GENOME-WIDE COMPUTATIONAL ANALYSIS OF SMALL NUCLEAR RNA GENES OF *ORYZA SATIVA* (INDICA AND JAPONICA)

M.SHASHIKANTH, A.SNEHALATHARANI, SK. MUBARAK AND K.ULAGANATHAN
Center for Plant Molecular Biology, Osmania University, Hyderabad, Andhra Pradesh, India.
kulaganathan@yahoo.com

Genome-wide computational analysis for small nuclear RNA (snRNA) genes resulted in identification of 76 and 73 putative snRNA genes from indica and japonica rice genomes, respectively. We used the basic criteria of a minimum of 70 % sequence identity to the plant snRNA gene used for genome search, presence of conserved promoter elements: TATA box, USE motif and monocot promoter specific elements (MSPs) and extensive sequence alignment to rice / plant expressed sequence tags to denote predicted sequence as snRNA genes. Comparative sequence analysis with snRNA genes from other organisms and predicted secondary structures showed that there is overall conservation of snRNA sequence and structure with plant specific features (presence of TATA box in both polymerase II and III transcribed genes, location of USE motif upstream to the TATA box at fixed but different distance in polymerase II and polymerase III transcribed snRNA genes) and the presence of multiple monocot specific MSPs upstream to the USE motif. Detailed analysis results including all multiple sequence alignments, sequence logos, secondary structures, sequences etc are available at http://kulab.org

1. Introduction

Most eukaryotic protein coding genes contain non-translated intronic sequences that are excised from the primary transcripts (pre-mRNA) by the process of splicing.[1] Splicing of nuclear pre-mRNA involves sequential trans-etherification reactions, which take place in a large complex called the spliceosome. The spliceosome is composed of five snRNAs, U1, U2, and U4, U5 and U6 as well as many proteins.[2] Some of these proteins are tightly associated with the snRNAs, forming small nuclear ribonucleoproteins (snRNPs) which assemble in a stepwise manner onto the pre-mRNA to form the spliceosome. Besides the snRNP subunits, a large number of proteins perform various functions during the splicing reaction.[3]

Four of the five snRNAs base pair with the pre-mRNA at various times during the splicing reaction. Interaction between U1-snRNA and the 5’ splice site and U2-snRNA and branch site are well established.[4,5,6,7] U5-snRNA interacts with exonic sequences immediately 5’ and 3’ to the splice junctions, while the U6-snRNA interacts with the 5’ splice site.[8,9,7,10] The snRNA genes are present in multiple copies and are synthesized from independent transcription units, which are transcribed either by RNA polymerase II or III. These genes differ from other classes of genes in having unique transcriptional factors.[11] Though transcribed by two polymerases, their promoters are structurally related.[11,12,13] The snRNA promoters are highly conserved within a species but show variation in different organisms except for certain conserved motifs. So analysis of snRNA genes and their expression may help in understanding the eukaryotic transcription.

Majority of the splicing related genes are known in human,[14] Yeast,[15] *Drosophila*[16] and *Arabidopsis thaliana.*[17] Unlike other organisms, in plants, the knowledge on splicing is scanty due to the non-availability of *in vitro* splicing system. In spite of this, the powerful comparative genomic approach can be employed to predict the genes involved in the splicing process which can be analyzed *in silico* to get a comprehensive picture of the splicing process in an organism like rice. Such analysis will help in prioritizing and better planning for wet-lab experiments to understand the process of splicing. Rice is a unique crop plant used as a model plant for genomic research due to the relatively small genome (ca. 450 MB) and the availability of two genomes, indica and japonica which facilitate comparative analysis. We carried out extensive search of indica and japonica rice genomes using other plant snRNA gene sequences and human splicing associated factors as the basis and predicted a total of 149 snRNA genes and more than 800 protein coding genes associated with splicing from indica and japonica genomes together. Here we provide information about the predicted snRNA genes from indica and japonica genomes and their analysis.

2. Methods

A total of 127551 scaffolds from the draft indica rice genome (super hybrid rice) sequence, downloaded from the Beijing Genomics Institute (http://btn.genomics.org.cn/rice/) were used for the analysis of Indica rice genome. Where as, genome sequences accessed from the Rice Genome Project (RGP) was used for the analysis of japonica rice genome. Rice genomic sequences (indica and japonica) in the form of contigs, cDNA and proteins were collected and a local database was created which can be searched by the key word, accession number or by homology search using BLAST[18] algorithm.

Sequences of splicing associated snRNAs of different plants were collected from various online databases and used to query the local database using the local blast search tool with different parameters. The hits with more than 70% identity to the query sequence were collected and used for the analysis. Based on the BLAST alignment, the open reading frame spanning the aligned sequence was extracted from the contig as the putative gene and verified by pairwise alignment with the query sequence using L-Align.[19]

To validate the prediction, the predicted gene sequences were aligned to other known snRNA genes (using ClustalW) from plants, human, mouse and *Drosophila* and the multiple sequence alignment was given as the first validation. Next, the predicted genes were searched against NCBI EST database and if rice plant ESTs were showing significant stretch of alignment with the predicted genes, then the Unigene cluster number was given as the second validation. The third and final validation of the predicted gene was carried out by analysis of the promoter sequence (upstream 500 bp). The extracted upstream sequences were aligned using ClustalW and the conserved TATA box and USE motif sequences, their location and the distance between them were analyzed. Location of monocot specific elements[20] from the upstream 500 bp sequence of putative SnRNA genes was carried out manually using the MSP consensus sequence, RGCCCR and the motif search function of Bioedit sequence analysis tool.

Validated putative snRNA gene sequences were further used for constructing the secondary structure of snRNAs based on published plant RNA secondary structures. Using predicted secondary structure and the multiple sequence alignment as the basis, the nucleotide variations and conserved features were identified. Using the online sequence logo rendering tool, the variations in different stem loop regions of snRNAs were analyzed and displayed.

3. RESULTS AND DISCUSSION

3.1. *snRNA genes in the draft rice genome*

Eukaryotic cells contain a large population of snRNAs, which play a major role in splicing of genes that lead to regulation of gene expression, differentiation and development of an organism. snRNAs are broadly classified into two different categories based on the transcription regulation. snRNAs such as U1, U2, U4 and U5 are transcribed by pol II, where as, U3 and U6 are Pol III transcribed. Using comparative genomic approach, we identified a total of 149 putative snRNA coding genes from the rice genome and validated them by identifying upstream conserved motifs and EST matching.

3.2. *Pol II specific snRNA genes*

To predict pol II specific snRNA genes (U1, U2, U4 and U5) in rice genomes, we used the pol II transcribed plant snRNA genes from *Zea mays*, *Triticum aestivum* and *Arabidopsis thaliana*. The U1 (S72336) and U2 snRNA (S72237) gene sequences were from *T.aestivum,* the U4 snRNA gene sequence (X67145) was from *A. thaliana* and the U5 snRNA gene sequence (Z14995) was from *Zea mays*. These sequences were used as queries for BLAST search against rice genomes with 0.01 e-value as the cut off score. In the BLAST results, the alignments with at least 70 % identity to the query sequence were chosen to identify the potential location of snRNA homologues. This resulted in identification of 58 and 52 putative pol II specific snRNA genes from indica and japonica genomes, respectively. There were 10, 23, 6 and 25 paralogs of U1, U2, U4, and U5 snRNA genes, respectively in indica genome. While in japonica genome, there were 16, 23, 6 and 13 paralogs of U1, U2, U4 and U5 snRNA genes, respectively.

The predicted U1-snRNA genes of indica genome ranged in between 160-164 nucleotide long, while, 157-163 nucleotides long in japonica genome. These genes showed 88-93 % identity to the query sequence. They also showed extensive sequence identity with human and *A. thaliana* snRNA genes (Table 1). All the predicted pol-II snRNA genes were finally validated by analyzing their upstream sequences. The genes that lack the characteristic USE motif were discarded (data regarding these genes not included). All the predicted pol-II transcribed U1-snRNA genes, in their upstream sequence contained the characteristic TATA box and one USE motif at a distance of 30-34 nucleotides upstream of the TATA box (Figure 2). The characteristic distance between the USE motif and the TATA box is a plant specific feature conserved in both indica and rice genomes. Unlike plants, the human, mouse and *Drosophila* snRNA genes transcribed by pol II are TATA box less and pol-III transcribed genes has the TATA box in their promoters. There were 1-5 monocot specific promoter elements found upstream to the USE motif of predicted U1-snRNA gene promoters (Table 1).

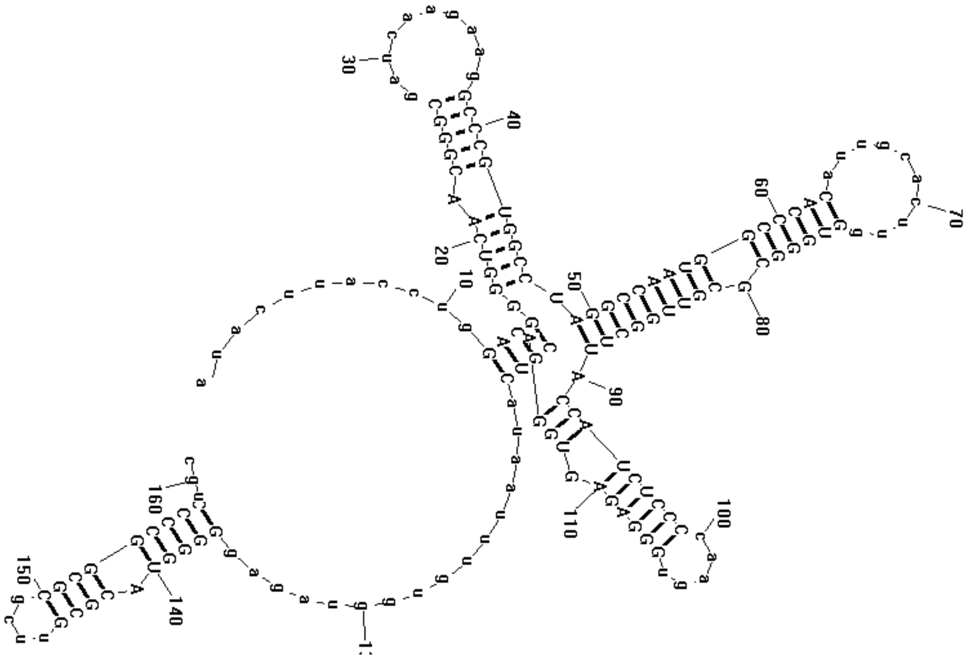

Fig.1. Predicted secondary structure of rice U1-snRNA gene consensus sequence

The secondary structure of *Phaseolus vulgaris* U1-snRNA[21] as the basis rice U1-snRNA secondary structure was drawn with the consensus sequence derived from the nucleotide sequence alignment of indica and japonica U1 snRNA sequence and with the help of RNA structure tool version 4.2. Overall, the structure is conserved in all the rice sequences predicted with variations seen in the stem regions, while the single stranded regions are well conserved (Fig-1). A more detailed explanation regarding the secondary structure is available in our research extension page at http://kulab.org

There were 23 putative U2-snRNA genes in both indica and japonica genomes which were 171-195 nucleotides in length. These genes showed 83.8 to 92.8 per cent sequence identity with the query sequence (*T.aesativum* U2-snRNA gene; S72337). The predicted

U2-snRNA genes in japonica genome showed extensive sequence identity with human (69.1-75.3 %) and *A.thaliana* (83.3-92.8 %) U2-snRNA sequences. EST sequence which belong to the Unigene cluster Os.11638 aligned with the predicted U2-snRNA genes. All the 46 genes contained the characteristic USE motif at 30-34 nucleotides upstream to the TATA box. Upstream to the USE motif there were 1 – 8 Monocot Specific Promoter elements present in the promoters of predicted U2-snRNA genes (Table 1). The secondary structure was drawn using the maize model[22] which showed that the overall structure is conserved. The branch point binding site (GUAGUA) and Sm site (AUUUUUUG) are absolutely conserved in all the 46 genes.

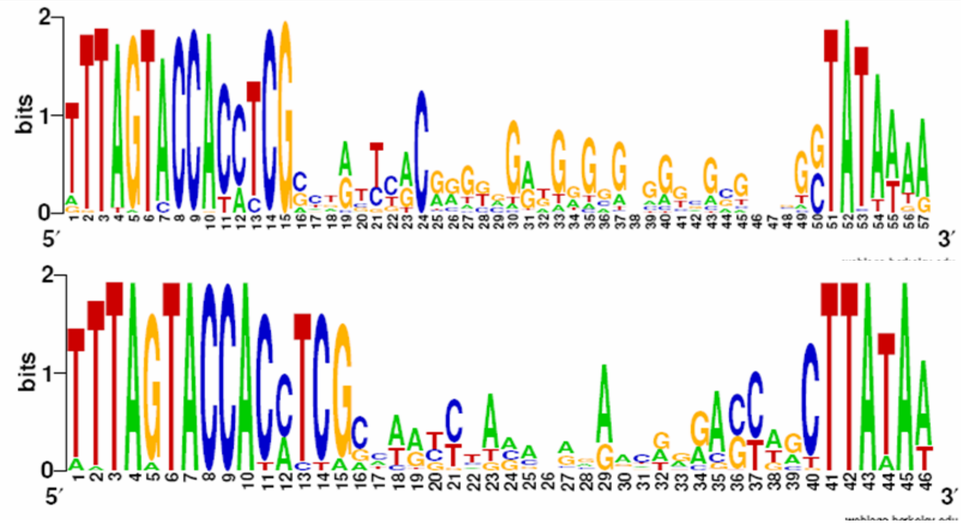

Figure 2. Sequence logos displaying the region between TATA box and the USE motif of pol-II transcribed (U1, U2, U4 and U5)(above) and pol-III transcribed (U3 and U6) (below) snRNA genes of *Oryza sativa* (indica and japonica together).

Each indica and japonica rice genome contained seven U4-snRNA genes (159-160 nucleotide length) showed 80.9 - 83.8 % identity with the query, *A.thaliana* U4-snRNA gene (X67145). All the 14 putative U4-snRNA genes aligned with EST sequences that belong to the Unigene cluster Sbi.6812. The characteristic TATA box and USE motifs were found upstream of all 14 genes. Further upstream to USE motif 1-6 MSPs found (Table 1) (Figure 3). The interacting secondary structure of U4 snRNA consensus and U6 snRNA consensus was made using secondary structure of interecting U4-U6 snRNAs of human.[23] There is overall conservation of the secondary structure of U4-snRNA (Secondary structure image available at http://kulab.org) with three stem loops and three single stranded regions. Two regions were predicted to interact with the U6-snRNA.

A total of 27 U5-snRNA genes were predicted, of which 12 and 15 genes were found in indica (104-107 nucleotides) and japonica (103-107 nucleotides) genomes, respectively. These putative U5-snRNA genes showed 78-95 % sequence identity with maize U5-snRNA gene (Z14995) as the query. These genes aligned with ESTs belonging to the Unigene cluster Os.37121. Upstrem 500 sequences of these putative genes showed the TATA box and the charcteristic USE motif which is located 35 nucleotides upstream to the TATA box. There were 1-9 MSPs located upstream of the USE motif in the promoters of the predicted U5-snRNA genes.

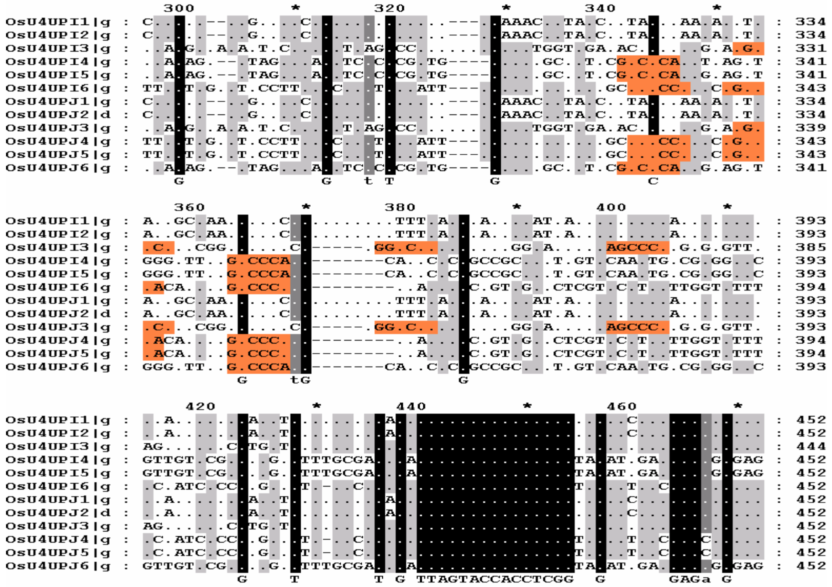

Figure 3. Multiple sequence alignment of upstream-500 sequence of U4 genes showing the conserved USE motif and MSPs.

The secondary structure was drawn using the U5-snRNA consensus sequence which showed that 11 nucleotides of loop-1 is conserved with that of human and yeast U-5 snRNAs. The Sm site and the 3' loop which are essential for Sm protein binding and Cap-tri-methylation are conserved as in human sequences.

Keefe and Newman, 1998, carried out deletion analysis to study the importance of stem loop I of U5 snRNA in yeast. This loop I interacts with 5' exon before first step of m-RNA splicing and with the 5' and 3' exons following the first step. The size of loop I was found to be critical for proper alignment of exons for the second catalytic step of splicing.[24] Rice U5 snRNA showed remarkable conservation of 11 residues (CGCCTTTTACT) among themselves and with other U5 snRNAs indicating that loop size is conserved in rice snRNAs.

3.3. *Pol III specific U3 and U6 snRNA genes*

Homology search for pol III transcribed snRNA genes, U3 and U6 was made using *T.aestivum* U3 (X63065.1) and U6 (X63066) genes as query sequences. This resulted in identification of seven and eight putative U3 snRNA genes in indica and japonica genomes, respectively. These putative U3-snRNA genes showed 80-86% identity and U6 snRNA has 91-99% sequence identity with the query sequence used There were 12 and 15 putative U6 snRNA genes in indica and japonica genomes, respectively (Table 1).

Similar to pol II specific genes, the predicted U3 and U6 snRNA genes of rice also contained three most important upstream promoter elements, the Upstream Sequence Element (USE, TTAGTACCACCTCG), the TATA box and one or more MSPs. The USE and TATA box were 23-26 bp apart (Figure 2) and the latter one lies 21-25 bp upstream of the transcription start site.

The U6 motif had the consensus GTTTAGTACCACCTCG and was present exactly 25 nucleotides upstream of TATA box. In few genes there is a 5 nucleotide deletion in between the TATA and USE motifs. The 14 putative U3 snRNA genes showed the conserved boxes, A, A^0, B, C and D which were identified in plant U3 snRNA genes[13]. There are no similar conserved boxes seen in U6 snRNA genes but they possessed extensive conserved regions. Both U3 and U6 snRNAs contained many copies of the MSP elements in their promoters, upstream to the USE motif. Unlike pol-II transcribed snRNA genes, in U3 and U6 there were fewer MSPs found (Table 1). Further, the predicted U3 and U6 snRNAs were found to align with EST sequences belonging to the Unigene clusters Os.39667 and At.47201, respectively.

Homology searching of indica and japonica genomes with plant snRNA sequences as queries resulted in the prediction of a number of putative snRNA genes. These putative regions were further analyzed only if they showed at least 70 % sequence identity with the query snRNA sequence used. Further, those predicted genes that lacked the characteristic USE motif were discarded which resulted in identification of 76 and 73 putative snRNA coding genes from indica and japonica rice genomes, respectively. Analysis of these snRNA genes showed that they are conserved as in case of other organisms like human, with respect to the sequence and overall secondary structure. These genes showed the following plant specific features: a) Conform to the overall predicted secondary structures of plant snRNA genes. b) All snRNA genes possessed TATA box in their promoters. c) Plant specific USE motif was found in the promoter upstream to the TATA box. d) Multiple MSP elements were found in the promoter upstream to the USE motif. e) The spacing between the USE and TATA box is the major determinant of pol specificity of plant snRNA genes. In the promoters of Pol II-specific genes the USE motif and the TATA box are centered approximately four DNA helical turns apart (30-34 bp), while in pol III-specific genes these elements are positioned one helical turn closer (20-25 bp). The TATA box is present 23-28 bp upstream in pol II genes and 21-25 bp upstream in pol III genes from the coding region. With respect to rice specific features, there are specific nucleotide variations seen in the snRNAs especially in the single stranded regions and stem of the various stem loops. Wherever variation in nucleotide is observed in the stem of loops, mostly they are compensatory (data not shown). There is variation in number of copies of different snRNA genes, between the two rice genomes with respect to U1, U5 and U6 while the number of copies with respect to U2, U3 and U4 is constant in both genomes. U5 snRNA, especially showed huge variation from 25 copies in indica genome to 13 copies in japonica genome which may be probably due to duplication in indica genome or deletion in japonica genome.

Further more, the predicted secondary structure is similar to the known structure of their counter parts in other plants,[25] indicating the structural conservation. Some of the predicted genes did not have the conserved USE in the promoter region (data not included) but their structure is absolutely conserved. These genes may be pseudogenes and may not be expressed. The Pol II and Pol III snRNA gene promoters of rice, in addition to mammals, frogs, dicot plants, and possibly nematodes,[26] represent one more example of promoters which are remarkably similar despite being recognized by two different RNA Polymerases. Our finding in rice regarding the conserved promoter regions further support that Pol II and Pol III transcription machineries are highly conserved.

Previously, snRNA genes have been predicted mostly from japonica genome of which the Rfam database had the maximum number of snRNAs of japonica rice.[27] Sequence alignment of those snRNAs with the snRNAs predicted in this work showed that many of them aligned perfectly without much variation (data not shown). But our prediction, though started initially with homology search of the genomes, is superior as we did not rely on simple sequence identity alone to predict the gene. We validated the predicted sequences by a) first with the presence of USE motif in the upstream sequences, then b) by the presence of TATA box in the upstream sequences and the distance between the USE motif and TATA box, c) by the presence of multiple MSP elements in the promoters, upstream to the USE motif, d) by alignment to expressed sequence tag sequences, and e) by conservation of structural features (stem loops, protein binding sites and interaction sites between snRNAs) based on the secondary structure.

References

1. C. Burge, T. Tuschl and P.A. Sharp. Splicing of precursors to mRNAs by the spliceosome. In R. F. Gesteland, T. R. Cech and J. F. Atkins (eds), *The RNA World*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 525-560, 1999.

2. M. J. Moore, C. C. Query and P.A. Sharp. Splicing of precursors to mRNAs by the spliceosome. In R. F. Gesteland, T. R. Cech and J. F. Atkins (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 303-357, 1993.
3. C. L. Will. and R. Luhrmann. snRNP structure and function. In Krainer, A.R. (ed.), *Eukaryotic mRNA processing.* IRL press at Oxford University Press, Oxford, UK, 130-173, 1997.
4. A. J. Newman and C. Norman. Mutations in yeast U5 snRNA alter the specificity of 5' splicing – site cleavage. C*ell*, 65: 115-123, 1991.
5. A. J. Newman and C. U5 snRNA interacts with exon sequences at 5' and 3' splicing sites. *Cell*, 68: 743-754, 1992.
6. J. J. Cortes, E. J. Sontheimer, S. D. Seiwert and J. A. Steitz. Mutations in the conserved loop of human U5 snRNA generate use of novel cryptic 5' splice sites in vivio. *EMBO J*, 12: 5181-5189, 1993.
7. E. Sontheimer and J. A. Steitz. The U5 and U6 small nuclear RNAs as active site components of spliceosome. *Science*, 262: 1989-1996, 1993.
8. H. Sawa and J. Abelson. Evidence for base-pairing interaction between U6 small nuclear RNA and 5' splicing site during the splicing reaction in yeast. *PNAS*, 89: 11269-11273, 1992.
9. D. A. Wassarman and J. A. Steitz. Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. *Science*, 257: 1918-1925, 1992.
10. J. S. Sun and J. L. Manley. A novel U2-U6snRNA structure is necessary for mammalian mRNA splicing. *Genes and Dev.*, 9: 843-854, 1995.
11. N.Hernandez. Transcription of vertebrate snRNA genes and related genes. In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 1, pp. 281-313, 1992.
12. P.C.H.Lo and S.M.Mount. *Drosophila melanogastor* genes for U1 snRNA variants and their expression during development. *Nucleic Acids Res.*, 18:6971-6979, 1990.
13. F.Waibel, W.Filipowicz. RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. *Nature*, 346:199–202, 1990.
14. J. Rappsilber,U. Ryder,A. I. Lamond, and M. Mann. Large-Scale Proteomic Analysis of the Human Spliceosome. *Genome Research*, 12:1231–1245, 2002.
15. C. W. Pikielny and M. Rosbash. Specific small nuclear RNAs are associated with yeast spliceosomes. *Cell*, 20: 869-877, 1986.
16. M. M. Stephen and K. S. Helen. Pre-messenger RNA processing factors in the *Drosophila* Genome. *The J. Cell Bio.*, 150: F37-F43, 2000.
17. B. B. Wang and V. Brendel. The ASRG database: identification and survey of *Arabidopsis thaliana* genes involved in pre-mRNA splicing. *Genome Biology*, 5: R102-102.23, 2004.
18. S. F.Altschul, W.Gish, W.Miller, E.W.Myers and D.J.Lipman. Basic local alignment search tool. *J. Mol. Biol.,* 215: 403-410, 1990.
19. X. Huang. and W. Miller. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* 12:337-357, 1991.
20. S. Connelly,C.Marshallsay,D.Leader, J.W. S. Brown, and W.Filipowiczi. 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. *Mol. Cell. Biol.,* 14:5910-5919, 1994.

21. V. L. Van Santen and R. A. Spritz. Nucleotide sequence of a bean (Phaseolus vulgaris) U1 small nuclear RNA gene: Implications for plant pre-mRNA splicing. *PNAS*, 84: 9094-9098, 1987.
22. J.W.S. Brown and R. Waugh. Maize U2 snRNAs: gene sequence and expression. *Nucleic Acid Research*, 17: 8991- 9001, 1989.
23. A. Mougin, A.Gottschalk, P. Fabrizio, R.Luhrmann and C. Branlant. Direct Probing of RNA Structure and RNA-Protein Interactions in Purified HeLa Cell's and Yeast Spliceosomal U4/U6.U5 Tri-snRNP Particles. *J. Mol. Biol.* 317: 631-649, 2002.
24. R. T. O'Keefe and A. J. Newman. Functional analysis of U5 snRNA loop I in the second catalytic step of yeast pre m-RNA splicing. *EMBO J.* 17: 565-574, 1998.
25. A.A.Patel and J.A.Steitz. Splicing double: Insights from second spliceosome. *Nat. Rev. Mol. Cell Biol.*, 4:960-970, 2003.
26. G.J.Goodall and W.Filipowicz. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *EMBO J.*,10: 2635–2644, 1991.
27. S. G. Jones, A. Bateman, M. Marshall, A. Khanna and S. R. Eddy. Rfam: an RNA family database. *Nucleic Acids Res*., 33:439-441, 2003.

Table -1. Putative snRNA genes in indica and japonica genomes (Detailed tables for each category of snRNAs are available online at http://kulab.org)

Name of the gene	No. of Copies	Length of the predicted gene	Percent identity to the query	EST matching: Unigene cluster number	Percent identity to Human gene	Percent identity to *A.thaliana* gene	Upstream USE motif	Number of MSPs found
Japonica genome								
U1 snRNA	16	157-163 S72336	(88-93) Ta.28789	J00318 (63-68)	X53175	(72-78)	TTTAGTACCACCTCG	1-5
U2 snRNA	23	171-195	S72337 (87-92)	Os.11638	X01408 (69-75)	X52312 (80-84)	TTTAGTACCACCTCG	1-7
U3 snRNA	6	228-311	X63065 (81-90)	Os.39667	X14945 (52-58)	X52629 (64-73)	TTTAGTACCACCTCG	1-3
U4 snRNA	6	159	X67145 (81-83)	Sbi.6812	X59361 (63-71)	X67145 (81-83)	TTTAGTACCACCTCG	1-5
U5 snRNA	13	107-103	Z14995 (86-91)	Os.37121	X04293 (72-81)	X13012 (86-96)	TTTAGTACCACCTCG	1-9
U6	15	102-105	X52315 (95-99)	At.47201	X07425 (81-86)	X63066 (96-99)	TTTAGTACCACCTCG	1-4
Indica genome								
U1 snRNA	10	160-164	S72336 (88-93)	Ta.28789	J00318 (66-69)	X53175 (70-77)	TTTAGTACCACCTCG	1-5
U2 snRNA	23	192-194	S72337 (94-96)	Os.11638	X01408 (70-84)	X52312 (72-89)	TTTAGTACCACCTCG	1-8
U3 snRNA	6	232-325	X63065 (82-90)	Os.39667	X14945 (52-54)	X52629 (66-75)	TTTAGTACCACCTCG	1-3
U4 snRNA	6	87-92	X67145 (93-96)	Sbi.6812	X59361 (70-73)	X67145 (93-96)	TTTAGTACCACCTCG	1-6
U5 snRNA	25	104-107	Z14995 (89-95)	Os.37121	X04293 (75-83)	X13012 (83-97)	TTTAGTACCACCTCG	1-6
U6 snRNA	12	97-101	X52315 (96-99)	At.47201	X07425 (68-86)	X63066 (76-85)	TTTAGTACCACCTCG	1-4