Structure of U2 small nuclear RNA genes of rice genome*

WANG Xiping(王喜萍), FENG Zhengyu (冯征宇), DONG Aiwu (董爱武),
SUN Chongrong(孙崇荣) and CAO Kaiming(曹凯鸣)
(Department of Biochemistry, School of Life Sciences, Fudan University, Shanghai 200433, China)

Received January 31, 2000; revised March 9, 2000

Abstract Three new U2 small nuclear RNA (U2snRNA) genes from rice (*Oryza sativa*) were isolated and sequenced. The upstream sequence elements, TATA elements, monocot-specific promoter elements and the conserved four-stem-loop secondary structure shared by all plant snRNA genes were identified . The rice U2snRNA genes are shown to be present as a multi - gene family in the genome. The homology of the rice U2snRNA genes is high, with the minimum of 80%. Two pairs of the U2snRNA genes with the same transcriptional direction are found to be linked. One pseudogene in rice was firstly discovered.

Keywords: **U2snRNA genes, gene structure, sequence analysis, rice(*Oryza sativa*).**

In eukaryotic cells, small nuclear RNAs (snRNAs) are involved in pre-mRNA splicing, and the genes encoding snRNAs are usually presented as gene families with certain number of genes linked. The amount of snRNA in plant cells is 10—100 times less than that of animal cells, but their genes in plant cells are more diversified[1]. Only UsnRNA gene families of several plants have been characterized, including 7 U2snRNA genes from *arabidopsis*[2], 8 U1snRNA genes from tomato[3] and 6 U2snRNA genes from potato[4]. The members of UsnRNA gene families in animal usually scatter at one locus in the genome with distance between genes ranging from several hundreds to several thousands of base pairs[5]. In plant, the structure of UsnRNA gene needs to be further studied. It has been known that monocots and dicots share high homology in the U2snRNA gene coding regions, while the upstream regulatory regions differ considerably[6, 7], this indicates that there are different regulation mechanisms of gene expression between these two types of plants. We previously reported the isolation and characterization of a rice U2snRNA gene[7], based on this work, we isolated in present study the other 13 positive clones containing rice U2snRNA genes using *arabdopsis* U2.2snRNA gene as the probe, among these clones there are two pairs of linked U2snRNA genes and one pseudogene. Five U2snRNA genes were sequenced and further characterized.

1 Materials and methods

1.1 Southern blot analysis

Genomic DNA was isolated from the etiolated seedling of rice Guang Lu Ai 4 (seeds were provided by Chinese Rice Research Institute). Of 10 μg *EcoR* I and *Hind* Ⅲ digested rice genomic DNA was separated on a 1% agarose gel by electrophoresis. The DNA was then transferred to Hybond N

* Project supported by the Chinese Rice Genome Project of Ministry of Science and Technology (Grant No. 101-09-05)

(Amersham) nylon membrane and probed with the [α-^{32}P]dATP-labeled *Hind* Ⅲ-*EcoR* Ⅰ fragment (about 500 bp) of *arabidopsis* U2.2snRNA gene.

1.2 Isolation of rice U2snRNA genes and sequence determination

The clones containing rice U2snRNA genes were isolated from a genomic library using *in situ* hybridization as described by Benton and Davis[8]. The probe used was the same as used in southern hybridization. The positive clones were completely digested with *EcoR* Ⅰ and *Hind* Ⅲ, then the fragments containing U2snRNA genes were cloned into the M13mp18/19 vector and sequenced by Sanger's method[9].

1.3 Linkage analysis of U2snRNA gene in rice genome

PCR method was used to analyze the linkage of rice U2snRNA genes. Two primers, U2d (5' TGC TGC AGA AAG GCC GAG AAA GGT AT 3') and U2u (5' ACG GAT CCA GTG TAG TAT CTG TTC 3') were designed based on the conserved sequence of plant U2snRNA genes[7]. PCR was performed using different combinations of the primers (U2d + U2u, U2d + U2d or U2u + U2u) to amplify the intergenic fragments from each positive clone isolated from the genomic library. The amplified fragments were analyed by restriction enzymes mapping, and partially sequenced at the both ends. The pattern of linkage was determined from the combinations of the primers, the sizes and partial sequences of the amplified fragments.

1.4 Sequence analysis of rice U2snRNA genes

Rice U2snRNA genes were analyzed with the FSTNSCAN program of the PCGENE6.9 package[1]. The sequence analysis included homology comparison between the coding regions, the variation of the upstream regulatory elements and the deduction of RNA secondary structures.

2 Results and discussions

2.1 Rice U2snRNA gene family

More than 10 bands were probed with the *arabidopsis* U2.2snRNA gene fragment when rice genomic DNA was completely digested. It suggests that U2snRNA genes present as a multi-gene family in rice genome(fig.1). According to the restriction maps of the individual rice U2snRNA genes and the facts that U2snRNA genes are linked to each other and there is no intron within the genes, we propose that there are at least 15 U2snRNA genes in rice.

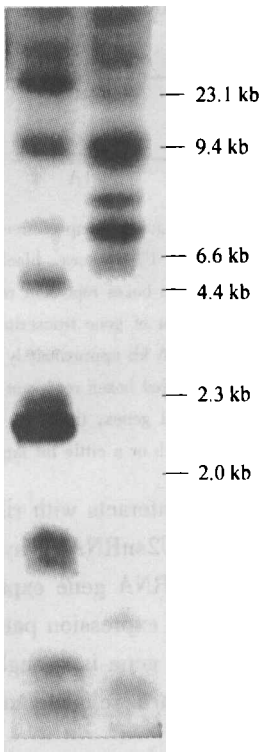

Fig. 1. Southern blot analysis of rice U2 snRNA gene. Rice genomic DNA was digested with *EcoR* Ⅰ (left lane) and *Hind* Ⅲ (right lane).

2.2 Linkage of rice U2snRNA genes

Thirteen positive clones hybridized with the *arabdpsis* U2.2 snRNA

1) Intelligenetics company

gene probe were isolated from a rice genomic library, named as U2-1—U2-13 respectively. Their restriction maps differ from each other (data not shown). Five clones (U2-1, U2-3, U2-4, U2-5 and U2-6) with the greatest variation were selected for further analysis. The sequence analysis shows that U2-1, U2-4 and U2-6 contain two linked U2snRNA genes, while U2-3 and U2-5 contain only one gene. U2-3 has the same sequence of U2-1A. PCR-based linkage analysis indicates that the distance between the linked genes is the same in U2-4 and U2-6, which is about 1.6 kb. In addition, the restriction maps of the amplified intergenic fragments from these two clones are identical, and the linked U2snRNA genes in both clones are arranged in tandem repeat. We consider that these two pairs of genes belong to one subgroup since the intergenic sequence is usually specific. We were unable to determine the distance between the two linked genes in U2-1. It is probably because these two genes are located apart and the ordinary PCR method cannot amplify such a large intergenic fragment. In consistent, no single positive band smaller than 10 kb in U2-1 was obtained in the restriction map. Moreover we observed different sizes of fragments when using rice genomic DNA as the template to analyze the linkage pattern of U2snRNA genes, besides an 1.6 kb band, a 3.2 kb and a 3.5 kb fragments were all observed. This provides an evidence that there are other types of linkage for U2snRNA genes in rice genome (figure 2).

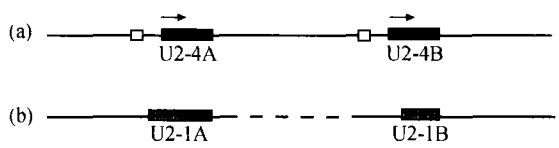

Fig. 2. Linkage map of rice U2 snRNA genes. (a) Linkage map of rice U2-4 genes, blackboxes represent coding region of genes, open boxes represent regulatory region, the arrows show the direction of gene transcription, the space between the two genes is 1.6 kb approximately. (b) Linkage map of rice U2-1 genes, shaded boxes represent coding region and regulatory region of U2-1 genes, the space between the two genes is estimated at 1.6 kb or a eittle bit larger.

2.3 Sequence analysis of rice U2snRNA genes

The homology of coding regions of U2-1A, U2-4A, U2-4B and U2-5 genes is 85.1%, while the homology between linked U2-4A and U2-4B is 95.5%. All the four U2snRNA genes share the typical features of plant U2snRNA genes[7]. Their first 124 nucleotides (nt) at 5' end are identical, and the difference in the sequence is only found at 3' end of the gene. Their RNA-RNA interaction region located at 5' end is highly conserved, while the stem-loop region which is located at 3' end and interacts with ribosome proteins varies greatly. It suggests that the binding of the proteins to different U2snRNAs may form specific splicesome performing different splicing[10]. The studies on plant UsnRNA gene expression by Hanley and schuler[11,12] have shown that at each developmental stage, the expression patterns of U2snRNA genes are distinct. Therefore, the expression of individual U2snRNA gene is strongly associated with the splicing of pre-mRNAs which are specifically expressed at different developmental stages.

Two elements specific for plant UsnRNA genes, USE (upstream sequence element) and TATA-like element, were identified from U2-1A, U2-4A, U2-4B and U2-5 genes, they are located at −70 nt and −30 nt respectively. In the upstream region of USE, there are 2 or 3 conserved G/ACCCA/G sequences located between −180 nt and −120 nt (fig.3). This element is termed MSP (monocot-specific promoter). Each of U2-1A, U2-4A and U2-5 contains two MSPs, with the sequence of GGCCCA. In U2-4B, three MSPs were identified, two with the sequence of GGCCCA and one with the sequence of AGCCCG. The difference at upstream regulatory regions between monocots

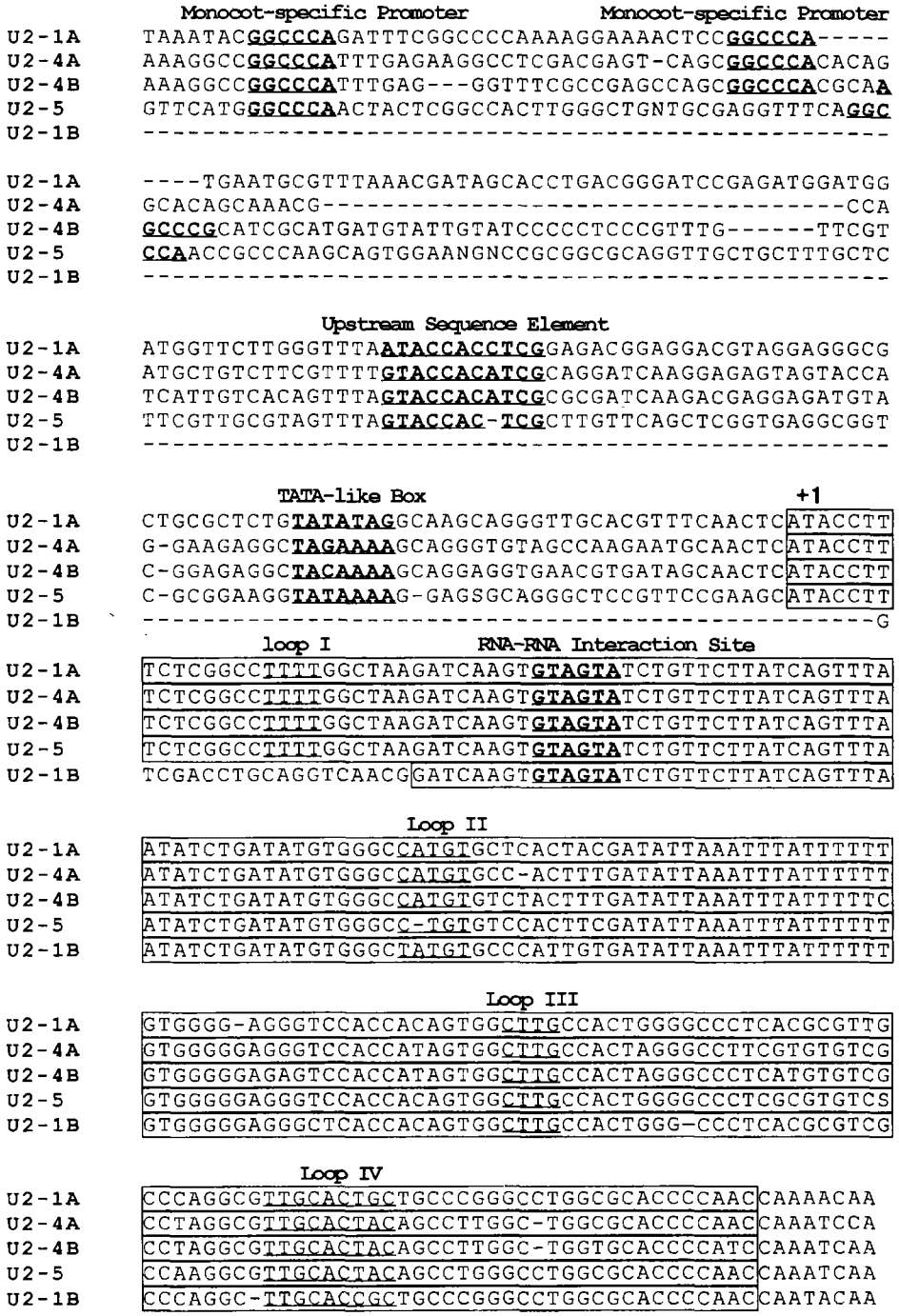

Fig. 3. Sequence homology of different U2snRNA from rice. The predicted coding region is within the frame, bold letters represent regulation elements or domains of genes.

and dicots indicates the difference in transcriptional machinery. It has been reported that monocot UsnRNA genes (even without MSP) can be expressed in dicot protoplasts. However, dicot UsnRNA genes generally cannot be expressed in monocot protoplasts[12]. Therefore, MSP is essential for the

expression of monocot UsnRNA genes and there is significant difference in the transcriptional factors required by UsnRNA gene expression in monocots and dicots.

2.4 Rice U2snRNA pseudogene

Among five identified rice U2snRNA genes in this study (fig. 3), U2-1B has no upstream regulatory region, besides, there is a deletion of 25 nt at its 5' end of the coding region, and its 3' end is distinct from any of the known rice U2snRNA genes. The pseudogenes are generally presented in UsnRNA gene families, which is the result of gene rearrangement[5,13—15]. The common features of these pseudogenes include the absence of upstream regulatory region and partial deletion of the gene. In our case, U2-1B has all of these properties so we conclude that U2-1B is a pseudogene.

The diversity of U2snRNA genes indicates the complexity of pre-mRNA splicing in plant development. Studies with the aim to characterize plant UsnRNA gene families will shed light on the molecular mechanisms in gene regulation during plant development.

References

1 Ro-Choi, T.S., Nuclear snRNA and nuclear function, *Crit. Rev. Eukaryote Gene Expression*, 1999, 9: 107.

2 Bankan, P., Filipowicz ,W., Structure of U2 snRNA genes of *Arabidopsis thaliana* and their expression in electroporated plant protoplasts, *The EMBO Journal*, 1988, 7: 791.

3 Abel, S., Kiss, T., Solymosy, F., Molecular analysis of eight U1 RNA gene candidates from tomato that could potentially be transcribed into U1 RNA sequence variants differing from each other in similar regions of secondary structure, *Nucl. Acids. Res.*, 1989, 16: 1042.

4 Waugh, R., Clark, G., Baux, P. *et al.*, Sequence and expression of potato U2 snRNA genes, *Nucl. Acids. Res.*, 1991, 19: 249.

5 Dahoberg, J.E., Lund, E., The genes and transcription of the major small nuclear RNAs, *in Collection of Small Nuclear Ribonucleoprotein Particles* (ed. Birnstiel,M.), Berlin: Springer-Verlag, 1988, 38: 70.

6 Musci, M. A., Egeland, D. B., Schuler, M. A., Molecular comparison of monocot and dicot U1and U2 snRNAs, *The Plant Journal*, 1992, 2: 589.

7 Xiping ,W., Zheng, Y. F., Chong, R. S. et al., Isolation and structure analysis of the rice U2 snRNA gene, *Acta Bot. Sin.*(in Chinese), 1997, 39: 601.

8 Benton, W. D., Davies, R. W., Screening λgt recombinant clones by hybridization to single phages *in situ*, *Science*, 1977, 196: 180.

9 Sanger, F., Nicxlen, S., Coulson, A.R., DNA sequencing with chain terminating inhibitors, *Proc. Natl. Acad. Sci.*, 1977,74: 5463.

10 Connelly, S., Marshallsay, C., Leader, D. et al., Small nuclear RNA genes transcribed by either RNA polymerase II or RNA polymerase III in monocot plants share three promoter elements and use a strategy to regulate gene expression different from that used by their dicot plant counterparts, *Molecular and Cellular Biology*, 1994, 14: 5910.

11 Hanley, B. A., Schuler, M. A., cDNA cloning of U1, U2, U4 and U5 snRNA families expressed in pea nuclei, *Nucl. Acids. Res.*, 1991, 19: 1861.

12 Hanley, B. A., Schuler, M. A., Developmental expression of plant snRNAs, *Nucl. Acids. Res.*, 1991, 19: 6319.

13 Kiss, T., Abel, S., Solymosy, F., A plant pseudogene for U1 RNA, *Plant. Mol. Biol.* 1989, 12: 709.

14 Kiss, T., Solymosy, F., Molecular analysis of a U3 RNA gene locus in tomato: transcription signals, the coding region, expression in transgenic tobacco plants and tandemly repeated pseudogenes, *Nucl. Acids. Res.*, 1990, 18: 1941.

15 Waugh, R., Clark, G., Brown, J. W. S., Sequence variation and linage of potato U2snRNA-encoding genes established by PCR, *Gene*, 1991, 107: 197.