

文章编号: 1674-7054(2013)04-0308-05

OsRRM1 基因标记开发及生物信息学分析

李德军^{1,2}, 夏志辉^{1,3}, 李晓兵¹, 朱立煌¹(1. 中国科学院遗传与发育生物学研究所, 北京 100101; 2. 中国热带农业科学院橡胶研究所, 海南 儋州 571737;
3. 海南大学农学院, 海南 海口 570228)

摘 要: 通过水稻杂种优势芯片中 OS009387_s_at Oligo 位点序列比对获得该位点对应的基因(*OsRRM1*), 基因结构分析表明, *OsRRM1* 具有 3 个内含子和 4 个外显子, 第 1 个内含子中序列插入/缺失可开发成分子标记。生物信息学分析结果显示, *OsRRM1* 蛋白由转运肽、酸性结构域和 RRM 结构域组成, *OsRRM1* 位点周围有影响水稻分蘖数、花粉育性和株高等多个性状的 QTLs, 以上研究为 *OsRRM1* 功能分析和分子标记在水稻中的应用奠定基础。

关键词: 水稻; *OsRRM1* 基因; 分子标记

中图分类号: Q 812

文献标志码: A

在生物个体的生长发育过程中, 基因的表达在转录、转录后和翻译 3 个水平上都受到严格的调控。转录调控是基因表达调控的第一步, 而转录后调控则决定着基因最终表达产物的类型和丰度。蛋白伴随着 RNA 从产生到降解的全部过程, RNA 结合蛋白是其中的一类重要蛋白。RNA 结合蛋白是一类包含一个或多个 RNA 结合结构域(RNA recognition motif, RRM)的蛋白。RRM 结构域一般包含约 80 个氨基酸, 形成 $\beta_1 - \alpha_1 - \beta_2 - \beta_3 - \alpha_2 - \beta_4$ 的二级结构。其中有 8 个和 6 个氨基酸组成的 RNP1 和 RNP2 两个结合 RNA 所必需且高度保守的模体, RNP1 和 RNP2 分别位于 RRM 二级结构 β_1 和 β_3 反向平行结构片层结构的中央位置(图 1)^[1]。RNA 结合蛋白是一个较古老且保守的蛋白, 在人、动物、植物、真菌和细菌等不同物种中已发现约 300 个含有 RRM 的 RNA 结合蛋白基因, 对 RNA 结合蛋白的研究主要集中在人、动物和微生物上, 在植物中研究相对较少。

在真核生物中, RNA 结合蛋白参与 RNA 分子的合成、加工、编辑、修饰和运输过程^[2-4]。此外, RNA 结合蛋白在遗传信息的解读、端粒保持和 X 染色体失活中发挥作用^[5-6]。与非植物 RNA 结合蛋白相比, 对植物 RNA 结合蛋白的研究相对较少。植物 RNA 结合蛋白参与多种逆境条件的反应^[7]、RNA 编辑^[8-11]、减数分裂^[12]、形态建成^[13]、调节小 RNA 丰度和加工^[14]、调节植物激素信号和调控植物开花^[15]等过程。在分析水稻杂种优势芯片结果时, 发现编号为 OS009387_s_at Oligo 的基因表达模式与 RT-PCR 的结果相同。通过 Oligo 序列回帖得到该基因(*OsRRM1*)对应的序列, 通过比对 *OsRRM1* 基因 cDNA 和基因组序列解析基因结构并开发分子标记。利用生物信息学方法分析 *OsRRM1* 蛋白结构域和 *OsRRM1* 位点已定位影响水稻性状的 QTLs 分析, 以上研究为 *OsRRM1* 功能分析和分子标记在鉴定水稻籼、粳稻品种时奠定基础。

收稿日期: 2013-10-22

基金项目: 国家转基因重大专项(2013ZX08009-001-005); 国家自然科学基金(31071173)

作者简介: 李德军(1975-), 男, 辽宁朝阳人, 中国热带农业科学院橡胶研究所副研究员, 硕士生导师。E-mail: djli.rricatas@gmail.com

通信作者: 朱立煌(1943-), 男, 上海人, 中国科学院遗传与发育生物学研究所研究员, 博士生导师。E-mail: lhzhu@genetics.ac.cn

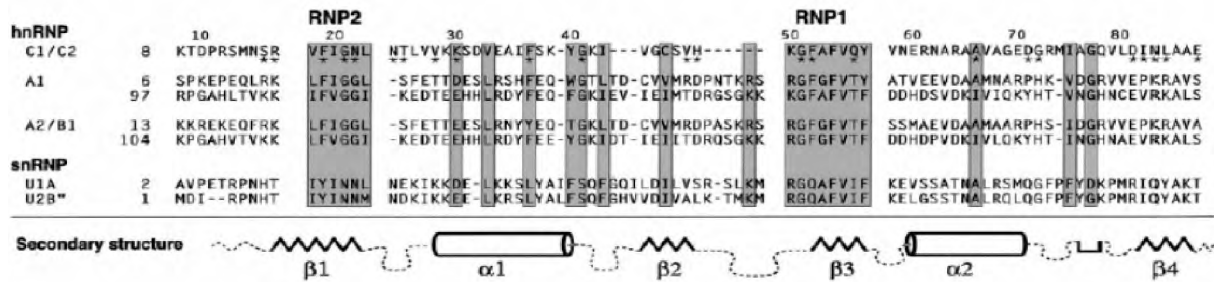


图1 RRM 结构域组成及二级结构示意图^[1]

RNP1 和 RNP2 为 RRM 结构域中 2 个保守的模体, α 和 β 分别代表二级结构中 α 螺旋和 β 折叠

Fig. 1 The diagram of domain and secondary structures in RNA recognition motif^[1]

RNP1 and RNP2 represent two conserved motifs in RRM, and α and β represent separately α -helix and β sheets of secondary structure

1 材料与方法

1.1 植物材料 本实验选用籼稻品种为 9311、南京 6、南京 11、明恢 63、窄叶青、特青、IR24 和桂朝 2 号; 粳稻品种为日本晴、Pai64s、京系 17、中花 11、台北 309、丽江、KL908、江南香糯和 SH55; 粳型野生稻为东乡普通野生稻和元江普通野生稻。实验材料均来自国家种质资源库。

1.2 水稻材料中 DNA 和 RNA 的提取 采用改进的 SDS 方法从水稻叶片中提取基因组 DNA^[14]。取大约 100 mg 水稻叶片, 经液氮冷冻, 在直径 5 cm 的小研钵中磨成粉状, 转移到 1.5 mL 离心管里提取 DNA, 获得的 DNA 沉淀溶解于 100 μ L ddH₂O 中。用 Qiagen plant RNA easy 试剂盒提取水稻 RNA, 具体步骤按说明书进行。

1.3 RT-PCR 及 PCR 反应 从提取的 RNA 中取 5 μ L 总 RNA 样品用于逆转录反应, 以逆转录后产物为模板, 高保真聚合酶扩增分离 *OsRRM1* cDNA。利用 KOME 提供的 *OsRRM1* 基因全长 cDNA 序列设计引物, F: 5'-GAGCCCCAAAACCTTATCTT-3' 和 R: 5'-CTAACAAGAGGATCTGTGCA-3'。PCR 反应条件为: 94 $^{\circ}$ C 3 min; 30 个循环的 94 $^{\circ}$ C 1 min, 58 $^{\circ}$ C 1 min, 72 $^{\circ}$ C 2 min; 72 $^{\circ}$ C 10 min, 4 $^{\circ}$ C 保温。PCR 产物加 A 纯化后连接 T-easy 载体并测序。

根据 *OsRRM1* 序列在水稻品种 93-11 和日本晴比对结果设计 InDel 引物, InDel 引物序列如下: F: 5'-GAGGAATCCACTAGATGAGG-3' 和 R: 5'-GGGAGTATACTCACGTATCC-3'。PCR 反应程序如下: 预变性 94 $^{\circ}$ C 5 min; 然后 94 $^{\circ}$ C 45 S, 56 $^{\circ}$ C 45 S, 72 $^{\circ}$ C 1.45 min, 共 35 个循环; 72 $^{\circ}$ C 10 min, 最后 4 $^{\circ}$ C 保温。PCR 产物在 $\omega = 1.5\%$ 琼脂糖凝胶上于 1 $\times$ TAE 缓冲液电泳检测。

1.4 生物信息分析 利用 NCBI 网站 BLAST(<http://www.ncbi.nlm.nih.gov/BLAST/>) 进行全基因组比对; 基因编码蛋白的保守结构域分析利用 NCBI 网站提供的保守结构域分析(<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>); 编码序列与水稻全长 cDNA 比对分析在 KOME 网站(<http://cdna01.dna.affrc.go.jp/cDNA/Wblast.html>) 进行; 基因所在染色体位置 QTL 定位分析在 GRAMME 网站中进行, 具体网址见 http://www.gramene.org/Oryza_sativa/contigview?chr=3&vc_start=3688516&vc_end=3691236&x=10&y=6。

2 结果与分析

2.1 RNA 结合蛋白基因的克隆及基因结构分析 笔者在验证水稻杂种优势芯片结果时发现编号为 OS009387_s_at Oligo 的基因表达模式与 RT-PCR 的结果一致, 通过 Oligo 序列回帖得到基因序列, 分析该基因编码蛋白序列发现蛋白包含 RRM 结构域, 在本研究将该基因命名为 *OsRRM1*。序列比对发现该基因在 KOME(<http://cdna01.dna.affrc.go.jp/cDNA/Wblast.html>) 中接受号为 AK061072。依 KOME 提供的全长 cDNA 序列设计引物, 通过 RT-PCR 方法从日本晴叶片总 RNA 中进行扩增并测序。测序及比对

结果表明,*OsRRM1* 序列与 KOME 中提供的全长 cDNA 完全相同。为进一步分析基因结构,用日本晴全长 cDNA 与日本晴基因组序列做比对(<http://biom3.univlyon1.fr/sim4.html/>),结果发现该基因含有 3 个内含子和 4 个外显子,3 个内含子的长度分别为 512,374 和 155 bp(图 2)。



图 2 *OsRRM1* 基因结构示意图

■ 为 5' 非翻译区, ■ 为 3' 非翻译区, ■ 为开放阅读框, □ 为内含子区, 箭头标识为日本晴和 93-11 插入/缺失位置

Fig. 2 The diagram of gene structures of *OsRRM1*

■ represent 5' untranslated region, ■ 3' untranslated region, ■ open reading frame and intron region of *OsRRM1*, respectively. □ The arrows show the insertion/deletion positions of *OsRRM1* between 93-11 and Nipponbare

2.2 *OsRRM1* 基因 InDel (Insertion and Deletion) 标记开发 为分析 *OsRRM1* 基因是否在籼、粳稻中存在差异,笔者用日本晴 *OsRRM1* 基因比对籼稻品种 93-11 获得该基因在 93-11 中 cDNA 和基因组序列, 比对 93-11 cDNA 和基因组序列结果表明 *OsRRM1* 基因在 93-11 中同样有 3 个内含子和 4 个外显子。此外,比对粳稻品种日本晴和籼稻品种 93-11 序列发现在第 2 个内含子处较日本晴缺失 139 bp, 93-11 *OsRRM1* 基因在第 1 个外显子中有 1 个 15 bp 的缺失。为方便标记应用,笔者验证了内含子中的缺失是否具有籼稻和粳稻品种特异性。在缺失区段的两端设计引物,对 11 个粳稻(其中包括东乡和元江普通野生稻) 和 8 个籼稻进行 PCR 扩增^[17-18]。PCR 结果表明,粳稻扩增带型同日本晴,而籼稻带型同 93-11(图 3),以上结果初步说明 *OsRRM1* 基因为籼、粳稻特异的 InDel 标记。

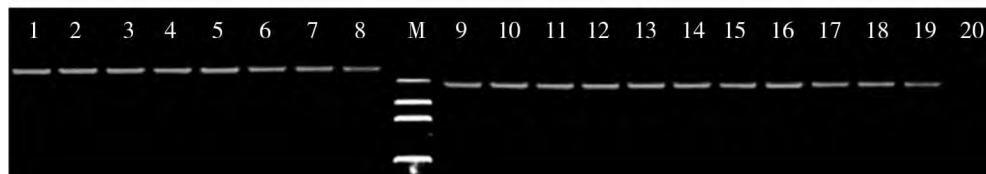


图 3 电泳检测 *OsRRM1* 基因在籼、粳稻品种中缺失和插入

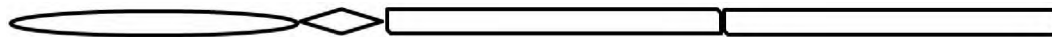
1~19 编号分别为 93-11、南京 6、南京 11、明恢 63、窄叶青、特青、IR24、桂朝 2 号、日本晴、Pai64s、京系 17、中花 11、台北 309、丽江、KL908、江南香糯、SH55、东乡普通野生稻和元江普通野生稻,20 编号为水对照,M 为 DNA 分子质量标记 DL2000

Fig. 3 The electrophoretogram of *OsRRM1* between indica and japonica rice varieties

1-19 represent separately 93-11, Nanjing6, Nanjing11, Minghui63, Zaiyeqing, Teqing, IR24, Guichao2, Nipponbare, Pai64s, Jingxi17, Zhonghua11, Taipei309, Lijing, KL908, Jiangnan Xiangnuo, SH55, Dongxiang Puye and Yuanjiang Puye; 20 and M indicate CK and molecular marker DL2000, respectively

2.3 *OsRRM1* 蛋白结构域分析 基因的功能最终是通过蛋白质实现的,笔者对 *OsRRM1* 编码的蛋白作了初步分析。*OsRRM1* 基因编码 265 个氨基酸,相对分子质量为 28.11 kD,等电点为 4.655。在 NCBI 保守结构域数据库中进行比对,发现 *OsRRM1* 蛋白含有保守的 RRM 结构域(<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>)。推测 *OsRRM1* 蛋白由 4 部分组成: N 端 61 个氨基酸富含羟基化氨基酸残基(约为 46%)且带有 5 个正电荷,因此,可能为转运肽;接下来 18(62-79)个氨基酸带有 10 个负电荷,推测为酸性结构域;C 端为两个保守的 RRM 结构域: 即 80-154 和 181-254 区域。根据 *OsRRM1* 蛋白结构分析,此蛋白可能为叶绿体 RRM 蛋白。

2.4 *OsRRM1* 位点定位影响水稻性状的 QTLs 分析 利用 GRAMME 网站(<http://www.gramene.org/>)提供的资源,分析了前人在 *OsRRM1* 基因位点所在染色体区段定位的 QTLs,此分析为确定 *OsRRM1* 基因影响的表型提供线索。通过分析得到此区段为 QTL 密集区,前人在 *OsRRM1* 区段共定位了分蘖数(CQAG17)、花粉育性(AQBI005)、雄性育性恢复(AQCG002)和水稻株高(CQE60-PTHT)等 17 个 QTLs,

图4 推测 *OsRRM1* 蛋白结构示意图

○ 为转运肽, ◇ 为酸性结构域, □ 为 RRM 结构域

Fig. 4 The diagram of *OsRRM1* protein structure

○ represent transit peptide, ◇ acidic domain, □ RRM respectively

该 QTL 定位结果表明 *OsRRM1* 可能为多功能基因,与水稻株高、分蘖数、花粉育性和雄性育性恢复等性状相关 (http://www.gramene.org/Oryza_sativa/contigview?chr=7&vc_start=26160028&vc_end=26312475&x=4&y=13)。

3 讨论

RRM 结合蛋白是生物中古老且保守的蛋白,它伴随着 RNA 从产生到降解的全部过程,在生物生长发育中发挥着重要作用^[1]。用水稻杂种优势芯片中编号为 OS009387_s_at Oligo 基因的 Oligo 序列回帖得到芯片相应点对应的 *OsRRM1* 基因序列。比对 *OsRRM1* cDNA 与基因组序列结果发现该基因含有 3 个内含子和 4 个外显子,通过比较籼稻品种 93-11 和粳稻品种日本晴 *OsRRM1* 基因序列发现 2 个水稻品种存在缺失/插入遗传变异,尽管在第 1 个外显子处出现遗传变异,但该插入不造成移码。该插入位于推测转运肽区,是否对蛋白定位造成影响有待于进一步研究。实验结果表明,*OsRRM1* 内含子处遗传变异可开发成鉴定水稻籼粳亚种的分子标记,因此,可在鉴定水稻籼粳亚种中应用。

序列分析结果表明 *OsRRM1* 基因编码 265 个氨基酸,相对分子质量为 28.11 kD,等电点为 4.655。蛋白结构分析显示 *OsRRM1* N 端 61 个氨基酸可能为转运肽,对蛋白的定位具有重要作用,当完成转运后被信号肽酶切除,变成成熟蛋白;接下来 18 个氨基酸推测为酸性结构域,该结构域在蛋白与蛋白互作中发挥作用;C 端为 2 个保守的 RRM 结构域:即 80-154 和 181-254 区域,它们在与 RNA 结合中发挥作用。*OsRRM1* 蛋白结构与 RNA 结合蛋白分类中的叶绿体 RRM 蛋白高度相似,已有研究显示叶绿体 RRM 蛋白多与叶绿体 mRNA 3' 末端产生、RNA 稳定和 RNA 编辑等过程相关^[8-11,19],推测 *OsRRM1* 在水稻中也可能与上述过程有关。此外,利用 GRAMME 网站推测 *OsRRM1* 区段为 QTL 密集区,前人在此区域共定位了包括分蘖数、花粉育性、雄性育性恢复和水稻株高等 17 个 QTLs。因此,*OsRRM1* 可能为 1 个多功能基因,其在水稻中的功能有待进一步研究。

参考文献:

- [1] EWAN B, SANJAY K, ADRIAN R K. Analysis of RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors [J]. Nucleic Acids Res., 1993, 21: 5803-5816.
- [2] COLE C N. Choreographing mRNA biogenesis [J]. Nat. Genet., 2001, 29: 6-7.
- [3] PROUDFOOT N J, FURGER A, DYE M J. Integrating mRNA processing with transcription [J]. Cell, 2002, 108: 501-512.
- [4] REED R, HURT E. A conserved mRNA export machinery coupled to pre-mRNA splicing [J]. Cell, 2002, 108: 523-531.
- [5] BACHAND F, TRIKI I, AUTEXIER C. Human telomerase RNA-protein interactions [J]. Nucleic Acids Res., 2001, 29: 3385-3393.
- [6] BOUMIL R M, LEE J T. Forty years of decoding the silence in X-chromosome inactivation [J]. Hum. Mol. Genet., 2001, 10: 2225-2232.
- [7] FREY N F D, MULLER P, JAMMES F, et al. The RNA binding protein tudor-SN is essential for stress tolerance and stabilizes levels of stress-responsive mRNAs encoding secreted proteins in *Arabidopsis* [J]. Plant Cell, 2010, 22: 1575-1591.
- [8] CHUSTER G, GRUISSEM W. Chloroplast mRNA 3' end processing requires a nuclear-encoded RNA-binding protein [J]. EMBO J., 1991, 10: 1493-1502.
- [9] HAYES R, KUDLA J, SCHUSTER G, et al. Chloroplast mRNA 3'-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins [J]. EMBO J., 1996, 15: 1132-1141.
- [10] HIROSE T, SUGIURA M. Involvement of a site-specific trans-acting factor and a common RNA-binding protein in the editing of chloroplast mRNAs: development of a chloroplast *in vitro* RNA editing system [J]. EMBO J., 2001, 20: 1144-1152.

- [1] SUN T, GERMAIN A, GILOTEAUX L, et al. An RNA recognition motif-containing protein is required for plastid RNA editing in *Arabidopsis* and maize [J]. PNAS, 2013, 4: E1169 – E1178.
- [2] NONOMURA K I, EIGUCHI M, NAKANO M, et al. A novel RNA-recognition – motif protein is required for premeiotic G1/S-phase transition in rice (*Oryza sativa* L.) [J]. PLoS Genet., 2011, 7(1): e1001265.
- [3] HONG F, ATTIA K, WEI C, et al. Overexpression of the rFCA RNA recognition motif affects morphologies modifications in rice (*Oryza sativa* L.) [J]. Biosci. Rep., 2007, 27: 225 – 234.
- [4] REN G D, XIE M, DOU Y C, et al. Regulation of miRNA abundance by RNA binding protein TOUGH in *Arabidopsis* [J]. PNAS, 2012, 109 (31): 12817 – 12821.
- [5] JUNG J H, SEO P J, A J H, et al. *Arabidopsis* RNA-binding protein FCA regulates microRNA172 processing in thermosensory flowering [J]. J. Biol. Chem., 2012, 287: 16007 – 16016.
- [6] 张凤娟, 张满良, 朱水芳. 一种改进的水稻总 DNA 的快速提取方法 [J]. 植物检疫, 2004, 18(6): 330 – 332.
- [7] SUN C Q, WANG X K, YOSHIMURA A, et al. Genetic differentiation from nuclear, mitochondrial and chloroplast genomes in common wild rice (*O. rufipogon* Griff.) and cultivated rice (*O. sativa* L.) [J]. Theor. Appl. Genet., 2002, 104: 1335 – 1345.
- [8] 朱作峰, 孙传清, 李自超, 等. 用 SSR 标记对水稻品种的分类研究 [J]. 农业生物技术学报, 2009, 9(1): 58 – 61.
- [9] LORKOVIC Z J, BARTA A. Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana* [J]. Nucleic Acids Res., 2002, 30: 623 – 635.

Developing of Molecular Marker and Bioinformatic Analysis of *OsRRM1*

LI Dejun^{1,2}, XIA Zhihui^{1,3}, LI Xiaobing¹, ZHU Lihuang¹

(1. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; 2. Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou 571737, China; 3. College of Agronomy, Hainan University, Haikou 570228, China)

Abstract: The *OsRRM1* was identified by aligning the sequences of OS009387_s_at Oligo with the genes developed from heterosis microarray in rice. The results from gene structure showed that *OsRRM1* contained three introns and four exons, and developed insertion/deletion (InDel) marker within the first intron. *OsRRM1* was composed of transit peptide, acidic domain and RNA recognition motif. The QTLs related to traits such as tiller number, pollen fertility, plant height, etc, were mapped around *OsRRM1* by some researchers. These studies laid foundation for studying the function of *OsRRM1* and utilizing the InDel marker of *OsRRM1* in rice.

Key words: rice; *OsRRM1*; molecular marker