

水稻基因组中 U2snRNA 基因的结构分析*

王喜萍 冯征宇 董爱武 孙崇荣 曹凯鸣

(复旦大学生命科学院生物化学系, 上海 200433)

摘要 对水稻 (*Oryza sativa*) U2snRNA 基因的结构分析结果表明单子叶植物水稻的 U2snRNA 基因以多基因家族的形式存在, 且 U2snRNA 基因中存在连锁现象. 3 个新的水稻 U2snRNA 基因和已报道的 FDRGU2-3 一样, 编码区上游除含有植物 UsnRNA 基因共有的 USE 和 TATA 元件外, 还含有单子叶植物 UsnRNA 所特有的 MSP 元件, 且具有保守的四茎环二级结构, 基因成员间的同源性最低为 85.1%. 其中两对基因连锁, 且转录方向相同, 此外还发现一例假基因的存在.

关键词 U2snRNA 基因 基因家族 结构分析 水稻

在高等真核生物中, 与 pre-mRNA 剪接相关的 UsnRNA 基因 (U1, U2, U4, U5 及 U6) 通常以多基因家族的形式存在, 且不同数目的基因常在一定程度上存在连锁. 植物细胞核中 snRNA 的丰度比动物的 snRNA 低 10~100 倍, 但个别成员的结构变化类型却远远多于动物^[1]. 在植物中, 被深入研究过的 UsnRNA 的基因家族并不多, 已知拟南芥有 7 个 U2snRNA 基因^[2], 番茄有 8 个 U1snRNA 基因^[3], 马铃薯有 6 个 U2snRNA 基因^[4]. 动物 UsnRNA 基因家族常常分散在基因组的某一区段, 间距几百到几千碱基对^[5]. 而植物 UsnRNA 基因在基因组的分布情况尚未见报道. 单子叶植物与双子叶植物的 U2snRNA 基因在编码区部分有较高同源性, 但在上游调控区的结构上不同^[6,7], 这意味着它们的 UsnRNA 基因在表达调控上会有所区别. 在完成水稻一种 U2snRNA 基因结构分析的基础上^[7], 为了解水稻中 U2snRNA 基因家族的组成情况、结构特征以及在基因组中的分布等, 我们以拟南芥 U2snRNA 基因为探针, 从水稻基因文库中分离了 13 个 U2snRNA 基因克隆, 在物理图谱分析基础上选择其中 5 个差异明显的 U2snRNA 基因进行结构测定, 结果表明它们都具有典型的 U2snRNA 基因的结构特征, 但又各不相同, 并在水稻中发现至少存在两对连锁的基因. 同时还发现在水稻中存在 U2snRNA 假基因, 这些结果显示出水稻 U2snRNA 基因家族成员结构较为丰富, 意味着水稻剪接机制的复杂性.

1 材料与方法

1.1 Southern 印迹分析

从水稻广陆矮 4 号 (种子由中国水稻研究所提供) 黄化苗中提取总 DNA, 各取 10 μ g 总 DNA 经酶切后在 1% 琼脂糖凝胶上电泳过夜, 电压为 1 V/cm, 将电泳后的凝胶进行 Southern 转

1999-08-17 收稿, 1999-11-15 收修改稿

* 国家水稻基因组研究计划资助项目 (编号: 101-09-05)

移,所用尼龙膜为 Hybond N (Amersham 产品). 以拟南芥 U2.2snRNA 基因的 *Hind*III-*Eco*R I 片段(约 500 bp)为探针,用 $[\alpha\text{-}^{32}\text{P}]$ dATP 标记后进行杂交和放射自显影.

1.2 水稻 U2snRNA 基因的分离及序列分析

以 $[\alpha\text{-}^{32}\text{P}]$ dATP 标记的拟南芥 U2.2snRNA 基因 *Hind*III-*Eco*R I 酶切片段为探针,按 Benton 和 Davis 的原位杂交法^[5],从水稻基因文库中筛选含 U2snRNA 基因的克隆^[2],制备重组 λ DNA,以 *Bam*H I, *Eco*R I, *Hind*III 等酶切进行酶谱分析, Southern 印迹分析后,将含有 U2snRNA 基因的酶切片段克隆于 M13mp18/19 载体后,进行序列分析.

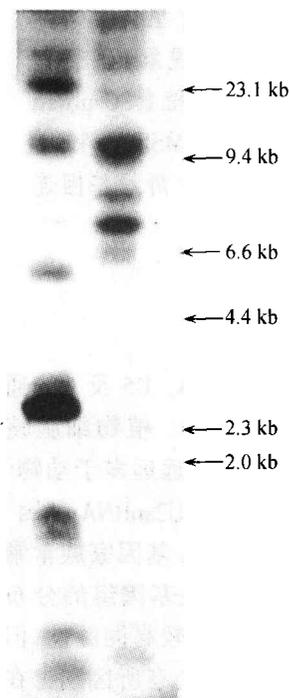


图 1 水稻 U2snRNA 基因的 Southern 印迹分析图谱
左边泳道为 *Eco*R I 酶切的水稻总 DNA,右边泳道为 *Hind*III 酶切的水稻总 DNA

1.3 水稻 U2snRNA 基因的连锁分析

以水稻基因文库中分离到的包含 U2snRNA 基因的重组 λ DNA (50 ng)为模板,根据植物 U2 基因的保守顺序设计并合成一对引物: U2d 和 U2u (5'-TGCTGCAGAAAGGCCGAGAAAGGTAT-3' 和 5'-ACGGATCCAGTGTAGT ATCT GTTC-3'). 分别以 U2u + U2d, U2u + U2u, U2d + U2d 为引物,用 PCR 法扩增基因间区,将扩增片段进行酶谱分析,并克隆于 M13mp18/19 载体中,用双脱氧法对扩增片段两端进行序列分析. 根据引物组合及扩增片段两端的 U2snRNA 特征序列,推测出基因的连锁方式.

1.4 水稻 U2snRNA 基因顺序的数据处理及同源性比较

应用计算机 PCGENE6.7FSINSCAN 软件系统,将各水稻 U2 基因编码区进行同源性比较,分析编码区上游调控元件的变化及 U2snRNA 二级结构的保守性.

2 结果与讨论

2.1 水稻 U2snRNA 基因的家数性

用 *Eco*R I, *Hind*III 酶切的水稻总 DNA 经 Southern 印迹分析,均有 10 余条杂交条带(图 1),证明水稻基因组中 U2snRNA 基因以多基因家族的形式存在. 结合水稻 U2snRNA 基因的酶切图谱和 U2snRNA 基因存在连锁且无内含子等特点(见下文),推测水稻 U2snRNA 基因的拷贝数至少为 15 个.

2.2 水稻中一些 U2snRNA 基因的连锁性分布

以拟南芥 U2.2 基因为探针,经噬菌体原位杂交,从水稻基因文库中筛选到 13 个阳性克隆,分别命名为 FDRGU2-1 ~ FDRGU2-13. 酶切图谱显示各克隆间均有一定的差异(结果未显示). 选取图谱差异最明显的 5 个克隆(U2-1, U2-3, U2-4, U2-5, U2-6)做进一步详细的分析和序列测定. 结果表明, U2-1, U2-4, U2-6 均含有 2 个 U2snRNA 基因,分别称为 A 和 B. U2-3 和 U2-5 为单基因,且 U2-3 与 U2-1A 的核苷酸序列相同. 进一步经 PCR 法扩增连锁基因的基因间区,并进行酶谱分析,结果显示 U2-4 和 U2-6 连锁间距均为 1.6 kb,间区 DNA 片段的酶切图谱亦相同,连锁基因均为顺式重复排列;由于基因间区的 DNA 序列具有较强的特异性,推测两者可能为同一组基因. 而 U2-1 的两个基因的连锁间距未能确定,推测可能是两个连锁基因的间距过长,难以用普通 PCR 方法扩增得到,酶谱分析和基因定

位中亦未发现短于 10 kb 的单一阳性片段, 从侧面证明了这一点. 此外, 以水稻总 DNA 为模板扩增连锁基因间区时发现除 1.6 kb 的片段外还有 3.2 kb 和 3.5 kb 的片段(数据未列出), 表明水稻基因组中 U2snRNA 基因存在不同的连锁方式(图 2).

2.3 水稻 U2snRNA 基因的结构特征

经序列分析, 水稻 U2-1A, U2-4A, U2-4B,

U2-5 基因编码区的同源性为 85.1%, 而相连锁的 U2-4A 和 U2-4B 的同源性为 95.5%, 表明连锁的基因具有更高的相似性. 4 个水稻 U2 基因皆具有典型的植物 U2snRNA 基因的特征^[7], 其编码区 5'

端 124 个碱基完全相同, 差异都发生在基因的 3' 端. 位于 U2snRNA 5' 端的 RNA-RNA 结合区域保守性极强, 而核糖核蛋白的结合区域在 3' 区的茎

环结构, 有较多变化, 这种差异意味着不同的 U2snRNA 与蛋白结合可形成特异的核糖核蛋白体^[8]. Hanley 等人^[9,10]对植物 UsnRNA 基因表达的研究表明, 在不同发育时期 UsnRNA 的表达模式不同, 不同的 UsnRNA 基因由于转录或转录后加工的不同而在发育的特殊时期表达. 由此推测, 不同的 U2snRNA 基因成员可能与生物体发育的各不同阶段特异表达的基因 mRNA 前体的剪接密切相关.

U2-1A, U2-4A, U2-4B, U2-5 基因的编码区上游 -70 区和 -30 区分别含有植物 UsnRNA 基因所特有的 USE (upstream sequence element) 元件和 TATA-like 元件(图 3). 除此之外, 在 USE 元件的上游, 从 -180 ~ -120 区分布着 2~3 个保守的 G/ACCCA/G 顺序(图 3), 此为单子叶植物 UsnRNA 基因所特有的调控顺序, 称为 MSP (monocot-specific promoter) 元件. 水稻 U2-1A, U2-4A 和 U2-5 基因中均含有两个 MSP, 顺序都是 GGCCCA, 而 U2-4B 中含有 3 个 MSP, 除两个是 GGCCCA 顺序外, 还有一个是 AGCCCG. 单子叶植物与双子叶植物 UsnRNA 基因上游调控区的不同, 反映出它们的转录系统存在差别. 据报道, 单子叶植物的 UsnRNA 基因(包括 MSP 缺失后的基因)在双子叶植物的原生质体中能够表达, 而双子叶植物的 UsnRNA 基因一般在单子叶植物的原生质体中不能表达^[10], 这证实了 MSP 元件是单子叶植物中 UsnRNA 基因表达所必需的, 也说明单子叶植物与双子叶植物的 UsnRNA 基因对表达所需的转录因子有着明显的不同.

2.4 水稻中 U2snRNA 假基因

5 个不同的 U2snRNA 基因中, U2-1B 无上游调控区, 与相应的 UsnRNA 相比, U2-1B 缺失 5' 端的 25 个碱基(图 3), 其余 5' 端的 99 个碱基顺序与其他的水稻 U2snRNA 相同, 3' 端不同于已测定的任何一个水稻 U2snRNA 基因, 是一个新的 U2snRNA 基因的部分片段. 在已报道的 UsnRNA 基因家族中, 普遍存在假基因^[5,11~13]. 据推测, 假基因可能是由于基因组的重组造成的, 表现为无上游调控元件或仅是基因的某个片段或基因发生部分缺失. U2-1B 是 U2snRNA 基因缺失了一部分, 故推测为假基因.

植物 U2snRNA 基因结构的多样性说明植物生长发育中 pre-mRNA 剪接加工的复杂性, 对植物 UsnRNA 基因家族的研究将为了了解植物生长发育的分子机制提供有益的信息.

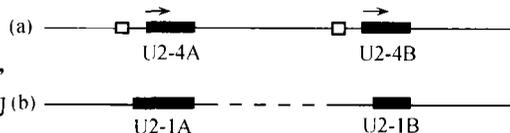


图 2 水稻 U2snRNA 基因连锁示意图

(a) 连锁基因 U2-4. 黑框为 U2-4A 和 U2-4B 基因编码区, 空框为基因上游调控区, 箭头表示基因的转录方向. 基因间区约为 1.6 kb. (b) 连锁基因 U2-1. 灰框表示 U2-1A 和 U2-1B 基因的调控区和编码区, 基因的排列方式尚未确定. 基因间区估计远大于 1.6 kb

	Monocot-specific Promoter	Monocot-specific Promoter
FDRGU2-1A	TAAATAC GGCCCA GATTTCCGGCCCCAAAAGGAAAAC TCCGGCCCA -----	----- GGCCCA -----
FDRGU2-4A	AAAGGCC GGCCCA TTTGAGAAGGCCTCGACGAGT-CAGC GGCCCA CACAG	----- GGCCCA -----
FDRGU2-4B	AAAGGCC GGCCCA TTTGAG---GGTTCCGCCGAGCCAGC GGCCCA CGCAA	----- GGCCCA -----
FDRGU2-5	GTTTCATG GGCCCA ACTACTCGGCCACTTGGGCTGNTGCGAGGTTTCAGGC	----- GGCCCA -----
FDRGU2-1B	-----	-----
FDRGU2-1A	----TGAATGCGTTTAAACGATAGCACCTGACGGGATCCGAGATGGATGG	
FDRGU2-4A	GCACAGCAAACG-----	-----CCA
FDRGU2-4B	CCCCG CATCGCATGATGTATTGTATCCCCCTCCCGTTG-----TTCGT	
FDRGU2-5	CCA ACCGCCAAGCAGTGAANGNCCGCGGCAGGTTGCTGCTTTGCTC	
FDRGU2-1B	-----	-----
	Upstream Sequence Element	
FDRGU2-1A	ATGGTTCTTGGGTTTAA ATACCACCTCG GAGACGGAGGACGTAGGAGGGCG	
FDRGU2-4A	ATGCTGTCTTCGTTTT GTACCACATCG CAGGATCAAGGAGAGTAGTACCA	
FDRGU2-4B	TCATTGTACAGTTT GTACCACATCG CCGATCAAGACGAGGAGATGTA	
FDRGU2-5	TTGTTGCGTAGTT GTACCAC -TCGCTTGTTCAGCTCGGTGAGGCGGT	
FDRGU2-1B	-----	-----
	TATA-like Box	+1
FDRGU2-1A	CTGCGCTCTGT TATATAG GCAAGCAGGGTTGCACGTTTCAACTC	ATACCTT
FDRGU2-4A	G-GAAGAGGC TAGAAA GCAGGGTGTAGCCAAGAATGCAACTC	ATACCTT
FDRGU2-4B	C-GGAGAGGC TACAAA GCAGGAGGTGAACGTGATAGCAACTC	ATACCTT
FDRGU2-5	C-GCGGAAGG TATAAA G-GAGSGCAGGGCTCCGTTCCGAAGC	ATACCTT
FDRGU2-1B	-----	G
	loop I	RNA-RNA Interaction Site
FDRGU2-1A	TCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTA	
FDRGU2-4A	TCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTA	
FDRGU2-4B	TCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTA	
FDRGU2-5	TCTCGGCCTTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTA	
FDRGU2-1B	TCGACCTGCAGGTCAACG GATCAAGT GTAGTATCTGTTCTTATCAGTTTA	
	Loop II	
FDRGU2-1A	ATATCTGATATGTGGGCCATGTGCTCACTACGATATTAATTTATTTTTT	
FDRGU2-4A	ATATCTGATATGTGGGCCATGTGCC -ACTTGATATTAATTTATTTTTT	
FDRGU2-4B	ATATCTGATATGTGGGCCATGTGCTCACTTTGATATTAATTTATTTTTT	
FDRGU2-5	ATATCTGATATGTGGGCC -TGTGTCCACTTCGATATTAATTTATTTTTT	
FDRGU2-1B	ATATCTGATATGTGGGCTATGTGCCATTTGTGATATTAATTTATTTTTT	
	Loop III	
FDRGU2-1A	GTGGGG-AGGGTCCACCACAGTGGCTTGCCACTGGGGCCCTCACGCGTTG	
FDRGU2-4A	GTGGGGGAGGGTCCACCATAGTGGCTTGCCACTAGGGCCCTTCGTGTGTCG	
FDRGU2-4B	GTGGGGGAGAGTCCACCATAGTGGCTTGCCACTAGGGCCCTCATGTGTGTCG	
FDRGU2-5	GTGGGGGAGGGTCCACCACAGTGGCTTGCCACTGGGGCCCTCGCGTGTCS	
FDRGU2-1B	GTGGGGGAGGGCTCACCACAGTGGCTTGCCACTGGG-CCCTCACGCGTTCG	
	Loop IV	
FDRGU2-1A	CCCAGGCGTTGCACTGCTGCCCGGGCCTGGCGCACCCCAAC CAAAACAA	
FDRGU2-4A	CCTAGGCGTTGCACTACAGCCTTGGC -TGGCGCACCCCAACCAAATCCA	
FDRGU2-4B	CCTAGGCGTTGCACTACAGCCTTGGC -TGGTGCACCCCAATCAAATCAA	
FDRGU2-5	CCAAGGCGTTGCACTACAGCCTGGGCCTGGCGCACCCCAAC CAAAATCAA	
FDRGU2-1B	CCCAGGC-TTGCACCGCTGCCCGGGCCTGGCGCACCCCAAC CAATACAA	

图 3 水稻 U2snRNA 基因的同源性比较

黑线框内为推测的基因的编码区,黑体为调控元件或基因结构域

参 考 文 献

- 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: 10 425
- 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: Birnstiel M, ed. *Small Nuclear Ribonucleo-protein Particles*. Berlin: Springer-Verlag, 1988. 38
- 6 Musci M A, Egeland D B, Schuler M A. Molecular comparison of monocot and dicot U1 and U2 snRNAs. *The Plant Journal*, 1992, 2: 589
- 7 王喜萍,冯征宇,孙崇荣,等. 水稻 U2snRNA 基因的分离及结构分析. *植物学报*, 1997, 39: 601
- 8 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: 5 910
- 9 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: 1 861
- 10 Hanley B A, Schuler M A. Developmental expression of plant snRNAs. *Nucl Acids Res*, 1991, 19: 6 319
- 11 Kiss T, Abel S, Solymosy F. A plant pseudogene for U1 RNA. *Plant Mol Biol*, 1989, 12: 709
- 12 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: 1 941
- 13 Waugh R, Clark G, Brown J W S. Sequence variation and lineage of potato U2snRNA-encoding genes established by PCR. *Gene*, 1991, 107: 197