doi:10.1016/j.jmb.2004.06.041 *J. Mol. Biol.* (2004) **341**, 669–683

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Different Expression Strategy: Multiple Intronic Gene Clusters of Box H/ACA snoRNA in *Drosophila melanogaster*

Zhan-Peng Huang, Hui Zhou, Dan Liang and Liang-Hu Qu*

Key Laboratory of Gene Engineering of the Ministry of Education, Biotechnology Research Center, Zhongshan University, Guangzhou 510275, People's Republic of China

**Corresponding author*

The high degree of rRNA pseudouridylation in *Drosophila melanogaster* provides a good model for studying the genomic organization, structural and functional diversity of box H/ACA small nucleolar RNAs (snoRNAs). Accounting for both conserved sequence motifs and secondary structures, we have developed a computer-assisted method for box H/ACA snoRNA searching. Ten snoRNA clusters containing 42 box H/ACA snoRNAs were identified from *D. melanogaster*. Strikingly, they are located in the introns of eight protein-coding genes. In contrast to the mode of one snoRNA per intron so far observed in all animals, our results demonstrate for the first time a novel polycistronic organization that implies a different expression strategy for a box H/ACA snoRNA gene when compared to box C/D snoRNAs in *D. melanogaster*. Mutiple isoforms of the box H/ACA snoRNAs, from which most clusters are made up, were observed in *D. melanogaster*. The degree of sequence similarity between the isoforms varies from 99% to 70%, implying duplication events in different periods and a trend of enlarging the intronic snoRNA clusters. The variation in the functional elements of the isoforms could lead to partial alternation of snoRNA's function in loss or gain of rRNA complementary sequences and probably contributes to the great diversity of rRNA pseudouridylation in *D. melanogaster*.

Keywords: box H/ACA snoRNA; intronic snoRNA gene cluster; rRNA pseudouridylation; *Drosophila melanogaster*

Introduction

In eukaryotes, a myriad of small nucleolar RNAs (snoRNAs) are enriched in the nucleoli in the form of small nucleolar ribonucleoprotein particles (snoRNPs).[1,2] Except for RNase MRP, all snoRNAs fall into two major families, i.e. box C/D and box H/ACA snoRNAs, on the basis of common sequence motifs and structural features.[3] A large number of snoRNAs characterized so far are box C/D snoRNAs that share two conserved motifs, the 5′ end box C (RUGAUGA) and the 3′ end box D (CUGA); whereas the box H/ACA snoRNAs exhibit a common hairpin-hinge-hairpin-tail secondary structure with the H (ANANNA) motif in the hinge region and an ACA triplet that is always three nucleotides from the 3′ end of the molecule.[4] Several snoRNAs, such as U3, snR30 and RNase MRP, are required for specific cleavage of pre-rRNAs.[1] However, the majority of box C/D snoRNAs function as guides for site-specific 2′-*O*-ribose methylation and most box H/ACA snoRNAs as guides for pseudouridylation in the post-transcriptional processing of diverse RNAs, such as rRNAs,[5,6] snRNAs[1,7] and even tRNA.[8]

During the last decade, research on snoRNAs has provided new, often quite unexpected, information about their genomic locations that has challenged some established principles on eukaryotic gene organization and expression. So far as we know, the genomic organization of snoRNAs has shown greatest diversity among eukaryotes. In the budding yeast *Sacharomyces cerevisiae*, most snoRNAs are independently transcribed as a monocistronic unit from their own promoters.[9]

Abbreviations used: snoRNA, small nucleolar RNA; snoRNPs, small nucleolar ribonucleoprotein particles.

E-mail address of the corresponding author: lsbrc04@zsu.edu.cn

Table 1. The 47 box H/ACA snoRNAs in the introns of eight host genes from *D. melanogaster*

snoRNA	Len (nt)	Chr	rRNA target		Location	Host gene
			5′ Hairpin	3′ Hairpin		
Ψ28S-1135a	148	X		28 S Ψ1135	2nd Intron	snoRNA host gene I
Ψ28S-1135b	145			28 S Ψ1135		(ribosomal protein S5 gene)
Ψ28S-1135c	143			28 S Ψ1135		
Ψ28S-1135d	142		28 S Ψ483	28 S Ψ1135		
Ψ28S-1135e	151			28 S Ψ1135		
Ψ28S-1135f	151			28 S Ψ1135		
Ψ18S-1854a	139		18 S Ψ1854	18 S Ψ1937		
Ψ18S-1854b	139		18 S Ψ1854	18 S Ψ1937		
Ψ18S-1854c	140		18 S Ψ1854	18 S Ψ1937		
Ψ28S-2876(snoR825)	140			28 S U2876	3rd Intron	
Ψ28S-1192a	140		28 S Ψ1192	28 S Ψ2533	4th Intron	
Ψ28S-1192b	138		28 S Ψ1192			
Ψ28S-1192c	134		28 S Ψ1192			
Ψ28S-1192d	135		28 S Ψ1192			
Ψ28S-2626	142	3R		28 S Ψ2626	3rd Intron	snoRNA host gene II
Ψ18S-1377a(snoR328)	139		18 S Ψ1279	18 S Ψ1377		(ribosomal protein S7 gene)
Ψ18S-1377b	138		18 S Ψ1279	18 S Ψ1377		
Ψ18S-1377c	137		18 S Ψ1279	18 S Ψ1377		
Ψ18S-1377d	136		18 S Ψ1279	18 S Ψ1377		
Ψ18S-1377e	137		18 S Ψ1279	18 S Ψ1377		
Ψ28S-2149(snoR143)	144		28 S U2149			
Ψ28S-3316a	132	2R	28 S Ψ3316		4th Intron	snoRNA host gene III
Ψ28S-3316b	141		28 S Ψ3316			(Dom gene)
Ψ18S-841a(snoR66)	144		18 S Ψ841			
Ψ28S-3378	150		28 S Ψ3378		6th Intron	
Ψ28S-3316c	133		28 S Ψ3316		7th Intron	
Ψ28S-3316d	132		28 S Ψ3316	28 S U584		
Ψ28S-3316e	132		28 S Ψ3316	28 S U584		
Ψ18S-841b	138		18 S Ψ841			
Ψ18S-841c	133		18 S Ψ841			
Ψ18S-841d	138		18 S Ψ841			
Ψ28S-1232	141	X		28 S U1232	1st Intron	snoRNA host gene IV
Ψ28S-1060	150		28 S Ψ1060		2nd Intron	(ribosomal protein L17 gene)
Ψ28S-3436a(snoR708)	141			28 S Ψ3436		
Ψ28S-3436b(snoR75)	143			28 S Ψ3436		
Ψ28S-1837a(snoR14)	137	3L	28 S Ψ1837	28 S Ψ3801	4th Intron	snoRNA host gene V
Ψ28S-1837b	137		28 S Ψ1837	28 S Ψ2938		(proliferation-associated protein gene)
Ψ28S-1837c	139		28 S Ψ1837	28 S Ψ2938		
Ψ18S-1397	141		18 S Ψ1397		6th Intron	
Ψ18S-1347a	137	2R	18 S Ψ1347	28 S Ψ1313	2nd Intron	snoRNA host gene VI
Ψ18S-1347b	143		18 S Ψ1347			(poly(A)-binding protein gene)
Ψ18S-1347c(snoR203)	146		18 S Ψ1347	28 S Ψ1313		
Ψ28S-3327a	140	3L	28 S Ψ3327	18 S U1920	4th Intron	snoRNA host gene VII
Ψ28S-3327b	140		28 S Ψ3327	18 S U1920		(ribosomal protein S4 gene)
Ψ28S-3327c(snoR586)	140		28 S Ψ3327	18 S U1920	5th Intron	
Ψ28S-2719(snoR734)	145	X	28 S Ψ2719		2nd Intron	snoRNA host gene VIII
Ψ18S-531	152		18 S Ψ531	28 S Ψ1850		(ribosomal protein L22 gene)

All snoRNAs are named after the coordination of their rRNA pseudouridylation sites, and the isoforms of each snoRNA are denoted by a, b, c, etc. snoRNAs previously identified by Yuan *et al.*[20] are indicated by their names in parentheses. Ψ represents rRNA pseudouridine sites that are conserved in *S. cerevisiae* and/or mammals,[22] and the known rRNA pseudouridine bases determined previously[19] or mapped here. U denotes a predicted pseudouridine site that has not been confirmed experimentally.

However, the vast majority of animal snoRNAs, as well as a small fraction of yeast snoRNAs, are nested within introns of protein-coding genes,[10,11] demonstrating an alternative pathway of maturation of snoRNAs and showing the important role of the introns in the expression of genetic information. This observation has been further emphasized by the characterization of some noncoding RNA genes that are exclusively transcribed for the expression of intronic snoRNAs.[11,12] Another genomic organization is the snoRNA gene cluster that was first discovered in plants[13,14] and then, intronic snoRNA clusters were found in prevalence in the rice genome.[15,16] Although the snoRNA cluster has also been found in yeasts[9,17] and protozoa,[18] this polycistronic organization of snoRNA has never been reported in metazoa.

The degree of rRNA pseudouridylation in *Drosophila melanogaster* is the highest among all

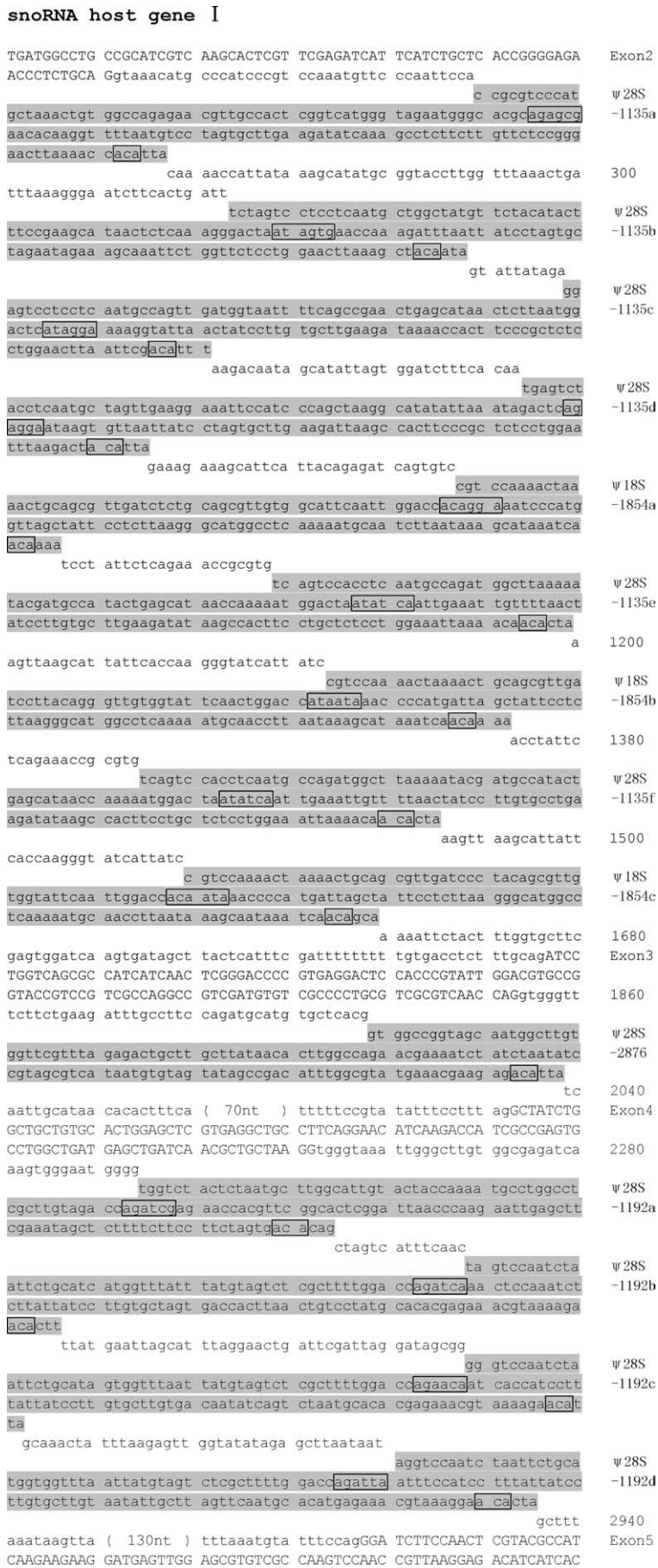


Figure 1 (gene I) (*legend on p.675*)

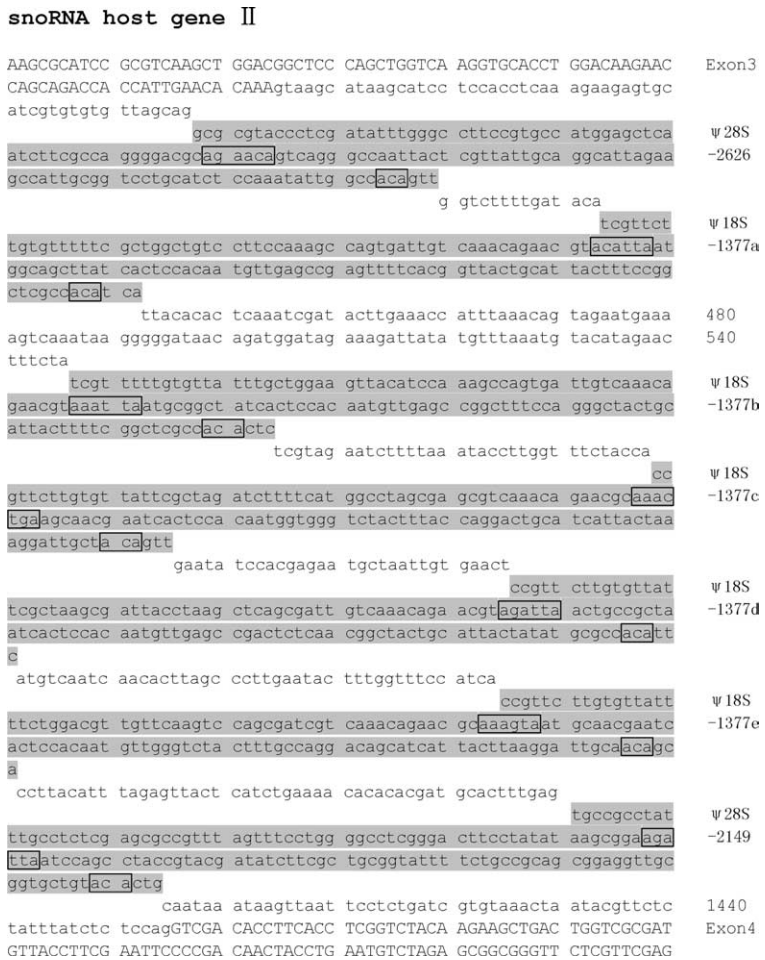


Figure 1 (geneII) (*legend on p.675*)

organisms,[19] which provides a good model for studying the diversity of box H/ACA snoRNAs and their genomic organization. A recent study of experimental RNomics in *D. melanogaster* has identified 20 box H/ACA snoRNAs that were predicted for about 30 pseudouridylation sites in rRNAs and snRNAs.[20] However, it seems that about 100 pseudouridylation sites are located in *D. melanogaster* rRNAs, and more than half of the box H/ACA snoRNAs remain to be found in this organism. Here, we searched for box H/ACA snoRNAs in *D. melanogaster* with a computer-assisted method, and subsequently identified ten novel box H/ACA snoRNAs with their numerous isoforms. Interestingly, a novel polycistronic organization of the box H/ACA snoRNAs was discovered. The results demonstrate the utilization of different strategies for the expression of two classes of snoRNA genes in *D. melanogaster*.

Results

Identification of ten intronic box H/ACA snoRNA gene clusters from *D. melanogaster*

Accounting for the conserved secondary structures, we have developed a computer-assisted method for box H/ACA snoRNA searching. The program searches for the candidates, ranging in size from 120 to 160 nt, that exhibit the hairpin-hinge-hairpin-tail structure with a box H (ANANNA) in the hinge region and box ACA (ACA or AUA) in the tail, and display an rRNA complementarity of at least 9 nt in the internal loop of the hairpin domain. The analytical approach was applied to all the intron sequences of eight protein-coding genes of *D. melanogaster*, from which ten box H/ACA snoRNA genes had been identified.[20] To our surprise, in addition to known snoRNAs, many new box H/ACA snoRNA candidates were also discovered in these introns. After sequence analyses, a total of 47 snoRNA coding regions corresponding to 19 different box H/ACA snoRNAs were identified from 15 introns of the host genes, respectively (Table 1 and Figure 1). Among the 19 snoRNAs, ten novel snoRNAs with 24 isoforms were first identified here. Notably, ten intronic clusters containing 42 snoRNA variants were characteristic of the genomic organization of the box H/ACA snoRNAs in *D. melanogaster*. The clusters are mainly formed by isoforms of the same snoRNA species, while the gene content of the clusters varies largely from

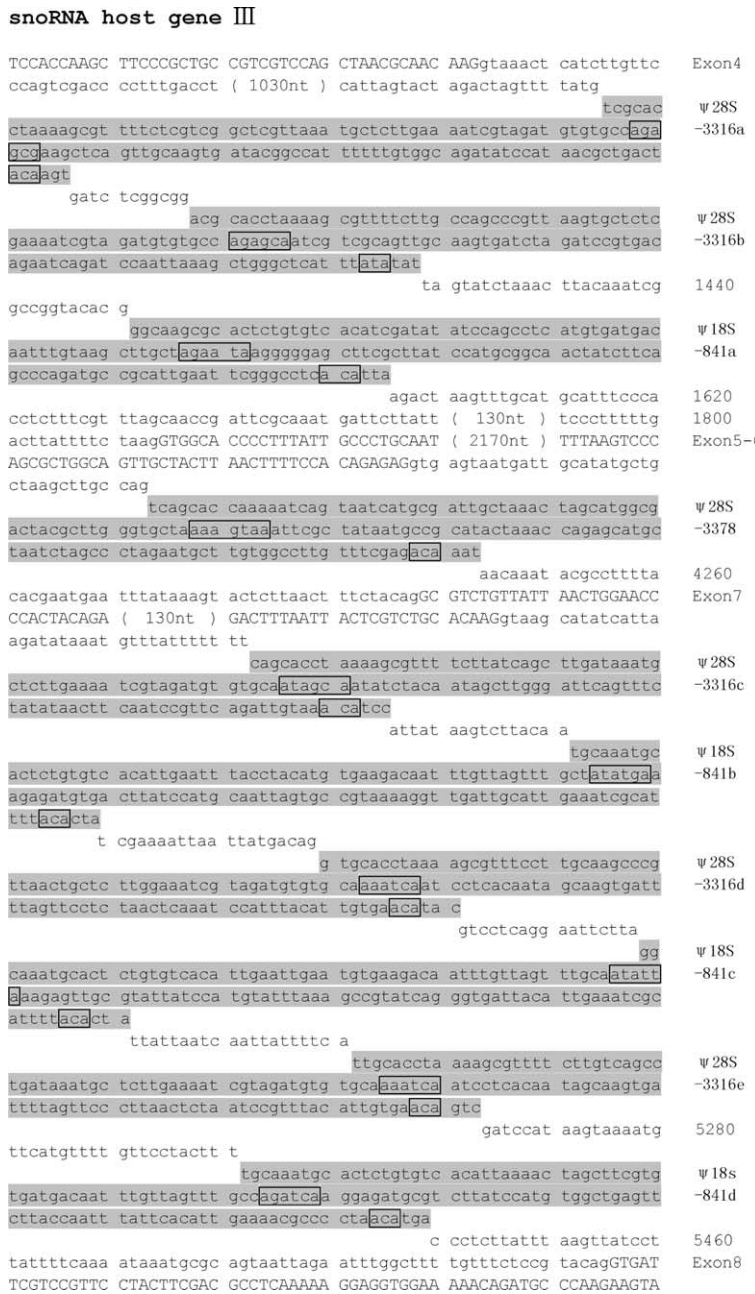


Figure 1 (gene III) (*legend on p.675*)

two to nine snoRNAs. Probably because of modest intron sizes in the fly, the clusters appear very compact. Intergenic spacers between the snoRNA genes are usually 20–40 nt but in some cases the short ones are less than ten nucleotides and a long spacer of 700 nt is also observed in host gene VIII. Among the eight host genes, five are ribosomal protein genes. All introns of snoRNA hosts possess standard boundary signals, i.e. GU at the 5′ end and AG at the 3′ end. The distance from the last snoRNA genes in the clusters to the 3′ end of the introns is at least 55 nt, slightly different from the 71–80 nt that is important for the processing of intronic snoRNAs in mammals.[21] On the other hand, the sequences between the first snoRNA gene in the cluster and the 5′ end of the intron are relatively short, many are less than 50 nt.

Positive detection of the ten novel H/ACA snoRNAs

Ten oligonucleotides were designed and synthesized according to the coding regions of ten corresponding snoRNAs, i.e. Ψ28S-2626, Ψ28S-3316, Ψ28S-1060, Ψ28S-1135, Ψ18S-531, Ψ18S-1854, Ψ28S-1192, Ψ18S-1397, Ψ28S-3378 and Ψ28S-1232, which were newly identified in this analysis. As shown in Figure 2A, all the target snoRNAs were positively detected by Northern blotting with the labeled probes. In each case, a unique and strong

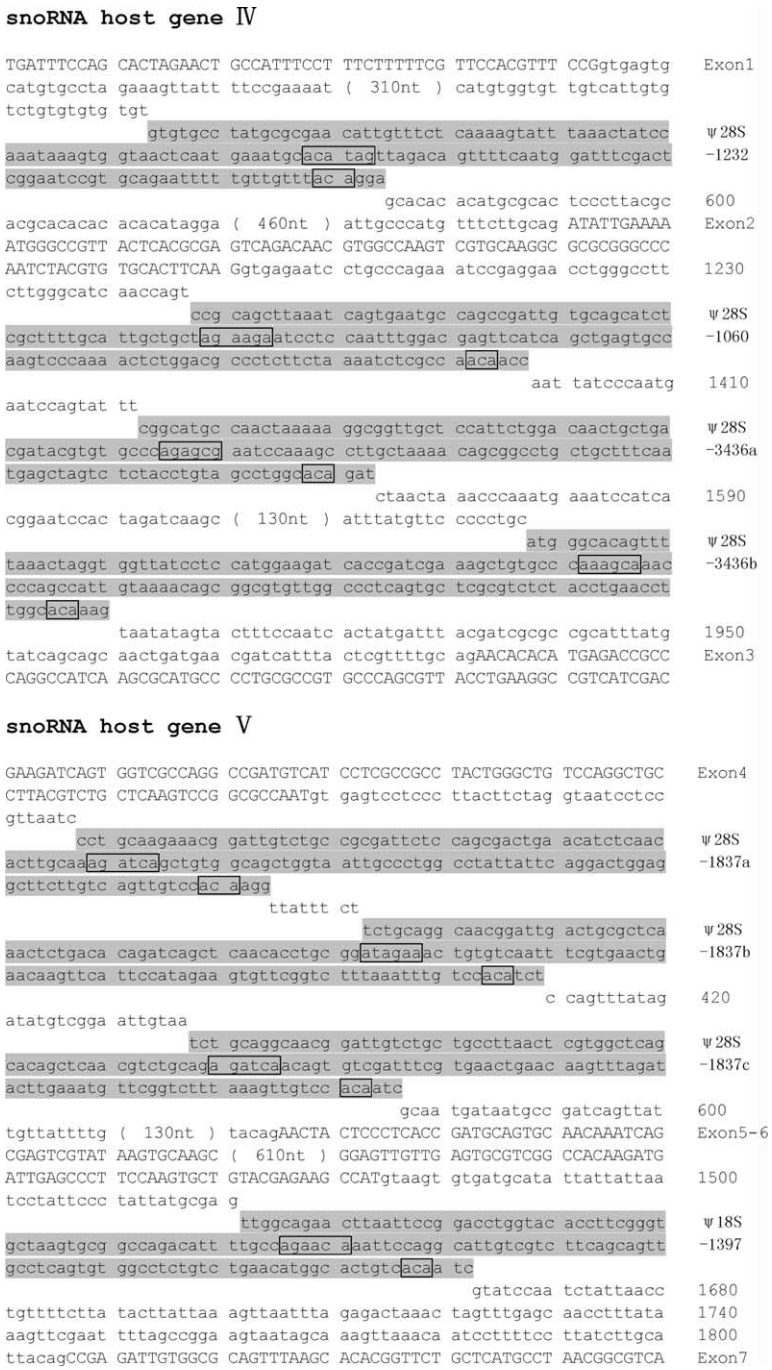


Figure 1 (genes IV and V) (*legend opposite*)

band was revealed under stringent conditions of hybridization and the size of each snoRNA was satisfactorily detected as expected. The 5′ end of the snoRNAs was further mapped by reverse transcription assays with the same labeled oligonucleotide. In most cases, a major cDNA product was obtained for each snoRNA (Figure 2B). The 5′ end of each snoRNA is essentially homogenous and is at the position predicted. Particularly in the case of Ψ28S-1135, the mapping assay revealed two bands that correspond in fact to two isoforms varying in length as expected for 151 nt and 142 nt, respectively.

To investigate the expression of different isoforms in the same intron, four oligonucleotides specific to Ψ28S-1192a, Ψ28S-1192b, Ψ28S-1192c and Ψ28S-1192d were designed and used to identify each isoform in total cellular RNA. The results of reverse transcription demonstrated that the polycistronic precursor from host gene IV was correctly processed to release four individual snoRNAs as expected (Figure 2C). In some cases, higher molecular mass bands corresponding to the polycistronic precursor were observed in the reverse transcription of snoRNAs (Figure 2D).

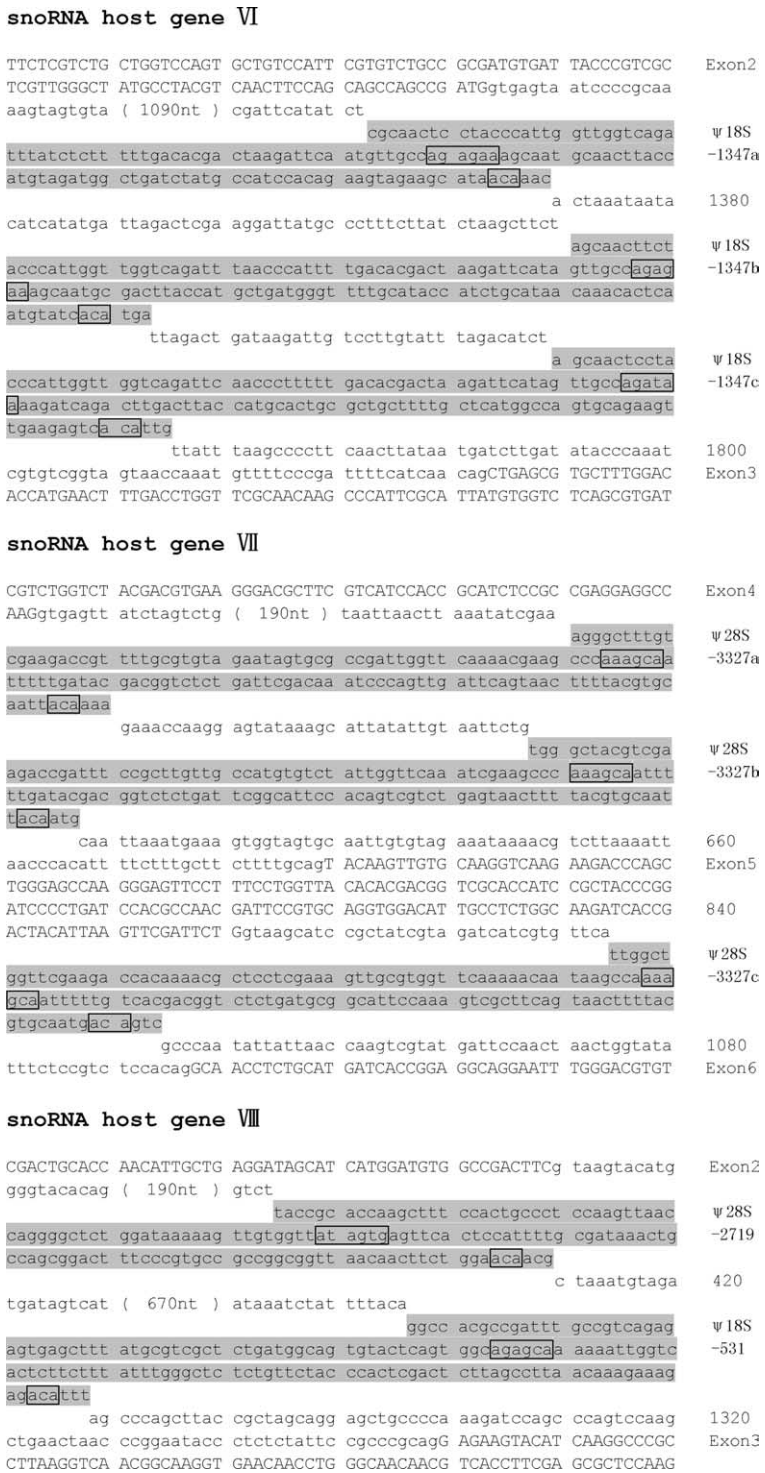


Figure 1. The sequences of ten intronic box H/ACA snoRNA gene clusters from *D. melanogaster*. Exons are in capital letters; introns are in lower case. The coding regions for snoRNAs are shaded, and box H/ACA are boxed.

Analysis of pseudouridylation sites predicted by the novel box H/ACA snoRNAs

In addition to nine snoRNAs whose functions have been described,[20] ten novel box H/ACA snoRNAs were predicted to guide 15 pseudouridylation sites (Figure 3) according to the relationship between the structure and function of this snoRNA gene family.[4] Among the 15 pseudouridylation sites, seven, i.e. Ψ1060, Ψ1135, Ψ1192, Ψ2533, Ψ2626, Ψ3316 and Ψ3378 in 28 S rRNA have been experimentally verified by Ofengand *et al*.[19] In addition, two pseudouridylation sites, Ψ531 and Ψ1397 in 18 S rRNA, are conserved in the yeast *Saccharomyces cerevisiae* and/or mammals.[22] Furthermore, the isoforms of nine known snoRNAs were further analyzed and predicted to guide four novel pseudouridylation sites based on their functional elements that are absent from other isoforms (Figure 3B).

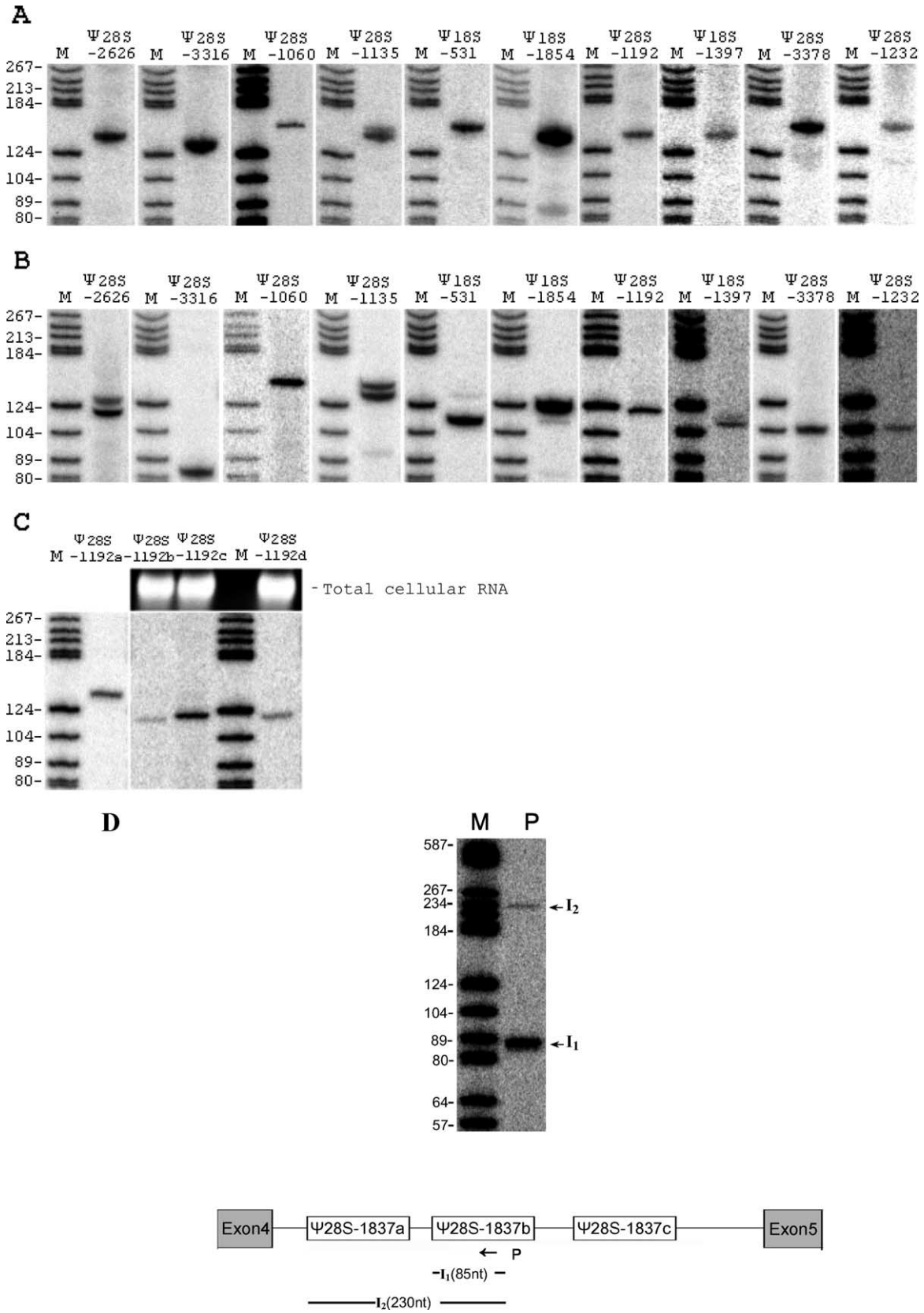


Figure 2. Positive detection and mapping of 5′ end of novel snoRNAs. A, Northern blot analyses. Aliquots of 30 μg of total cellular RNA were separated in each lane and hybridized with the labeled oligonucleotide probes described in Materials and Methods. Lane M, molecular mass markers (pBR322 digested with HaeIII and 5′-end-labeled with [γ-^{32}P]ATP). B, Reverse transcription analyses of ten novel snoRNAs with the same primers as Northern blots. C, Reverse transcription analyses of the expression of four isoforms Ψ28S-1192 in the intron 4 of snoRNA host gene I by primers specific to each isoform. These primers are complementary to the divergent regions of 3′ end of each

CMC-alkali-treated *D. menalogaster* RNA was applied to determine the new rRNA pseudouridylation sites. Six sites, Ψ1854, Ψ1937 in 18 S rRNA and Ψ483, Ψ1837, Ψ1850, Ψ2938 in 28 S rRNA, were precisely mapped (Figure 4). However, U1232 in 28 S rRNA, a pseudouridylation site predicted by Ψ28S-1232, was found to be unmodified while two known pseudouridylation sites, Ψ1227, Ψ1240, were clearly shown in the same assay (Figure 4E). In addition, two pseudouridylations, Ψ1347 in 18 S rRNA and Ψ3801 in 28 S rRNA, predicted by Yuan *et al*.[20] were also determined in the primer extension. For some unknown reason, the extension of primer Dm28S584, which was applied to analyze Ψ584 in 28 S rRNA, failed to show any results (data not shown).

Large number of box H/ACA snoRNA isoforms in *D. menalogaster*

Multiple isoforms of the box H/ACA snoRNA genes were identified from the *D. menalogaster* genome. Of the 19 different box H/ACA snoRNA genes, ten have at least one variant (Table 1). For example, Ψ18S-1377 has five isoforms in the same intron, while five isoforms of Ψ28S-3316 are distributed in the two introns of host gene III. Evidently, multiple isoforms of one snoRNA gene are frequently found within one intron, suggesting the important role of local duplications for the cluster formation. The sequence alignment of six Ψ28S-1135 isoforms and three Ψ18S-1854 isoforms clearly demonstrates the duplication of the two snoRNA genes, especially in the three regions. They all consist of different isoforms of Ψ28S-1135 and Ψ18S-1854, in the second intron of host gene I (Figure 5). The degree of sequence similarity between the isoforms varies from 99% to 70%, implying the duplication events in different periods and a trend of enlarging the intronic snoRNA clusters.

In many cases, although mutations including insertions or deletions had occurred frequently, the secondary structures of the isoforms remained unchanged and the isoforms kept in common at least one functional element. Interestingly, the accumulation of mutation in the isoforms would lead to partial alternation of snoRNA's function in loss or gain of rRNA complementary sequences. For example, the variations in the 5′-hairpin of 28 SΨ1135d and 3′-hairpin of 28 SΨ1192a result in novel complementary sequences for 28 S rRNA at U483 and U2533 pseudouridylation, respectively. On the other hand, the sequence variation among the isoforms can also cause the loss or change of the guide sequence for rRNA pseudouridylation, such as the 3′-hairpin of 18 SΨ1347b and 28 SΨ1837a (see Table 1).

Discussion

Ten intronic snoRNA gene clusters were identified from *D. melanogaster* in this study, demonstrating for the first time a novel polycistronic organization of snoRNA genes in animals. In fact, all intronic snoRNA gene clusters identified so far from plants consist of homogeneous box C/D snoRNA genes with few exceptions, such as the rice hsp70 gene where two box H/ACA snoRNA genes are mixed with four other box C/D snoRNA genes in the first intron of the host.[15] Here, we show homogeneous box H/ACA snoRNA gene clusters without any intervening box C/D snoRNA. Therefore, the intronic box H/ACA snoRNA gene clusters are first demonstrated not only in *D. melanogaster* but also in eukaryotes.

Up to now, more than 30 box C/D snoRNA gene variants including three methylation guides for snRNA and five orphan guides with unknown target have been identified from *D. melanogaster* (GenBank accession number U40615 for Z1 and AJ010684 for Z5 sno(RNA)).[20,23] Similar to mammals, all the methylation guide snoRNAs characterized so far in *D. melanogaster* are intron-encoded and arranged strictly in the mode of one snoRNA per intron. For example, DUHG1, a human UHG-like non-coding RNA gene in *D. melanogaster*, encodes 16 box C/D snoRNAs in their 16 introns, respectively (Figure 6), representing an outstanding gene organization and expression strategy for the C/D snoRNA gene family in *D. melanogaster*.[24] Interestingly, we have shown that box H/ACA snoRNAs in *D. melanogaster* possess a different genomic organization, that is, intronic gene clusters. A typical case is snoRNA host gene I, in which 14 box H/ACA snoRNA genes are nested within three introns of a ribosomal protein gene RpS5. Among 47 snoRNAs analyzed here, 42 were found in the intronic clusters. The high proportion of box H/ACA snoRNAs in the intronic clusters suggests that this kind of gene organization may be prevalent in the *D. melanogaster* genome. The polycistronic snoRNAs in an intron imply a processing mechanism that involves both endonucleolytic and exonucleolytic cleavages, substantially different from processing a singleton of intronic snoRNA

isoform, respectively. The sequence of the primer for Ψ28S-1192a is completely different from the three others, and there are at least six nucleotide differences among the primers for Ψ28S-1192b, Ψ28S-1192c and Ψ28S-1192d. D, Detection of a polycistronic snoRNA precursor in the reverse transcription of total RNA. Lane P, reverse transcription with primer PΨ28S-1837b, which is specific antisense to the 3′ end of Ψ28S-1837b. Lane M, molecular mass markers. I_1 and I_2 indicate cDNA bands from mature Ψ28S-1837b and the precursor containing two box H/ACA snoRNAs, respectively, as illustrated below.

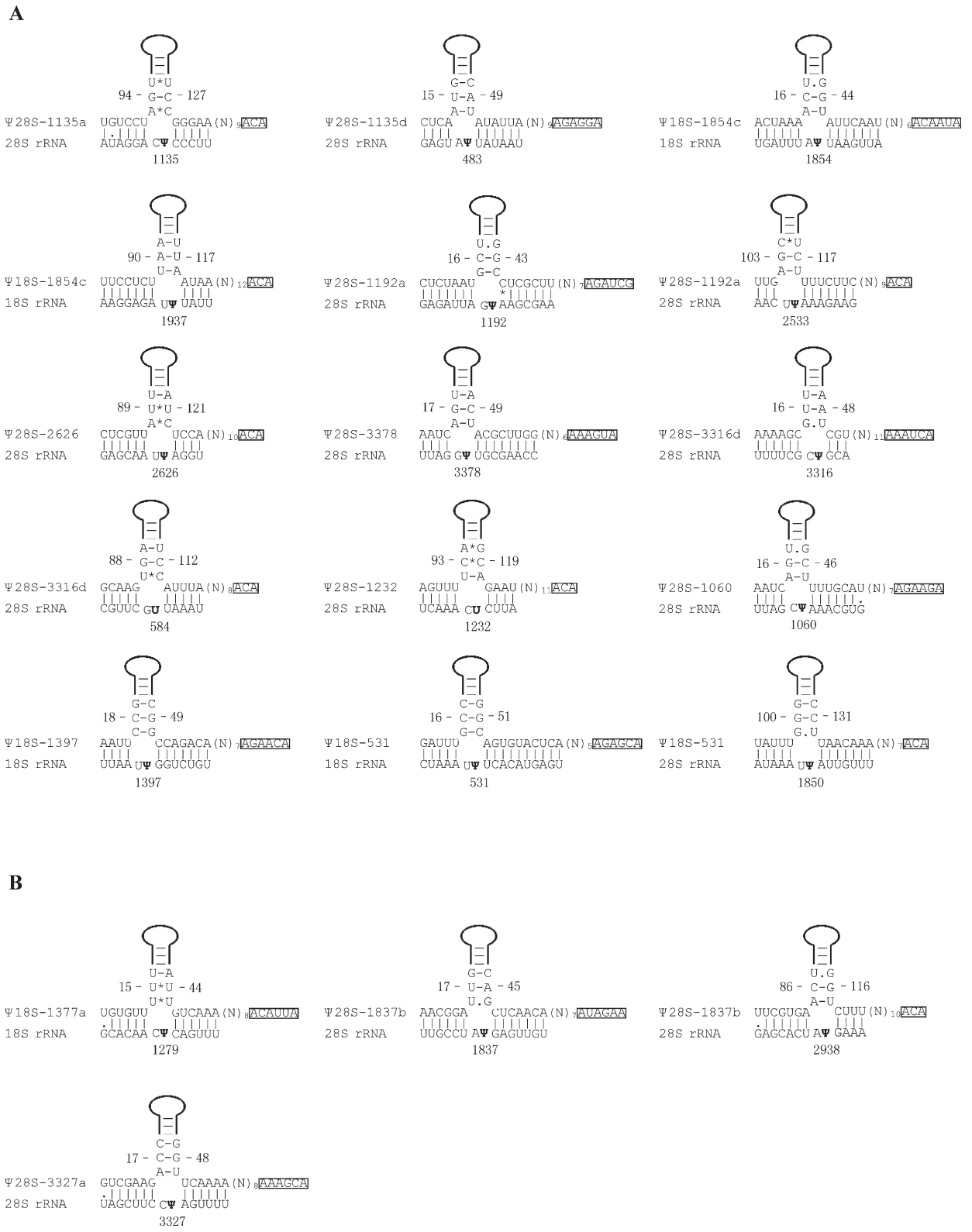


Figure 3. Predicted pseudouridylation guide duplex between snoRNAs and rRNAs. The snoRNA sequences in a 5′ to 3′ orientation are shown in the upper strands, while rRNA sequences in a 3′ to 5′ orientation are shown in the lower strands. The two sequence motifs are boxed and the upper parts of the hairpins are represented by continuous lines. The positions of pseudouridine bases are indicated by numbers. Ψ represents rRNA pseudouridine sites that are conserved in *S. cerevisiae* and/or mammals,[22] and the known rRNA pseudouridine bases determined previously[19] or mapped here. U denotes a predicted pseudouridine site that has not been confirmed experimentally. Only one isoform is shown if the snoRNA has more isoforms. A, Pseudouridylation sites predicted by novel snoRNAs. B, Four novel sites predicted from isoforms of snoRNAs identified previously.

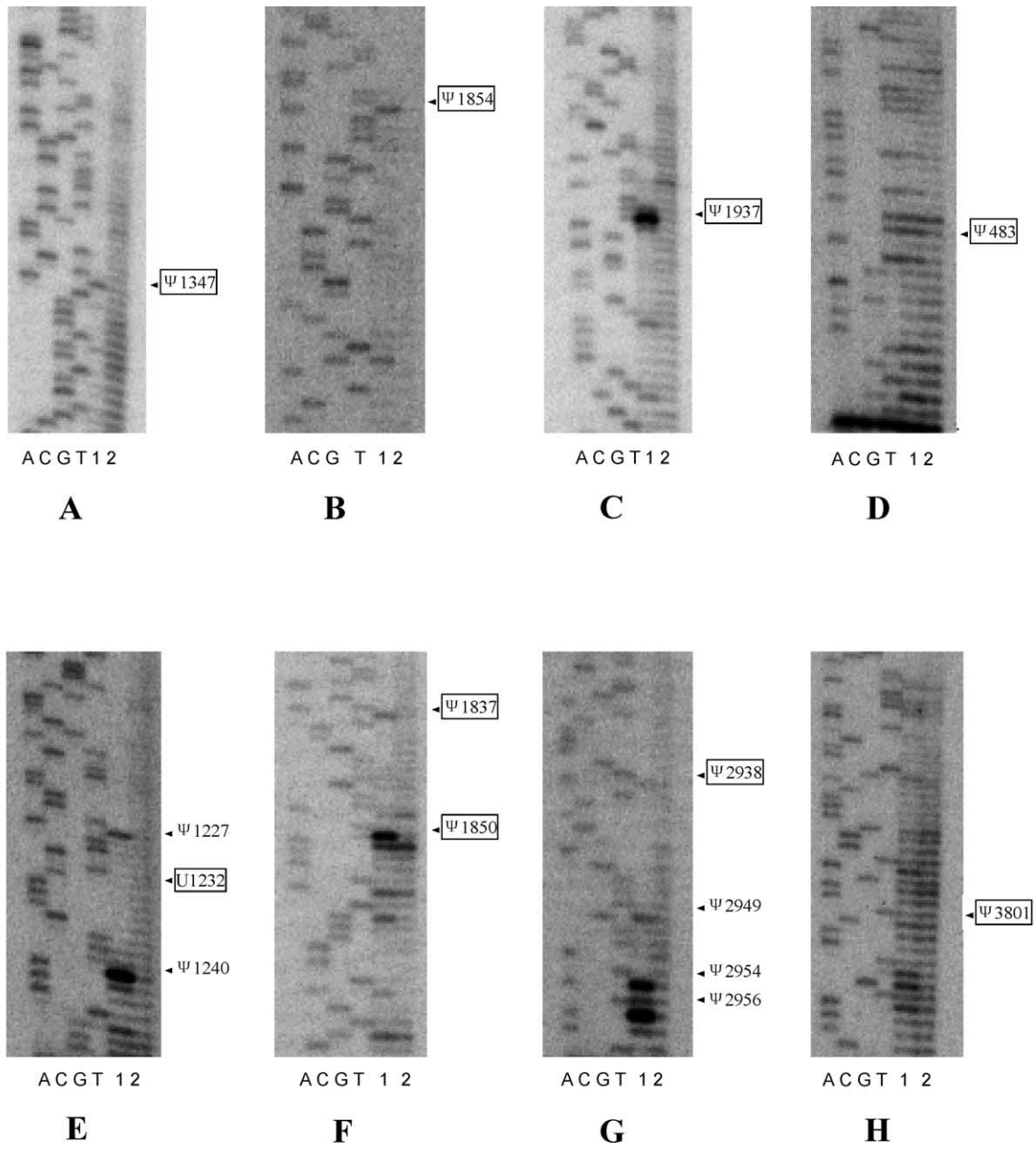
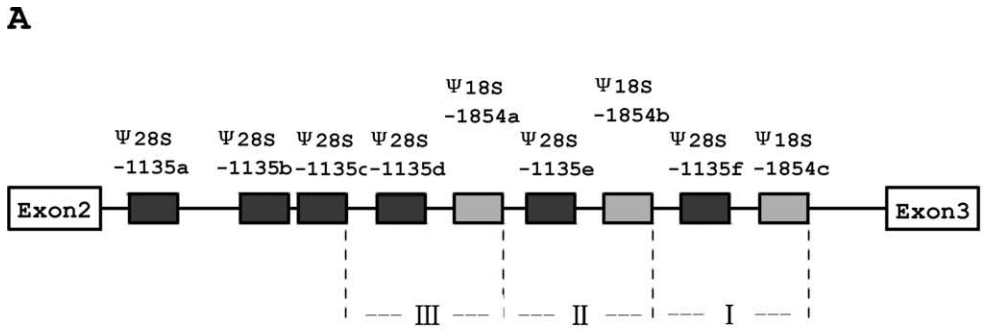


Figure 4. Mapping of rRNA pseudouridylation sites predicted by the novel snoRNAs. Lane 1, reverse transcription with CMC-treated total RNA; lane 2, reverse transcription reaction with CMC-untreated total RNA. Lanes A, C, G and T, the rDNA sequence ladder; positions of pseudouridine residues are indicated by arrows. Pseudouridylation sites boxed are predicted in this work, and those without boxes were previously identified by Ofengand and Bakin.[19]

(Figure 6).[14,17] The adoption of different organizations and expression strategies for the snoRNA genes may reflect intrinsic differences in gene regulation and functional evolution between the two classes of snoRNAs in *D. melanogaster*. It is worth noting that most intronic snoRNAs are produced by a splicing-dependent processing pathway involving exonucleolytic trimming of the debranched lariat.[25] However, another minor, splicing-independent mode for intron-encoded box C/D snoRNAs has been found in yeast and mammals,[26–28] suggesting the complexity of intronic snoRNA biogenesis in eukaryotes. In Figure 6, we describe a splicing-dependent pathway for snoRNA release from host introns, but the possibility that the pre-mRNA is a substrate of endonucleolytic cleavage, particularly, when the splicing efficiency is reduced in the organism, cannot be ruled out.

Although endonucleolytic activity is absolutely necessary for processing the polycistronic transcript of snoRNAs, diverse endonuclease(s) may be involved in the maturation of the snoRNA. Endonuclease RNase III that recognizes a strong potential secondary structure has been proved to play a key role for processing the polycistronic transcript of snoRNAs in the budding yeast.[17,29]



B

Ψ28S-1135

```
I    *********a cctattctca gaaaccgcgt g*TCAGTCCA CCTCAATGCC AGATGGCTTA    50
II   *********T ---------- ---------- -*-------- ---------- ----------    50
III  aagacaatag -a----agtg --tcttt-ac aa-G----T- ---------T --T--A**AG    58

I    AAAATACGAT GCCATACTGA GCATAACCAA AAATGGACTA ATATCAATTG AAATTGTTTT   110
II   ---------- ---------- ---------- ---------- ---------- ----------   110
III  G----T-C-- C---GCTAAG -----T**** **--TA-A-- GAC---GAG- --TAA--G--   112

I    AACTATCCTT GTGCCTGAAG ATATAAGCCA CTTCCTGCTC TCCTGGAAAT TAAAACAACA   170
II   ---------- ----T----- ---------- ---------- ---------- ----------   170
III  --T------A ----T----- --*------- -----C---- --------T- ---G--T---   171
```

Ψ18S-1854

```
I    CTAaagttaa gcattattca ccaagggtat cattatCCGT CCAAAACTAA AACTGCAGCG   230
II   ---------- ---------- ---------- ---------- ---------- ----------   230
III  T--g-a**-g aa-gc----- tt-ca-ag-- --g-g----- ---------- ----------   229

I    TTGATCCCTA CAGCGTTGTG GTATTCAATT GGACCACAAT AAACCCCATG ATTAGCTATT   290
II   -------T-- ---G------ --------C- ------T--- ---------- ----------   290
III  ------T--G ---------- -C-------- --------GG ---T------ G---------   289

I    CCTCTTAAGG GCATGGCCTC AAAAATGCAA CCTTAATAAA GCAATAAATC AACAGCA      347
II   ---------- ---------- ---------- ---------- ---*------ ----AA-     346
III  ---------- ---------- ---------- T--------- ---*------ ----AA-     345
```

Figure 5. Duplication of snoRNA-coding region in the second intron of snoRNA host gene I. A, snoRNA genes are represented by boxes. Three duplication regions for two snoRNA genes are indicated. B, Sequence alignment of the duplication regions. snoRNA genes are in upper-case letters and indicated by arrowheads. Nucleotide identities are denoted by hyphens and those absent from either sequence by asterisks.

Recently, dicistronic tsnoRNA, a heterogeneous cluster consisting of a box C/D snoRNA and tRNA, was found in plants.[30] Dicistronic precursors transcribed from the cluster were processed by endonuclease RNase Z that specifically recognizes 3′ ends of tRNAs, instead of intergenic sequences which were less than ten nucleotides between the two RNAs.[30] Intergenic spacers between the box H/ACA snoRNA genes in *D. melanogaster* are remarkably short and rich in A and T, so they can hardly form secondary structures similar to those that are involved in RNase III cleavage. Furthermore, the analysis of sequences reveals no conserved element among 33 intergenic spacers from the snoRNA genes. It is very likely that the processing signals recognized by endonuclease are not in the intergenic spacer regions but in box H/ACA snoRNAs, especially in the high structure of the snoRNAs. Endonuclease that is involved in the processing of polycistronic transcripts of box H/ACA snoRNAs may contribute to the amplification of intronic clusters of this snoRNA family and the high diversity of rRNA pseudouridylation in *D. melanogaster*.

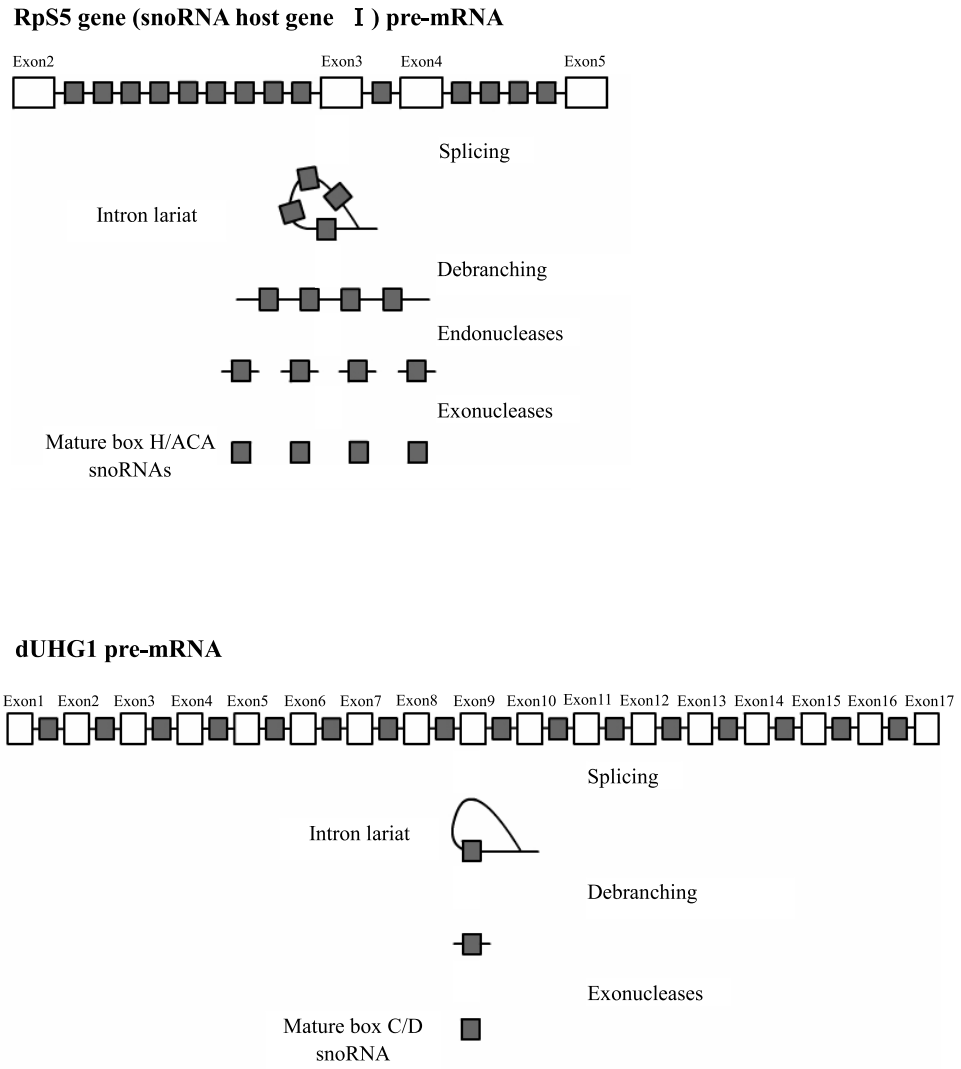


Figure 6. Different gene organizations and expression strategies for the two classes of snoRNAs in *D. melanogaster*.

Materials and Methods

Computational search for box H/ACA in the *D. melanogaster* genomic database

Complete sequences of eight *D. melanogaster* genes, CG8922, CG1883, CG9696, CG3203, CG10576, CG5119, CG11276 and CG7434, were obtained from the Flybase.[31] All the intron sequences in the genes were analyzed by using our computer program that takes into account both the sequence motifs and secondary structures in box H/ACA snoRNAs. The search for isoforms of the snoRNAs in the *D. melanogaster* genome was performed by using BLAST[32] and FASTA[33] programs. Sequence alignments and comparison of snoRNAs were performed by using Clustal X 1.8 and DNAstar packages.

RNA extraction and analyses

Fresh wild-type *D. melanogaster* larva were cultured and collected for the RNA extraction. Total cellular RNA was isolated and purified according to the method of guanidine thiocyanate/phenol/chloroform.[34] An aliquot of 30 μg of total RNAs was analyzed by electrophoresis on 8% (w/v) acrylamide/7 M urea gels. Electrotransfer onto nylon membrane (Hybond-N + ; Amersham) was followed by UV irradiation for five minutes. Hybridization with 5′-labeled probes was performed as described.[23] Reverse transcription was carried out in a 20 μl reaction mixture containing 20 μg of *D. melanogaster* total RNA and 20 ng of 5′-labeled primer in the presence of 500 μM dNTPs. After denaturation at 65 °C for five minutes, the mixture was cooled to 42 °C for ten minutes, and then 200 units of MMLV reverse transcriptase (Promega) was added and incubated at 42 °C for 60 minutes. The cDNA synthesized by reverse transcription was analyzed by electrophoresis on 8% acrylamide/7 M urea gels.

Mapping of ribose pseudouridylation by CMC-primer extension method

Mapping of *D. melanogaster* ribosomal pseudouridines was performed essentially as described by Bakin and Ofengand.[35] The *D. melanogaster* 28 S and 18 S rDNA

were amplified by PCR with the primer pairs Dm28SL1/Dm28SR1, Dm28SL2/Dm28SR2 and Dm18SL/Dm18SR, then cloned into the SmaI site of plasmid pUC18. An rDNA sequence ladder was prepared with the same primer used for rRNA pseudouridylation mapping and run in parallel with the reverse transcription reaction as a molecular mass marker.

Oligodeoxynucleotides

Oligonucleotides were synthesized and purified by Sangon Co. (Shanghai, China). The sequences of oligonucleotide probes and primers used for Northern blotting and reverse transcription were as follows: 5′-GCAGGACCGCAATGGCTTCT-3′ (PΨ28S-2626); 5′-GCT ATTGTGAGGATTGATTTTGC-3′ (PΨ28S-3316); 5′-GCG AGATTTTAGAAGA GGGCGTC-3′ (PΨ28S-1060); 5′-TTAAATTCCAGGAGAGCGGG-3′ (PΨ28S-1135); 5′-TAGAACAGAGAGCCCAAATAAAGAA-3′ (PΨ18S-531); 5′-TTTATTAAGGTTGCAT TTTTGAGG-3′ (PΨ18S-1854); 5′-TTTCTCGTGTGCATTAGACTGAT-3′ (PΨ28S-1192c); 5′-CCACACTGAGGCAACTGCTGAA-3′ (PΨ18S-1397); 5′-GCTCTGGTTTAGTATGC GGC-3′ (PΨ28S-3378); 5′-CGAAATCCATTGAAAACTGTCTAA-3′ (PΨ28S-1232); 5′-TTCACGAAATTGACACAGTTTCTA-3′ (PΨ28S-1837b). The following oligonucleotides were used for identification of Ψ28S-1192 isoforms: 5′-GTCACTAGAAGGAA GAAA AGAGC-3′ (PΨ28S-1192a); 5′-GTGTGCATAGG ACA GTTAAGTGG-3′ (PΨ28S-1192b); 5′-TTTCTCGTGT GC ATTA GACTGAT-3′ (PΨ28S-1192c); 5′-CTCATGTGC AT TGAACTAAGCAAT-3′ (PΨ28S-1192d). Primers used for rRNA pseudouridylation mapping were as follows: 5′-AACCAGACAAATCGCTCCAC-3′ (Dm18S1347); 5′-AA ACAACCGTAACACGCAAGG-3′ (Dm18S1854); 5′-TGA TCCTTC CGCAGGTTCACC-3′ (Dm18S1937); 5′-CAAT GTCCTTATATGGAAAAAATGC-3′ (Dm28S483); 5′-CAC TGTAATCATATAAATCTATCAGCACTT-3′ (Dm28S584); 5′-GATCT TCATATCAAGAAAGTTAAGGTTC-3′ (Dm28S1232); 5′-TTTATGGTCGTTCCTGTTG CC-3′ (Dm28S1850); 5′-CGTTTTATTAAAGAATTTGTTTGCG-3′ (Dm28S2938); 5′-ACCACTTACAACACCTTGCCTG-3′ (Dm28S3801). Primers used for PCR of *D. melanogaster* 28 S and 18 S rDNA were: 5′-GGTTATGTTATTATTCTT CGTTGGTT-3′ (Dm28SL1); 5′-AATTCGCTTTGTTTATAT AGTTAGGC-3′ (Dm28SR1); 5′-GCCTAACTA TATAAAC AAAGCGAATT-3′ (Dm28SL2); 5′-AATTGATGACGAG CTGTTTGG-3′ (Dm28 SR2); 5′-TGGTTGATCCTGCCAG TAGTTAT-3′ (Dm18SL); 5′-AACCCATCTTCGTTTTA TTTTGA-3′ (Dm18SR). The primers and probes used in reverse transcription and rDNA sequencing were 5′-end-labeled with [γ-^{32}P]ATP (Yahui Co.) and submitted to purification according to standard laboratory protocols as described.[36]

Database accession codes

All snoRNA gene sequences identified in this study have been deposited in the EMBL database under accession numbers from AJ629193 to AJ629216 and from AJ629256 to AJ629278.

Acknowledgements

We are grateful to Xiao-Hong Chen for technical assistance and Quan-Sheng Du for *D. melanogaster* culture. We also thank Professor Mohsen Ghadessy for revising the English text. This research was supported by the National Natural Science Foundation of China (key project 30230200, 30170216), and by a Fund from the Ministry of Education of China to L.H.Q.

References

1. Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
2. Filipowicz, W. & Pogaèiæ, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell. Biol.* **14**, 319–327.
3. Balakin, A. G., Smith, L. & Fournier, M. J. (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.
4. Ganot, P., Bortolin, M. L. & Kiss, T. (1997). Related site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell*, **89**, 799–809.
5. Smith, C. M. & Steitz, J. A. (1997). Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*, **89**, 669–672.
6. Bachellerie, J. P., Cavaille, J. & Qu, L. H. (2000). Nucleotide modifications of eukaryotic rRNAs: the world of small nucleolar RNA guides revisited. In *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions* (Garrett, R. A., Douthwaite, S., Liljas, A., Matheson, A., Moore, P. B. & Noller, H., eds), pp. 191–203, ASM Press, Washington, DC.
7. Tycowski, K. T., You, Z. H., Graham, P. J. & Steitz, J. A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell*, **2**, 629–638.
8. Clouet d'Orval, B., Bortolin, M. L., Gaspin, C. & Bachellerie, J. P. (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. *Nucl. Acids Res.* **29**, 4518–4529.
9. Lowe, T. M. & Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168–1171.
10. Maxwell, E. S. & Fournier, M. J. (1995). The small nucleolar RNAs. *Annu. Rev. Biochem.* **35**, 897–934.
11. Bachellerie, J. P., Cavaille, J. & Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie*, **84**, 775–790.
12. Tycowski, K. T., Shu, M. D. & Steitz, J. A. (1996). A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, **379**, 464–466.
13. Leader, D. J., Sanders, J. F., Waugh, R., Shaw, P. & Brown, J. W. (1994). Molecular characterisation of plant U14 small nucleolar RNA genes: closely linked genes are transcribed as polycistronic U14 transcripts. *Nucl. Acids Res.* **22**, 5196–5203.
14. Brown, J. W., Echeverria, M. & Qu, L. H. (2003). Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci.* **8**, 42–49.
15. Liang, D., Zhou, H., Zhang, P., Chen, Y. Q., Chen, X., Chen, C. L. & Qu, L. H. (2002). A novel gene organization: intronic snoRNA gene clusters from *Oryza sativa*. *Nucl. Acids Res.* **30**, 3262–3272.
16. Chen, C. L., Liang, D., Zhou, H., Zhou, M., Chen,

Y. Q. & Qu, L. H. (2003). The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. *Nucl. Acids Res.* **31**, 2601–2613.
17. Qu, L. H., Henras, A., Lu, Y. J., Zhou, H., Zhou, W. X., Zhu, Y. Q. *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.
18. Dunbar, D. A., Chen, A. A., Wormsley, S. & Baserga, S. J. (2000). The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistronic RNA. *Nucl. Acids Res.* **28**, 2855–2861.
19. Ofengand, J. & Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *J. Mol. Biol.* **266**, 246–268.
20. Yuan, G., Klambt, C., Bachellerie, J. P., Brosius, J. & Huttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucl. Acids Res.* **31**, 2495–2507.
21. Hirose, T. & Steitz, J. A. (2001). Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc. Natl Acad. Sci. USA*, **98**, 12914–12919.
22. Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K. & Lane, B. G. (1995). The pseudouridine residues of ribosomal RNA. *Biochem. Cell. Biol.* **73**, 915–924.
23. Zhou, H., Chen, Y. Q., Du, Y. P. & Qu, L. H. (2002). The *Schizosaccharomyces pombe* mgU6-47 gene is required for 2′-O-methylation of U6 snRNA at A41. *Nucl. Acids Res.* **30**, 894–902.
24. Tycowski, K. T. & Steitz, J. A. (2001). Non-coding snoRNA host genes in *Drosophila*: expression strategies for modification guide snoRNAs. *Eur. J. Cell. Biol.* **80**, 119–125.
25. Tollervey, D. & Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell. Biol.* **9**, 337–342.
26. 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 exo- and endonucleolytic activities. *Mol. Cell. Biol.* **18**, 3367–3383.
27. Cavaille, J., Vitali, P., Basyuk, E., Huttenhofer, A. & Bachellerie, J.-P. (2001). A novel brain-specific box C/D small nucleolar RNA processed from tandemly-repeated introns of a non-coding RNA gene in rat. *J. Biol. Chem.* **276**, 26374–26383.
28. Hirose, T., Shu, M.-D. & Steitz, J. A. (2003). Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol. Cell*, **12**, 113–123.
29. Chanfreau, G., Legrain, P. & Jacquier, A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. *J. Mol. Biol.* **284**, 975–988.
30. Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S. *et al.* (2003). Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.* **22**, 621–632.
31. The FlyBase Consortium (2003). The FlyBase database of the *Drosophila* genome projects and community literature. *Nucl. Acids Res.* **31**, 172–175. http://flybase.org/.
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
33. Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
34. Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 732–735.
35. Bakin, A. & Ofengand, J. (1993). Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry*, **32**, 9754–9762.
36. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Edited by M. Yaniv

(Received 5 March 2004; received in revised form 24 May 2004; accepted 14 June 2004)