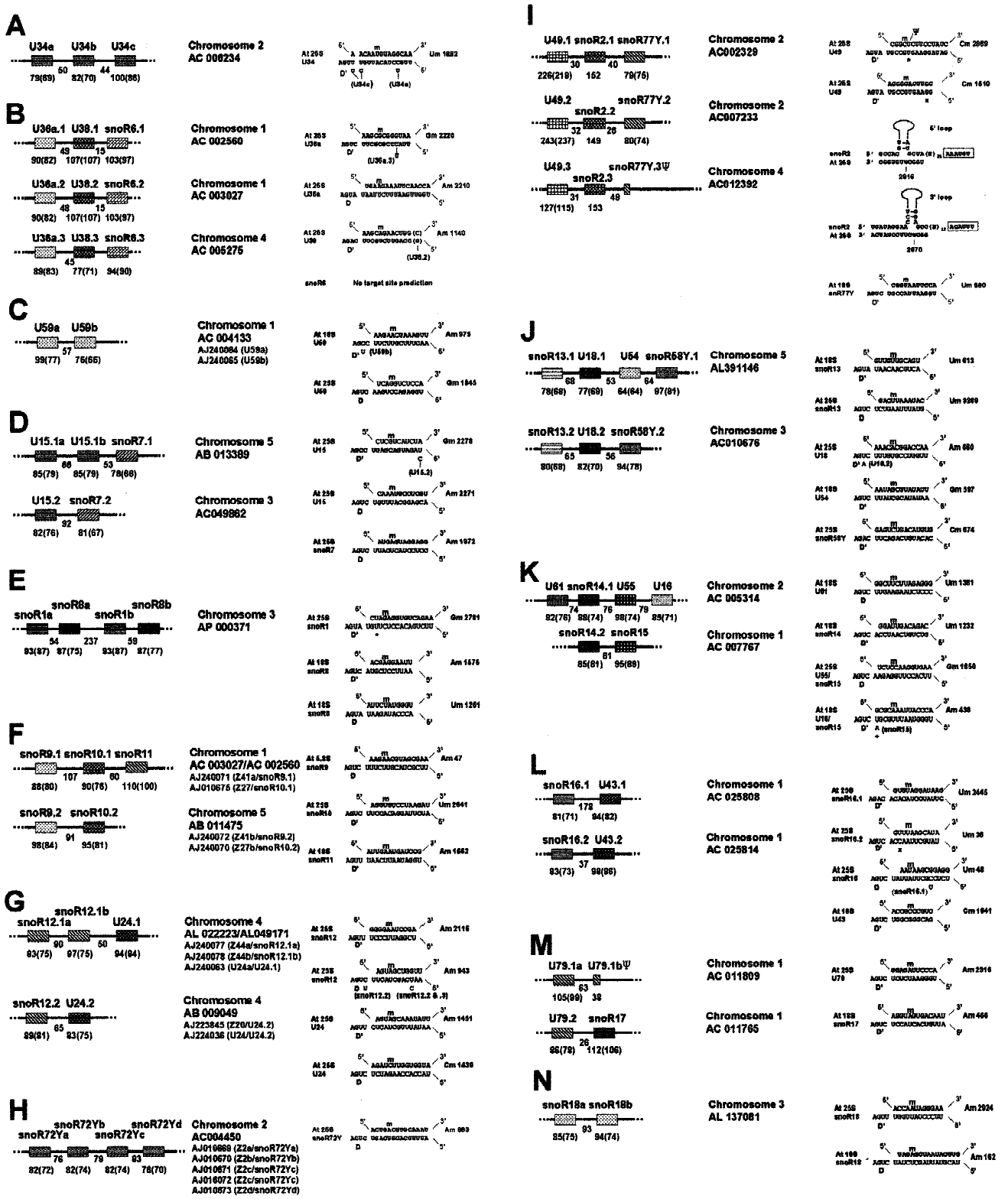

FIGURE 6. *Arabidopsis* gene clusters identified by computer assisted searches. Schematic diagrams of gene clusters giving intergenic distances and the sizes of the predicted snoRNAs with and without terminal stems, are accompanied by the chromosomal location and GenBank accession numbers of BAC clones. The accession number of snoRNA genes from Qu et al. (2001) are indicated. The complementary regions and predicted methylation sites are shown with sequence variation among variants in D or D′ boxes or complementary regions labeled. Asterisks mark conserved U-U mismatches in plant *snoR1* genes, and plant and human *U49* genes, x indicates other mismatches in complementary regions. Pseudouridines and a 3-methyl uridine (m^3) at mismatch sites are shown. **A–W**: gene clusters; **X**: putative intronic snoRNA gene; **V** and **Y**: single genes. (*Figure continues on facing page.*)

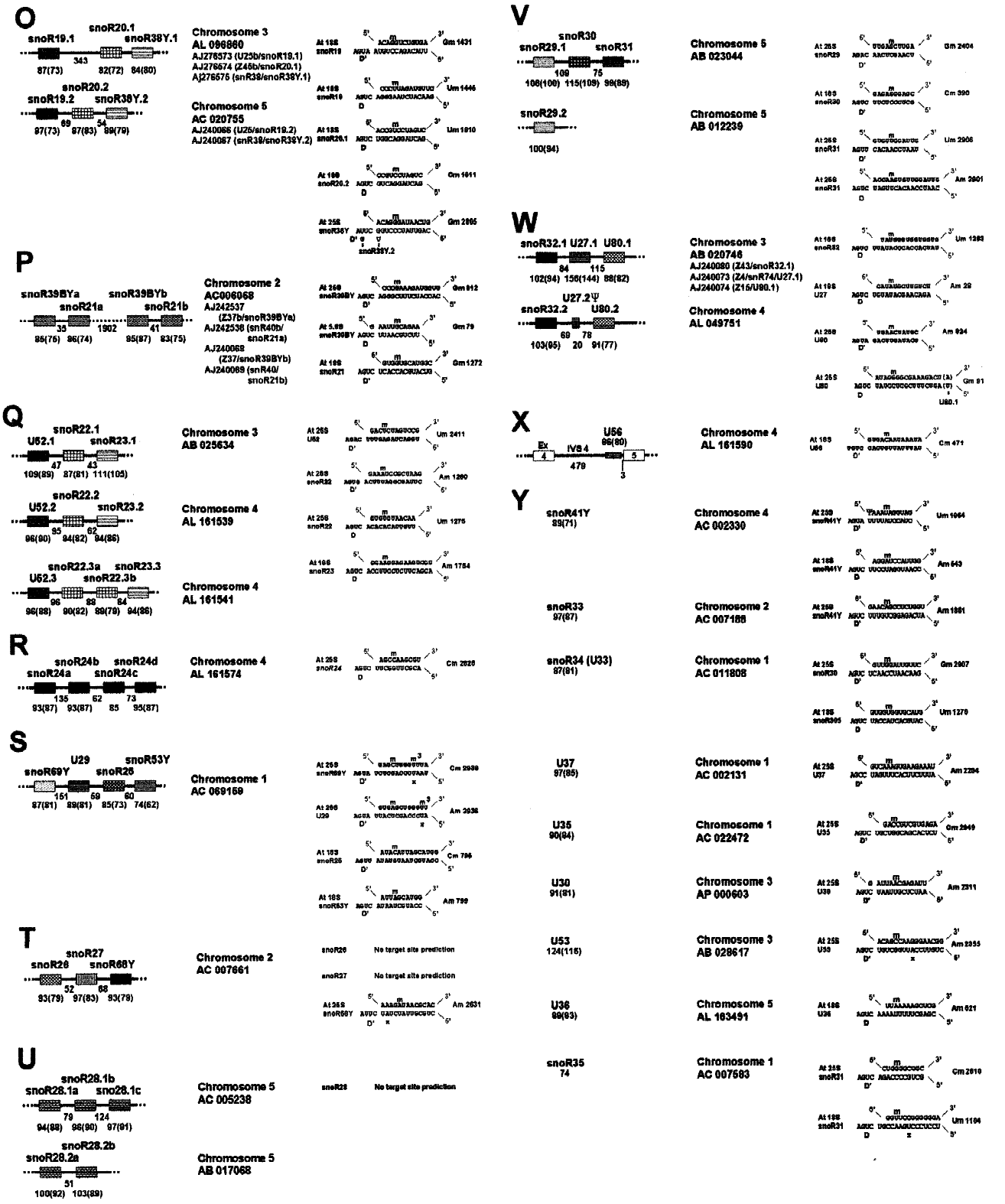

FIGURE 6. *Continued.*

Where changes occurred in the complementary sequences, internal changes often produced G:U or G:C base pairs, or led to loss or gain of base pairing at either end of the complementary region. It is unlikely that these changes would affect the function of these guide snoRNAs in base pairing with specific rRNAs sequences. However, in the snoR20 variants (Fig. 6O), sequence changes between the box D sequence and

complementary regions cause the snoRNAs to guide methylation at adjacent sites (18S: Um1010 and Cm1011), mapped by Qu et al. (2001). Clear differences in the degree of sequence variation among different genes was also observed. Some were extremely similar (e.g., *U14* and *snoR72Y*) whereas in others there were substantial differences (e.g., *U49*). Accumulation of sequence changes or rearrangements can lead to loss of function and eventually gene loss. Two clusters contained gene fragments or pseudogenes of *snoR72Y* and *U27* (Fig. 6I,M,W). For example, the pseudogene *snoR72Y.3* contained only the 5′ half of the terminal stem, box C, and 15–17 bp of gene sequence. Clearly, gene duplication and loss have occurred, but there is also evidence of gain or loss of functional sequences such as rRNA complementary sequences. For example, *U33* contains a guide sequence that corresponds to vertebrate *U33* (Fig. 2A), but there is also a single gene, *snoR34*, that contains two guide sequences, including the U33 sequence (Fig. 6Y). Similarly, the two variants of *snoR16* contain a common guide sequence (25S: Um48), but each also contains a different guide sequence (Fig. 6L). Finally, on chromosome 2, there is a cluster of four genes including *U55* and *U16*, each with a single guide sequence (Figs. 6K and 7), but on chromosome 1 there is a pair of genes, one of which, *snoR15*, contained the guide sequences of both *U16* and *U55* in the same gene (Figs. 6Y and 7). The clusters are related by the presence of *snoR14*, and alignment of the *U16* with *snoR15* gene sequences shows

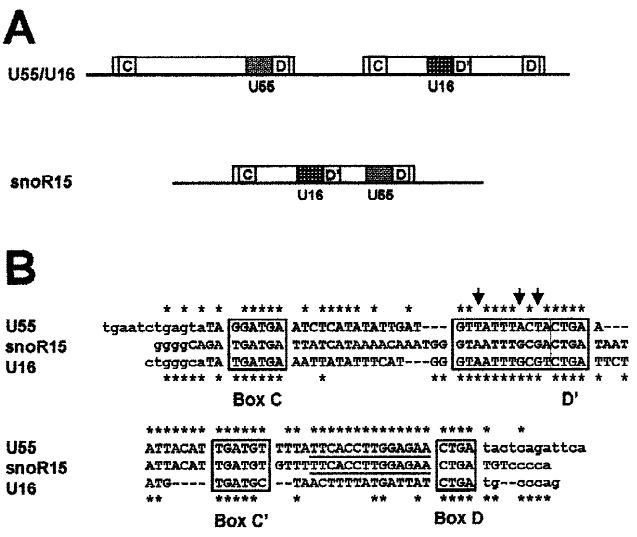

FIGURE 7. *U55*/*U16* snoRNA gene arrangements. **A**: Schematic representation of the *U55* and *U16* gene pair, and the *snoR15* gene containing complementary regions of *U55* and *U16*. **B**: Sequence alignments of *U55*, *snoR15*, and *U16*. Boxes C and D are boxed. The *U55* complementary regions adjacent to box D are underlined. Box D′, the *U16* complementary regions in *U16* and *snoR15*, and a "cryptic" *U16* complementary sequence are boxed. Arrows indicate the three mismatches in this sequence in *U55*. Asterisks indicate sequence identity between *U55* and *snoR15* (above alignment) and *snoR15* and *U16* (below alignment).

that the terminal repeats and *U16* complementary regions are conserved, whereas alignment of *U55* with *snoR15* shows the 3′ two-thirds of the genes to be very closely related. In addition, *U55* contains an internal sequence that is virtually identical to the *U16* complementary region—the sequence has a D′ box and 8 of 11 bases matching the *U16* complementary region (Fig. 7B). Due to this feature, one hypothesis is that *snoR15* was originally a duplicated pair of *U55* and *U16*, and that an intact *U16* complementary region was generated by accumulation of the three nucleotide mutations in the "cryptic" *U16* complementary region (Fig. 7B), and the second gene was lost. Alternatively, the *U16* and *U55* genes may have been originally derived from a single *snoR15* ancestor that was duplicated locally, with each duplicated gene losing one of the complementary regions due loss of selective pressure.

DISCUSSION

To examine the gene organization of plant snoRNA genes and their function, we have isolated and identified 136 box C/D snoRNA genes and have carried out an initial characterization of 2′-*O*-ribose methylation sites in *Arabidopsis* and maize rRNAs. Until now, few plant snoRNA genes have been characterized, but with the determination of the *Arabidopsis* genome sequence (The Arabidopsis Genome Initiative, 2000), it has been possible to identify putative snoRNA genes through computer-assisted searches. The identification of 43 gene clusters in *Arabidopsis* is consistent with our earlier finding of maize gene clusters consisting of multiple, different snoRNA genes. Taken with the recent report of 10 gene clusters from *Arabidopsis* (Qu et al., 2001), it is clear that plant snoRNA gene organization differs from that of yeast and vertebrates in that polycistronic genes are the most prevalent type of organization. In addition, mapping of 42 putative methylation sites demonstrates that there is conservation of many methylation sites and their respective guide sequences, but also novel, plant-specific methylation sites and snoRNA genes. Furthermore, this detailed examination of snoRNA gene and gene cluster structure provides a great deal of information on snoRNA gene evolution and mechanisms involved.

The majority of *Arabidopsis* snoRNA genes have polycistronic organization

Most plant snoRNA genes are found in polycistronic clusters. Our previous work with potato and maize snoRNA genes (Leader et al., 1994, 1997) and that of Qu et al. (2001) with one of the *Arabidopsis* clusters demonstrated the presence of polycistronic pre-snoRNA transcripts. In addition, pre-snoRNAs were detected in nucleoli and Cajal bodies, suggesting a role for the

latter in transport or processing (Shaw et al., 1998). Thus, it is highly likely that each of the *Arabidopsis* gene clusters identified here is transcribed from an upstream promoter as a polycistronic pre-snoRNA, from which individual snoRNAs are processed (Leader et al., 1997, 1998, 1999; Chanfreau et al., 1998b; Qu et al., 1999). Some evidence for polycistronic transcription of the *Arabidopsis* clusters comes from rare EST sequences that contain portions of snoRNAs. In particular, ESTs covered the 3′ half of U51, the intergenic region and snoR5 (AV544620), and U31, snoR4, U33, and part of U51 (AV552970). Although such sequences are likely to be artefacts of cDNA cloning, they suggest that the snoRNA genes are transcribed together as a polycistronic transcript.

One of the key questions in snoRNA gene organization is why the majority of plant snoRNA genes are found in polycistronic clusters with very few intronic genes, whereas vertebrates contain no polycistronic snoRNAs and the majority of genes are intronic. In general terms, polycistronic organization of functionally related genes in eukaryotes may have evolved for coordinated expression (Lawrence, 1999). In vertebrates and yeast, this has been achieved by some snoRNA genes being located in introns of genes encoding proteins involved in ribosome assembly or nucleolus formation, and by others within multiple introns of non-protein coding host genes (Tycowski et al., 1996; Bortolin & Kiss, 1998; Pelczar & Filipowicz, 1998; Smith & Steitz, 1998). However, in vertebrates and yeast, there is only a single snoRNA gene within any given intron, and they are usually processed from pre-mRNAs in a splicing-dependent manner (Kiss & Filipowicz, 1995; Maxwell & Fournier, 1995). In plants, the two snoRNAs in introns of fibrillarin genes (Barneche et al., 2000) and the intronic *U56* predicted here have this organization. However, the rice *U51*-containing gene clusters are both found in the single intron of the rice *hsp70* gene (Fig. 1C). This situation is fundamentally different from that of most vertebrate and yeast intron-encoded snoRNAs, where processing usually depends on exonucleolytic processing from linearized intron lariats (Kiss & Filipowicz, 1995; Bortolin et al., 1998). However, we have shown previously that both box C/D (*U14*) and box H/ACA (*snoR2*) single and polycistronic snoRNAs can be processed from both nonintronic or intronic transcripts and therefore that processing in plants is splicing independent (Leader et al., 1997, 1999). It is likely, therefore, that the single intronic snoRNAs and the clustered snoRNAs in the rice *hsp70* intron can be processed from either the pre-mRNA or from the spliced intron, and probably from both. Clearly, endonucleolytic cleavage of the pre-mRNA for snoRNA production would be mutually exclusive with mRNA production. Currently, there is no information on the relative efficiencies of the pathways or any advantage that one may have over the other. However, the splicing state of the cell may regulate the degree to which each pathway is utilized. The lack of dependency on splicing means that, under reduced splicing conditions (e.g., during heat shock when splicing activity is rapidly reduced and the distribution of splicing components among subnuclear structures is drastically altered; Beven et al., 1995), snoRNAs could still be produced from polycistronic or even unspliced transcripts. Thus, splicing-independent production of plant snoRNAs may have evolved to allow normal ribosome production under extreme conditions that could adversely affect the splicing competency of cells. Therefore, the key difference in snoRNA gene organization between plants and yeast/vertebrates may reflect the large variations in environmental conditions that plants generally have to tolerate.

2′-*O*-ribose methylation of plant rRNAs

Together with the 42 methylation sites mapped here, and those mapped by Qu et al. (2001) and Barneche et al. (2000), a total of 54 ribose methylation sites have now been mapped in *Arabidopsis*, of which 12 are novel to plant rRNAs. Prior to the mapping of these sites, ribose methylation in plant rRNAs has been analyzed biochemically. 2′-*O*-ribose methylation protects the phosphodiester bond from alkali hydrolysis and analysis of the composition of alkaline-resistant di- and trinucleotides showed differences between plants and other organisms. For example, yeast 5.8S is not methylated whereas human 5.8S is methylated in two positions (Um 14 and Gm 75; Maden, 1990). Wheat 5.8S rRNA contains two alkali-resistant dinucleotides, AmA and GmG, pointing to two methylation sites, the first of which is clearly different from human 5.8S (Lau et al., 1974). Both the novel site in plant 5.8S rRNA (Am 47) and the conserved site (Gm 79) have been mapped and cognate snoRNAs identified here and by Qu et al. (2001). A more notable difference in methylation pattern between plants and other organisms was the number of sites in plant 25S rRNA where adjacent nucleotides are methylated. Wheat 25S rRNA contained eight alkali-resistant trinucleotides pointing to eight positions where adjacent nucleotides are ribose methylated, NmNmN (Lane, 1965; Lau et al., 1974). In contrast, human 28S rRNA has two sites where two adjacent nucleotides are methylated and one site where three adjacent nucleotides are methylated. Yeast 25S rRNA has four sites where two adjacent nucleotides are methylated, of which two correspond to those of human 28S rRNA (Maden, 1990; Lowe & Eddy, 1999; Samarsky & Fournier, 1999). Mapping of methylation sites here identified three possible double methylation sites matching the sequences of those identified biochemically: Am1451/Gm1452/C, Um2641/Gm2642/U (UmGmU), and Gm2942/Um2943/U (GmUmU).

Different mechanisms may be involved in the methylation of adjacent nucleotides. First, each methylation site may require a different snoRNA complementary region. For example, snoR10 in plants and U58 in human are predicted to guide methylation of Um and Gm respectively in the UmGmU sequence. If plants have a U58 and human an snoR10 equivalent, then the UmGmU sites may be determined by two snoRNAs. Second, there is evidence from yeast that single snoRNAs may be responsible for methylation of adjacent nucleotides. Depletion of U24 and snR13 in yeast each resulted in the loss of a double methylation site: Am1448 and Gm1449 (Kiss-László et al., 1996) and Am2278 and Gm2279 (Lowe & Eddy, 1999), suggesting that these snoRNAs determine methylation of both sites. Third, the AmGm double methylation in yeast and plants corresponds to part of the triple methylation site in vertebrate 28S rRNA (AmGmCm). In human, U76 and U24 are predicted to guide methylation of the A and C respectively (Kiss-László et al., 1996; Smith & Steitz, 1998), such that either U76 or U24 could determine methylation of the G nucleotide. Alternatively, this unique triple methylation site has been shown previously to be methylated in vitro by a methyltransferase (Eichler & Craig, 1995). Similarly, whereas the majority of ribose methylations occur on the pre-rRNA transcript, the G residue of the conserved UmGmΨ site is methylated later in processing (Maden, 1990). Thus, in a small number of cases, methylation may occur by alternative processes other than by guide boxC/D snoRNAs. Identification of all plant box C/D snoRNAs and methylation sites will be required to examine the processes by which the increased number of adjacent nucleotide modifications found in plant 25S rRNA are produced.

Novel plant snoRNAs

The majority of box C/D snoRNAs are involved in guiding 2′-*O*-ribose methylation of rRNAs, and are also thought to have a general function as RNA chaperones in regulating folding of rRNAs (Weinstein & Steitz, 1999). SnoR4 is a box C/D snoRNA containing a sequence complementary to the 3′ end of 5.8S rRNA, which predicts methylation at position Gm 155. However, it is known from early biochemical work that this site is not methylated (see above), suggesting that snoR4 may have another function. The base-pairing interaction between snoR4 and 5.8S rRNA overlaps the region that base pairs with 25S rRNA in the ribosome that forms the ITS2-proximal stem that is essential in processing of pre-rRNAs (Peculis & Greer, 1998; Côté & Peculis, 2001). In yeast, where the pathways of pre-rRNA processing are best studied, 5.8S and 26S rRNAs are processed first by cleavage in ITS2, and the 3′ end of 5.8S rRNA is formed by 3′ to 5′ exonucleolytic processing in two stages (Venema & Tollervey, 1999). It is therefore possible that snoR4 acts as an RNA chaperone and regulates the timing of formation of the 5.8S/25S base-pairing interaction during pre-rRNA processing. Alternatively, in vertebrates, *U8* snoRNA may base pair with the 5.8S/ITS2 junction, possibly forming a substrate for cleavage or acting as an RNA chaperone for correct folding of pre-rRNA (Michot et al., 1999). Yeast does not contain *U8* and, to date, a plant *U8* has not been identified, raising the possibility that snoR4 has taken over some of the functions of *U8*.

Besides *snoR4*, there are 30 other plant-specific snoRNA genes. Of these, *snoR6*, *snoR26*, *snoR27*, and *snoR28* do not appear to have complementarity to rRNAs, snRNAs, or tRNAs. However, for *snoR6* and *snoR28*, where sequence alignments of variants are possible, there are conserved regions with adjacent D or D′ boxes that could act as complementary sequences. Of the other novel genes, two have complementary regions found in other genes (e.g., *snoR15* and *U16/U55*; *snoR16* and *U33*), whereas the rest guide novel methylation sites in the conserved core regions of rRNAs, as is the case in other organisms (Maden, 1990), with the exception of Um3289 in a nonconserved region of 25S rRNA.

Duplication and sequence variation of *Arabidopsis* snoRNA genes

Numerous single gene duplications have occurred in either homogeneous clusters where two to four copies of the same gene exist as a cluster, or in heterogeneous clusters where a single gene within a pair or more of different genes has been duplicated (Fig. 6). In addition, there are duplications of gene pairs (Fig. 6E,P), and one gene clusters of five genes giving two clusters 8 kb apart (Fig. 1B). For such duplicated clusters to be expressed, either the promoter region must be duplicated along with the structural genes, or possibly both are transcribed from the same promoter as an extended polycistronic pre-snoRNA. Currently, we have no information on the nature of the promoters driving expression of snoRNA gene clusters or of termination signals, so it is not possible to predict how the duplicated clusters are expressed. As well as relatively closely linked cluster duplications, there are examples where duplicated clusters are found in very different regions of the same chromosome or on different chromosomes. The interchromosomal locations and the long range intrachromosomal locations of duplicated gene clusters may have their origin in early polyploidization events and subsequent chromosomal rearrangements in *Arabidopsis* (Vision et al., 2000, and see below).

At the level of gene sequences, different variants exhibit different degrees of sequence variation, presumably reflecting the age of the duplication events and subsequent accumulation of mutations. Whereas in a number of cases gene variants are virtually iden-

tical, suggesting that local duplications have occurred relatively recently, many others show numerous sequence changes and small insertions and deletions. Although limited sequence variation is seen in conserved sequences (e.g., boxes C, D′, and D, and complementary sequences), the degree of variation in the rest of the molecule is far more extensive. This is consistent with internal sequences (excluding box D′ and associated complementary regions), being nonfunctional such that snoRNAs with large internal deletions are still able to form stable snoRNAs (Watkins et al., 1996; Leader et al., 1998). In some cases, mutation has led to loss of conserved sequences or gene fragments giving rise to pseudogenes or nonfunctional genes, which ultimately could lead to loss of the entire gene. For example, the pseudogenes of *snoR77Y*, *U79*, and *U27* may represent intermediates in gene loss, whereas *snoR9-11* or *snoR14/U18/U54/snoR58Y* may be examples where genes (*snoR11* and *U54*) have been completely lost from one cluster. On the other hand, mutations may alter rRNA complementary regions to produce novel sequences that could be selected, leading to establishment of novel methylation sites. For example, the two variants of *snoR20* appear to guide methylation of adjacent sites. In the *snoR68Y* cluster (Fig. 6S), two pairs of genes each guide methylation of sites lying 2 and 4 nt apart, suggesting that they have arisen through duplication and mutation. Finally, significant sequence deletions or insertions could increase the possibility of generating a novel box D or D′ and complementary region. For example, *U49.3* is around 100 nt shorter than *U49.1* and *U49.2*, due to a central deletion. Such gene reductions or expansion events could give rise to new combinations of sequence adjacent to box D or D′ sequences ultimately producing different snoRNAs guiding methylation of the same site (e.g., *U55*, *U16*, and *snoR15*; *U33* and *snoR34*).

The degree of sequence variation seen among the snoRNA gene variants reflects the fact that snoRNA genes are nonprotein coding, and so do not have the constraints of maintaining open reading frames, and that their function may be nonessential. In yeast snoRNA knockouts, none of methylation guide functions tested to date have been shown to be essential (Lowe & Eddy, 1999). Taken with the high degree of gene duplication and thereby gene redundancy, and against a background of no or low selection pressure, they are likely to be able to accumulate numerous mutations and to be evolving more rapidly than protein-coding genes. Thus, plant snoRNA genes may provide a useful model for analyzing gene evolution mechanisms.

SnoRNA gene evolution

The evolution of snoRNAs in Archae, yeast, and vertebrates is thought to have occurred through repeated series of duplication, mutation, and selection for their ability to associate into stable snoRNPs and to influence ribosome assembly and function (Lafontaine & Tollervey, 1998). Extensive analyses of snoRNAs in Archae suggest that these organisms already contained multiple snoRNA genes prior to the divergence of Archae from eukaryotes (Gaspin et al., 2000; Omer et al., 2000). Higher eukaryotic box C/D snoRNA genes can therefore be viewed as a large multigene family encoding small RNAs with related structure and function, which have evolved from a set of related ancestral genes. The key differences between plants and other higher eukaryotes are, first, the high occurrence of polycistronic clusters of genes, and, second, the number of gene and gene cluster duplications (Leader et al., 1997; Qu et al., 2001; this study).

Duplicated genes and multigene families are a feature of protein-coding genes in plants. In *Arabidopsis*, 65% of genes exist in more than one copy and 17% are arranged in tandem arrays (The Arabidopsis Genome Initiative, 2000). Much of the gene duplication in plants is explained by polyploidization, which is a major force in plant evolution with around 70% of species estimated to have undergone hybridization events at some stage (Wendel, 2000). The *Arabidopsis* genome contains large duplicated regions reflecting a number of major duplication events, including polyploidization, at different times in its evolutionary history (Blanc et al., 2000; Ku et al., 2000; Vision et al., 2000). Further rearrangements through chromosomal inversions and reciprocal translocations, and extensive gene loss have also occurred (Ku et al., 2000; Vision et al., 2000). One consequence of gene duplication is gene redundancy, such that mutations in or loss of one copy of the gene can be tolerated, or more rarely, give rise to novel gene functions. On the other hand, some mutations, such as promoter region mutations, could generate tissue specificity or altered transcript levels of different alleles that might be advantageous. Indeed, a model to explain the preservation and maintenance of duplicated genes (the DDC or duplication/divergence/complementary model), suggests that in many cases, following gene duplication, genes can diverge such that they each take on subfunctions of the original gene. Thus, they become complementary, providing the selection for fixing the duplicated versions in the genome (Force et al., 1999; Lynch & Force, 2000). Such subfunctions could be expression in different cells and tissues now shared by the duplicated genes. In the absence of allele-specific expression analysis, it is currently not possible to determine the functional redundancy or otherwise of the snoRNA gene clusters.

The organization of related gene clusters in different species show variation in gene number, order, and organization, suggesting that plant snoRNA genes have undergone substantial reorganization and transposition during the evolution of different plant species. For

example, *U33*, *U51*, and *snoR5* in the *Arabidopsis* clusters (Fig. 1B) correspond to genes *C*, *D*, and *F* in rice (Fig. 1C). However, in *Arabidopsis*, there are two more genes upstream (*U31* and *snoR4*) that are absent in the rice sequences, and the rice genes are intronic whereas those in *Arabidopsis* are not. This suggests that the initial gene order was established before the divergence of dicotyledonous and monocotyledonous plants. In maize, *snoR1*, *U49*, *snoR2*, *U14*, and *snoR3* are linked in clusters (Leader et al., 1997), but in *Arabidopsis*, *snoR1*, *U49*, and *U14* are unlinked. A complete characterization of snoRNA gene organization in *Arabidopsis* and other angiosperms (such as rice), and of lower plants will greatly enhance the understanding of how and when plants evolved polycistronic snoRNA gene organization, and how this differs within the plant kingdom.

MATERIALS AND METHODS

Isolation of genomic clones

The *U14* probe was a 450-bp *Eco*RI/*Hin*dIII fragment containing the potato *U14.1* gene (Leader et al., 1994). Rice *U33* and *U51* homologs are encoded in the first (and only) intron of the rice heat shock protein 70 cognate (*hsc70*) gene (Qu et al., 1997). The *U51* probe used to screen the library was a 370-bp fragment generated by PCR with primers OSHSC3 (5′-GTTTCGAGGAGAAGTGATGATGTCTTATCC-3′) and OSHSC4 (5′-AATGAGTCAGATACCAAATGGC-3′), and total rice leaf DNA. The PCR products were cloned and sequenced to confirm the presence of rice *U33* and *U51*, and produced sequences of two *U33*/*U51*-containing fragments, one of which was identical to the *hsc70* intron and one which differed in five nucleotide positions, suggesting that rice contains at least two clusters of *U33*/*U51* genes.

An *A. thaliana* genomic library in λFIX was obtained from the ABRC Stock Center (Ohio, USA; Stock No. CD4-8; donated to ABRC by D. Voytas). The library was screened with the *U14* and *U51* probes using standard procedures (Sambrook et al., 1989). Hybridizing fragments from positive clones were cloned into pGEM3Zf(+) (Promega) and sequenced. Sequences were analyzed using programs of the University of Wisconsin Genetics Computer Group and Daresbury Seqnet.

RNase A/T_1 protection mapping

Gene-specific probes for RNase A/T_1 protection mapping were prepared by PCR amplification of the snoRNA coding sequences and cloning into pGEM3Zf(+). Gene fragments consisting of the coding region and varying lengths of 5′ and 3′ flanking regions were amplified by PCR using the following primers:

U31-5′: 5′-CGTCTCGAGTTTTTGAGGAAAGTGATGA-3′;
U31-3′: 5′-CATAGTCGACGATTTGAGAGAATCAGAC-3′;
SnoR4-5′: 5′-GTTCTCGAGGTTGTGTATGGATATGATG-3′;
SnoR4-3′: 5′-CAGAAAGCTTAGGATCAGAAAATATGGC-3′;
U33-5′: 5′-GTTTAGATCTTTGAAAGGGAGGAATG-3′;
U33-3′: 5′-GGGTGATCAATAAAACAGAGGTCAGATGG-3′;
U51-5′: 5′-ACCCTCGAGTTAGTTGTCTGCAATCAGTG-3′;
U51-3′: 5′-GCGTTCGAAAAACTGCAATCTCAGCCG-3′.

Fragments were cloned using introduced restriction sites (underlined sequences) and confirmed by DNA sequencing. Antisense [^{32}P]-labeled RNA probes were produced by in vitro transcription with SP6 or T7 RNA polymerase following linearization of plasmids with appropriate restriction enzymes. Gel purified probes were used in RNase A/T_1 mapping as described by Goodall et al. (1990).

In situ hybridization

In situ labeling of *Arabidopsis* root tips was carried out as described previously (Boudonck et al., 1998) using gene-specific transcripts from the plasmids described above. Labeling of probes with digoxygenin, detection by cy3- or fluorescein-conjugated antibodies, confocal microscopy, and image preparation were as described previously (Beven et al., 1995).

Mapping of methylation sites rRNA

Mapping of 2′-*O*-ribose methylation was carried out by primer extension using rRNA-specific primers in three concentrations of dNTPs (0.2, 0.04, and 0.004 mM), based on a method described previously (Maden et al., 1995; Kiss-László et al., 1996). Labeled products were separated on 8% polyacrylamide-urea gels and autoradiographed.

Computer-assisted identification of putative box C/D snoRNA genes

The *Arabidopsis* genomic sequence was searched in three ways: (1) using the box C/D snoRNA search algorithm (Lowe & Eddy, 1999), (2) using complementary sequences from human, yeast, and other plant snoRNAs, and (3) using the region around putative, mapped methylation sites. Flanking sequences of putative snoRNAs were also examined for other box C/D snoRNAs. Putative snoRNA genes found by the box C/D snoRNA search algoritm are posted at http://rna.wustl.edu/snoRNAdb/Arabidopsis/ and a new *Arabidopsis* snoRNA database is currently under construction.

ACKNOWLEDGMENTS

This research was supported by grant-in-aid to the Scotish Crop Research Institute from the Scottish Executive Environment and Rural Affairs Department. We thank Dr. Tamas Kiss for identifying *U51* in the rice *hsp70* intron, Kurt Boudonck and Peter Shaw for carrying out the in situ hybridizations, and David Marshall and Chris Kelly for searching the *Arabidopsis* genomic sequence with known complementary sequences.

Received August 2, 2001; returned for revision August 30, 2001; revised manuscript received September 21, 2001

REFERENCES

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

Bachellerie JP, Cavaillé J. 1997. Guiding ribose methylation of rRNA. *Trends Biochem Sci 22*:257–261.

Bachellerie JP, Cavaillé J, Qu L-H. 2000. Nucleotide modifications of eukaryotic rRNAs: The world of small nucleolar RNA guides revisited. In: Garrett RA, Douthwaite SR, Matheson AT, Moore PB, Noller HF, eds. *The ribosome: Structure, function, antibiotics and cellular interactions.* Washington, DC: ASM Press. pp 191–203.

Balakin AG, Smith L, Fournier MJ. 1996. The RNA world of the nucleolus: Two major families of small nucleolar RNAs defined by different box elements with related functions. *Cell 86*:823–834.

Barneche F, Steinmetz F, Echeverria M. 2000. Fibrillarin genes encode both a conserved nucleolar protein and a novel small nucleolar RNA involved in ribosomal RNA methylation in *Arabidopsis thaliana*. *J Biol Chem. 275*:27212–27220.

Beven AF, Lee R, Razaz M, Leader DJ, Brown JWS, Shaw PJ. 1996. The organization of ribosomal RNA processing correlates with the distribution of nucleolar snRNAs. *J Cell Sci 109*:1241–1251.

Beven A, Simpson GG, Brown JWS, Shaw P. 1995. The organisation of spliceosomal components in the nuclei of higher plants. *J Cell Sci 108*:509–518.

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M. 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell 12*:1093–1101.

Bortolin ML, Ganot P, Kiss T. 1999. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J 18*:457–469.

Bortolin ML, Kiss T. 1998. Human U19 intron-encoded snoRNA is processed from a long primary transcript that poses little potential for protein coding. *RNA 4*:445–454.

Boudonck K, Dolan L, Shaw PJ. 1998. Coiled body numbers in the *Arabidopsis* root epidermis are regulated by cell type, developmental stage and cell cycle parameters. *J Cell Sci 111*:3687–3694.

Caffarelli E, Fatica A, Prislei S, De Gregorio E, Fragapane P, Bozzoni I. 1996. Processing of the intron-encoded U16 and U18 snoRNAs: The conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *EMBO J 15*:1121–1131.

Cavaillé J, Bachellerie JP. 1996. Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: An exonucleolytic process directed by the common terminal stem-box structure. *Biochimie 78*:443–456.

Cavaillé J, Bachellerie JP. 1998. SnoRNA guided ribose methylation of rRNA: Structural features of the guide RNA duplex influencing the extent of the reaction. *Nucleic Acid Res 26*:1576–1587.

Cavaillé J, Nicoloso M, Bachellerie JP. 1996. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature 383*:732–735.

Chanfreau G, Legrain P, Jacquier A. 1998a. Yeast RNase III as a key processing enzyme in small nucleolar RNA metabolism. *J Mol Biol 284*:975–988.

Chanfreau G, Rotondo G, Legrain P, Jacquier A. 1998b. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J 17*:3726–3737.

Côté CA, Peculis BA. 2001. Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast. *Nucleic Acids Res 29*:2106–2116.

Dunbar DA, Chen AA, Wormsley S, Baserga SJ. 2000. The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistron. *Nucleic Acids Res 28*:2855–2861.

Dunbar DA, Wormsley S, Lowe TM, Baserga SJ. 2000. Fibrillarin-associated box C/D small nucleolar RNAs in *Trypanosoma brucei*. *J Biol Chem 275*:14767–14776.

Eichler DC, Craig N. 1995. Processing of eukaryotic ribosomal RNA. *Prog Nucl Acids Res Mol Biol 49*:179–239.

Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics 151*:1531–1545.

Ganot P, Bortolin ML, Kiss T. 1997b. Site-specific pseudouridine formation in eukaryotic pre-rRNAs is guided by small nucleolar RNAs. *Cell 89*:799–809.

Ganot P, Caizergues-Ferrer M, Kiss T. 1997a. The family of box ACA snoRNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes & Dev 11*:946–972.

Gaspin C, Cavaillé J, Erauso G, Bachellerie JP. 2000. Archael homologs of eukaryotic methylation guide small nucleolar RNAs: Lessons from *Pyrococcus* genomes. *J Mol Biol 297*:895–906.

Goodall GJ, Wiebauer K, Filipowicz W. 1990. Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol 181*:148–161.

Kastenmayer JP, Green PJ. 2000. Novel features of the XRN-family in *Arabidopsis*: Evidence that AtXRN4, one of several orthologues of nuclear Xrn2p/Rat1p, functions in the cytoplasm. *Proc Natl Acad Sci USA 97*:13985–13990.

Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996. Site-specific ribose methylation of pre-ribosomal RNA: A novel function for small nucleolar RNAs. *Cell 85*:1077–1088.

Kiss-László Z, Henry Y, Kiss T. 1998. Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J 17*:797–807.

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

Ku HM, Vision T, Liu J, Tanksley SD. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proc Natl Acad Sci USA 97*:9121–9126.

Lane BG. 1965. The alkali-stable trinucleotide sequences and chain termini in 18S and 28S ribonucleates from wheat germ. *Biochemistry 4*:212–219.

Lau RY, Kennedy TD, Lane BG. 1974. Wheat embryo ribonucleates: III. Modified nucleotide constituents in each of the 5.8S, 18S and 26S ribonucleates. *Can J Biochem 52*:1110–1123.

Lawrence J. 1999. Selfish operons: The evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr Opin Genet Dev 9*:642–648.

Leader DJ, Clark GP, Boag J, Watters JA, Simpson CG, Watkins NJ, Maxwell ES, Brown JWS. 1998. Processing of vertebrate box C/D small nucleolar RNAs in plants. *Eur J Biochem 253*:154–160.

Leader DJ, Clark GP, Watters JA, Beven AF, Shaw PJ, Brown JWS. 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.

Leader DJ, Clark GP, Watters JA, Beven AF, Shaw PJ, Brown JWS. 1999. Splicing-independent processing of plant box C/D and box H/ACA small nucleolar RNAs. *Plant Mol Biol 39*:1091–1100.

Leader DJ, Sanders JF, Waugh R, Shaw PJ, Brown JWS. 1994. Molecular characterization of plant U14 small nucleolar RNA genes—closely linked genes are transcribed as polycistronic U14 transcripts. *Nucleic Acids Res 22*:5196–5203.

Lowe TM, Eddy SR. 1999. A computational screen for methylation guide snoRNAs in yeast. *Science 283*:1168–1171.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics 154*:459–473.

Maden BEH. 1990. The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol 39*:241–301.

Maden BEH, Corbett ME, Heeney PA, Pugh K, Ajuh PM. 1995. Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie 77*:22–29.

Maxwell ES, Fournier MJ. 1995. The small nucleolar RNAs. *Annu Rev Biochem 35*:897–934.

Michot B, Joseph N, Mazan S, Bachellerie JP. 1999. Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with snoRNA U8 detected by comparative analysis of new mouse sequences. *Nucleic Acids Res 27*:2271–2282.

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

Nicoloso M, Qu LH, Michot B, Bachellerie JP. 1996. Intron-encoded, antisense small nucleolar RNAs: The characterization of nine

novel species points to their direct role as guides for the 2′-*O*-ribose methylation of rRNAs. *J Mol Biol 260*:178–195.
Ofengand J, Fournier MJ. 1998 The pseudouridine residues of rRNAs: *n*umber, location, biosynthesis and function. In: Grosjean H, Benne R. eds. *Modification and editing of RNA: The alteration of RNA structure and function.* Washington, DC: ASM Press. pp 229–253.
Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP. 2000. Homologs of small nucleolar RNAs in Archae. *Science 288*:517–522.
Ooi SL, Samarsky DA, Fournier MJ, Boeke JD. 1998. Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: Intron length effect and activity of a precursor snoRNA. *RNA 4*:1096–1110.
Peculis BA, Greer CL. 1998. The structure of the ITS2-proximal stem is required for pre-rRNA processing in yeast. *RNA 4*:1610–1622.
Pelczar P, Filipowicz W. 1998. The host gene for intronic U17 small nucleolar RNAs in mammals has no protein-coding potential and is a member of the 3′ terminal oligopyrimidine gene family. *Mol Cell Biol 18*:4509–4518.
Petfalski E, Dandekar T, Henry Y, Tollervey D. 1998. Processing of the precursors to pre-snoRNAs and rRNAs requires common components. *Mol Cell Biol 18*:1181–1189.
Qu LH, Henras A, Lu YJ, Zhou H, Zhou WX, Zhu YH, Zhoa J, Henry Y, Caizergues-Ferrer M, Bachellerie JP. 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 LH, Meng Q, Zhou H, Chen YQ. 2001. Identification of 10 novel snoRNA gene clusters from *Arabidopsis thaliana*. *Nucleic Acids Res 29*:1623–1630.
Qu LH, Zhong L, Shi S, Lu Y, Zhou H, Fang R, Wang Q. 1997. Two snoRNAs are encoded in the first intron of the rice hsp70 gene. *Prog Natural Sci 7*:371–377.
Samarsky DA, Fournier MJ. 1999. A comprehensive database for the small nucleolar RNAs from *Saccharomyces cerevisiae*. *Nucleic Acids Res 27*:161–164.
Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning. A laboratory manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Shaw PJ, Beven AF, Leader DJ, Brown JWS. 1998. Localization and processing from a polycistronic precursor of novel snoRNAs in maize. *J Cell Sci 111*:2121–2128.
Smith CM, Steitz JA. 1998. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol Cell Biol 18*:6897–6909.
Tycowski KT, Shu MD, Steitz JA. 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature 379*:464–466.
Tycowski KT, Steitz JA. 2001. Non-coding snoRNA host genes in *Drosophila*: Expression strategies for modification guide snoRNAs. *Eur J Cell Biol 80*:119–125.
Venema J, Tollervey D. 1999. Ribosome synthesis in *Saccharomyces cerevisiae*. *Annu Rev Genet 33*:261–311.
Villa T, Ceradini F, Presutti C, Bozzoni I. 1998. Processing of the intron-encoded U18 small nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exonucleolytic and endonucleolytic activities. *Mol Cell Biol 18*:3376–3383.
Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science 290*:2114–2117.
Watkins NJ, Leverette RD, Xia L, Andrews MT, Maxwell ES. 1996. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA 2*:118–133.
Weinstein LB, Steitz JA. 1999. Guided tours: From precursor snoRNA to functional snoRNP. *Curr Opinion in Cell Biol 11*:378–384.
Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol Biol 42*:225–249.