遗传学报 2007 34 (3): 247-257 ISSN: 0379-4172 CN: 11-5450/R

研究论文

水稻NBS-LRR基因选择性剪接的全基因组检测及分析

顾连峰，郭荣发

广东海洋大学农学院，湛江524088

收稿日期 2006-5-8 修回日期 2006-7-29 网络版发布日期 2007-3-13 接受日期

摘要

选择性剪接是促进基因组复杂性和蛋白质组多样性的一种主要机制，但是对水稻NBS-LRR序列选择性剪接的全基因组分析却未见报道。通过隐马尔柯夫模型搜索，从TIGR数据库里得到了855条编码NBS-LRR基序的序列。利用这些序列在KOME、TIGR基因索引及UniProt三个数据库中进行同源搜索，获得同源的完整cDNA序列、假设一致性序列和蛋白质序列。再利用Spidey 和SIM4程序把完整cDNA序列和假设一致性序列联配到相应的BAC序列上来预测选择性剪接。蛋白质序列和基因组序列之间的联配使用tBLASTn。在这875个NBS-LRR基因中，119个基因具有选择性剪接现象，其中包括71 内含子保留，20个外显子跳跃，25个选择性起始，16个选择性终止，12个5′端的选择性剪接和16个3′端选择性剪接。大多数选择性剪接都为两个和多个转录本所支持。可以通过访问http://www.bioinfor.org查询这些数据。进而通过生物信息学分析剪接边界发现外显子跳跃和内含子保留的‘GT…AG’的规则不如组成型的保守。这暗示了它们是通过不同的调控机制来指导剪接变构体的形成。通过分析内含子保留对蛋白质的影响，发现选择性剪接的蛋白更倾向于改变其C端氨基酸序列。最后对选择性剪接的组织分布和蛋白质定位进行分析，结果表明选择性剪接的最大类的组织分布是根和愈伤组织。超过1/3剪接变构体的蛋白质定位是质膜和细胞质。这些选择性剪接蛋白可能在抗病信号转导中起到重要作用。

关键词 选择性剪接；生物信息学；NBS-LRR同源序列；水稻

分类号

Genome-wide Detection and Analysis of Alternative Splicing for Nucleotide Binding Site-Leucine-Rich Repeats Sequences in Rice

Lianfeng Gu, Rongfa Guo

College of Agriculture, Guangdong Ocean University, Zhanjiang 524088, China

Abstract

<P>Alternative splicing is a major contributor to genomic complexity and proteome diversity, yet the analysis of alternative splicing for the sequence containing nucleotide binding site and leucine-rich repeats (NBS-LRR) domain has not been explored in rice (Oryza sativa L.). Hidden Markov model (HMM) searches were performed for NBS-LRR domain. 875 NBS-LRR-encoding sequences were obtained from the Institute for Genomic Research (TIGR). All of them were used to blast Knowledge-based Oryza Molecular Biological Encyclopaedia (KOME), TIGR rice gene index (TGI), and Universal Protein Resource (UniProt) to obtain homologous full-length cDNAs (FL-cDNAs), tentative consensus sequences, and protein sequences. Alternative splicing events were detected from genomic alignment of FL-cDNAs, tentative consensus sequences, and protein sequences, which provide valuable information on splice variants of genes. These sequences were aligned to the corresponding BAC sequences using the Spidey and Sim4 programs and each of the proteins was aligned by tBLASTn. Of the 875 NBS-LRR sequences, 119 (13.6%) sequences had alternative splicing where multiple FL-cDNAs, TGI sequences and proteins corresponded to the same gene. 71 intron retention events, 20 exon skipping events, 16 alternative termination events, 25 alternative initiation events, 12 alternative 5′ splicing events, and 16 alternative 3′ splicing events were identified. Most of these alternative splices were supported by two or more transcripts. The data sets are available at http://www.bioinfor.org. Furthermore, the bioinformatics analysis of splice boundaries showed that exon skipping and intron retention did not exhibit strong consensus. This implies a different regulation mechanism that guides the expression of splice isoforms. This article also presents the analysis of the effects of intron retention on proteins. The C-terminal regions of

alternative proteins turned out to be more variable than the N-terminal regions. Finally, tissue distribution and protein localization of alternative splicing were explored. The largest categories of tissue distributions for alternative splicing were shoot and callus. More than one-thirds of protein localization for splice forms was plasma membrane and cytoplasm. All the NBS-LRR proteins for splice forms may have important function in disease resistance and activate downstream signaling pathways.</P>

Key words alternative splicing；bioinformatics；NBS-LRR homologous sequence； rice

DOI:

通讯作者 郭荣发 guorf@gdou.edu.cn