SnoRNA gene clusters

J.W.S. Brown, G. Clark, G. Thow & C.G. Simpson

Small nucleolar RNAs (snoRNAs) are found in the major sub-compartment of the nucleus – the nucleolus. In the nucleolus, ribosomal RNA (rRNAs) is transcribed, processed, and complexed with ribosomal proteins to form ribosomal subunits for translation. SnoRNAs are involved in the processing of rRNAs. Apart from specific cleavages to form the 18S, 5.8S and 25/28S rRNAs, the two major types of modifications are 2'-O-ribose methylation and pseudouridylation, each of which occurs at around 100 sites in rRNAs in plants and animals. These modifications are thought to fine tune the structure of the ribosome, making the translation process more efficient. The sites of 2-O-ribose methylation in rRNAs are determined by a class of snoRNAs called box C/D snoRNAs. These snoRNAs contain two terminal conserved sequences, box C and D, and each contains either one or two sequences complementary to rRNAs which determine the methylation site via base pairing of the snoRNA to rRNA (Fig. 1a).

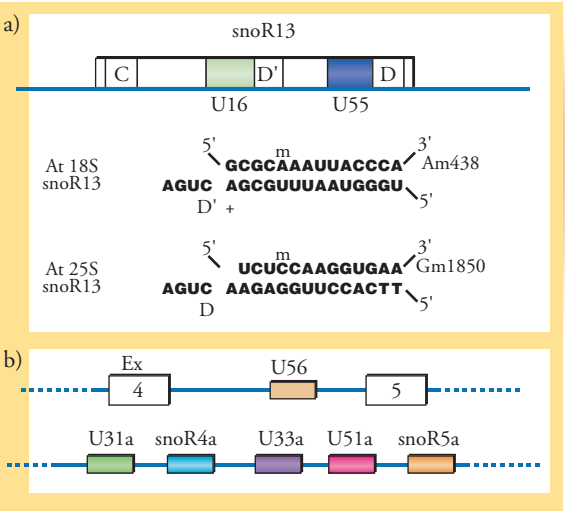

Figure 1 Structure and function of box C/D snoRNAs. (a) Schematic structure of box C/D snoRNAs showing the terminal box C and D sequences, an internal D' sequence and sequences complementary to rRNAs (shaded boxes). SnoR13 contains two regions which base-pair with specific rRNA sequences and guide methylation at position Am438 in 18S rRNA and Gm1850 in 25S rRNA. (b) Intron-encoded and polycistronic snoRNAs from *Arabidopsis*. Large open boxes – exons; small boxes – different snoRNA genes.

The way in which the majority of snoRNAs are expressed differs in the three eukaryotic organisms in which they have been most widely studied. The main modes of snoRNA gene organisation are intron-encoded and polycistronic (Fig. 1b). Intron encoded snoRNAs are found in introns of protein coding genes and are processed following splicing. Polycistronic snoRNAs consist of closely linked clusters of genes which are transcribed together as a precursor snoRNA from which they are processed. In vertebrates, there are no cases of polycistronic snoRNA genes and the majority are intron-encoded. In yeast, there are five polycistrons and the majority of the remaining snoRNA genes are intron-encoded. In contrast, in plants, the vast majority of snoRNA genes are polycistronic[1] and there are only three examples of intron-encoded plant snoRNAs to date.

We have mapped methylation sites in plant rRNAs and have searched the *Arabidopsis* genome sequence for snoRNA genes using complementary sequences. To date, 24 snoRNA gene clusters each containing 2-5 snoRNA genes have been discovered. Some of the clusters are related, containing the same genes, and are found in different positions in the *Arabidopsis* genome. The organisation of the different gene clusters illustrates a number of mechanisms of gene and genome evolution. Clearly, there are local gene duplications of either single genes (Fig. 2a), single genes

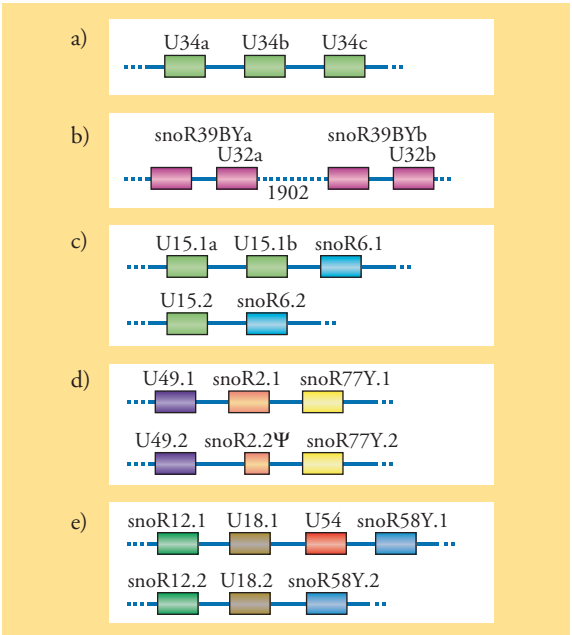

Figure 2 *Arabidopsis* snoRNA gene clusters showing: (a) a homogeneous gene cluster; (b) locally duplicated pairs of genes; (c) gene clusters at different chromosomal locations with single gene duplication; (d) gene clusters at different locations with partial gene loss forming pseudogenes; and (e) gene clusters at different locations showing gene loss. The genes are shown by coloured boxes.

within a cluster (Fig. 2c), or pairs or clusters of genes (Fig. 2b). The presence of related gene clusters on different chromosomes and chromosome areas (Fig. 2c-e) could reflect duplication events, including polyploidisation, dispersed by chromosomal translocations or inversions. In some cases, mutation has led to loss of conserved sequences, leading to pseudogenes (snoR11 in Fig. 2d) and other related clusters suggest that genes have been lost completely from clusters (U54 in Fig. 2e). At the level of gene sequences, different alleles exhibit different degrees of sequence variation. The sequence variation can take the form of base substitutions, insertions or deletions.

The evolution of snoRNAs from *Archae* and yeast to higher eukaryotes is thought to have occurred through a series of gene duplications, mutation and selection, both in terms of ability to associate into stable snoRNPs, and functional advantage of the modifications to the ribosome. Thus, some sites of modification are conserved among yeast, plants and vertebrates while others are novel. rRNAs of different organisms differ in their rRNA expansion segments (regions of rRNA which are not conserved among species). The expansion segments can influence rRNA/ribosome structure such that the effects of particular modifications can be variable. This leads to differences in the modification patterns of rRNAs among different organisms.

In plants, polyploidy is an important influence of genome structure and gene evolution. It is estimated that around 70% of plant species have undergone hybridisation events during their history. Gene duplication through polyploidisation and intra- and inter-chromosomal rearrangements means that plants can often tolerate gene mutations which in single genes would otherwise be disadvantageous or detrimental and lost from the gene pool. Although there are examples of protein-coding gene variants with different structural or expression properties, the large number of snoRNA genes, in the box C/D snoRNA gene family, and the non-protein coding nature of the genes makes these genes an excellent model for analysing gene evolution mechanisms in plants.

Reference

[1] Leader, D.J., Clark, G.P., Watters, J.A., Beven, A.F., Shaw, P.J. & Brown, J.W.S. (1997). *EMBO J.* **16**, 5742-5751.