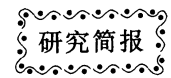


DOI: 10.3724/SP.J.1006.2009.00566



玉米自交系 B73 全基因组 NBS 类型抗病基因分析

汪结明 江海洋 赵阳 项艳 朱苏文 程备久*

安徽农业大学生命科学学院, 安徽合肥 230036

摘要: 核苷酸结合位点(NBS)类型抗病基因是植物抗病基因中最大的一个类别, 也是近年来植物抗病分子育种研究的一大热点。本研究对玉米自交系 B73 全基因组中含有 NBS 结构的候选抗病基因进行了基因总数、类型、系统进化关系等分析。在 B73 全基因组中含有 165 个 NBS 结构的基因, 远远少于水稻中的同类基因, 按照 N-端结构和亮氨酸富集区(LRR)结构, 将其分为 153 个标准结构和 12 个非标准结构 NBS 基因。其中, 标准结构基因又分为 CC-NBS-LRR、CC-NBS、NBS、NBS-NBS、NBS-LRR、NBS-NBS-LRR、NBS-X、X-NBS 等 8 个类型。系统进化树分析表明, NBS 类型抗病基因存在明显的两大分支, 与水稻的星状发散型分布有很大差异。通过基因家族分析, 还发现了玉米 NBS 类型基因的复制现象, 但发生复制的基因比例低于水稻, 可能是造成玉米 NBS 抗病基因数目较少的原因之一。

关键词: 玉米; 生物信息学; 抗病基因; 核苷酸结合位点(NBS); 系统进化树

Genome-Wide Analysis of NBS-Encoding Disease Resistance Genes in Maize Inbred Line B73

WANG Jie-Ming, JIANG Hai-Yang, ZHAO Yang, XIANG Yan, ZHU Su-Wen, and CHENG Bei-Jiu*

School of Life Science, Anhui Agricultural University, Hefei 230036, China

Abstract: Nucleotide-binding site (NBS) disease resistance gene is a largest category in plant disease resistance genes, which is a focus in recent studies on molecular breeding of plant disease resistance. Using maize (*Zea mays* L.) inbred line B73, the complete set of disease resistance candidate genes that encode NBS was identified in the genome. The putative NBS genes were characterized with respect to structural diversity, phylogenetic relationships and so on. One hundred and sixty-five NBS-coding sequences were identified into two types: nonregular (12) and regular NBS genes (153). The amount of NBS genesis much smaller in maize than in *japonica* rice (*Oryza sativa* L.). The 153 regular NBS genes were categorized into eight classes, including CC-NBS-LRR, CC-NBS, NBS, NBS-NBS, NBS-LRR, NBS-NBS-LRR, NBS-X, and X-NBS, according to N-terminal motif and leucine-rich repeat (LRR) domains motif. The 165 NBS genes showed two remarkable branches in the phylogenetic tree, differing from the radiation structure in *japonica* rice. Gene duplication event was observed based on gene family analysis of the NBS disease-resistance genes in maize; however, the ratio of gene duplication was smaller than that in rice. This might be one of the reasons for less NBS disease-resistance genes in maize than in rice.

Keywords: *Zea mays* L.; Bioinformatics; Disease resistance gene; Nucleotide binding site; Phylogenetic tree

玉米(*Zea mays* L.)作为重要的粮食、饲料和能源作物在世界范围内广泛种植^[1], 但由于病害较为普遍, 往往造成大量减产、品质下降。阐明病原菌与寄主植物之间的相互作用关系, 揭示植物抗病机制一直是植物病理学及植物抗病育种研究中的热点问题^[2]。植物对病原菌的抗病性表现多种途径, 首先是植物本身的某些结构障碍如细胞壁的角质、蜡质、栓质、木质素及特殊的气孔结构等; 其次是植物本身的某些化学成分具有抗病功能, 如小分子

抗病物质、影响病原菌细胞透性的蛋白质和核糖体失活蛋白等^[3]; 此外还有在病原菌侵染时植物表现的受抗病基因影响的抗病反应。根据 Flor 的“基因对基因”假说, 植物与病原菌之间存在着共进化的关系, 寄主植物的每一个抗性基因 (resistance gene, *R* 基因), 在病原菌中就存在相应的无毒基因(avatar gene, *avr* 基因), 与 *avr* 基因相对应的是病原菌的毒性基因(virulence gene, *vir* 基因), 含有 *avr* 和 *vir* 基因的病原菌对含 *R* 基因的寄主分

本研究由国家高技术研究发展计划(863 计划)项目(2008AA10Z408 和 2006 AA10Z1B4)和教育部科学技术研究重点项目(206065)资助。

* 通讯作者(Corresponding author): 程备久, E-mail: cbj@ahau.edu.cn

第一作者联系方式: E-mail: wjmmjw2000@163.com

Received(收稿日期): 2008-08-20; Accepted(接受日期): 2008-10-06.

别表现出非亲和性与亲和性^[4]。

作为最大的一个抗病基因类别, NBS 类型抗病基因在保护植物免受病原菌和有害物侵害中起十分重要的作用, 其编码蛋白包含一个核苷酸结合位点(NBS)和亮氨酸的富集区(LRR)^[5], 这个区域包含一些非常保守的基序, 因而被广泛地用于植物抗病基因的识别和分类。根据系统发生学对 NBS 抗病基因的 8 个主要基序进行的研究显示^[6-7], 以 NBS-LRR 结构编码的这一类基因在植物中主要被分成两大类。一类是在编码蛋白 N 端包含一个 TIR (toll and interleukin-1 receptor, 是与果蝇 Toll 蛋白及哺乳动物白细胞介素-1 受体的胞外域相似的区域) 结构, 称作 TIR-NBS-LRR 类型基因或简称 TNL 类型基因^[8]。另一类并不编码 TIR 结构区域, 而往往在 N-端带有一个卷曲螺旋结构(coiled-coil, CC), 所以也称作 CC-NBS-LRR 类型基因^[9]。

抗病基因作为玉米基因组中的一个大家族, 对其进行分子生物学研究具有十分重要的意义。目前, 在玉米全基因组 NBS 类型抗病基因的分类和系统进化关系等方面的研究明显落后于模式植物拟南芥(*Arabidopsis thaliana*) 和水稻(*Oryza sativa* L.)。美国华盛顿大学科学家今年 3 月份宣布玉米基因组草图绘制完成, 并且已经通过网站免费向世界提供草图数据^[10], 其功能基因组的研究已经起步。因此利用现代生物信息学的手段, 在玉米全基因组水平上研究含有核苷酸结合区(NBS)结构的候选抗病基因已成为可能。本文在玉米基因组草图公布后, 首次通过生物信息学方法研究玉米 B73 全基因组中 NBS 类型的抗病基因总量, 并对其进行分类和系统发生学分析, 旨在为玉米抗病基因的克隆、抗病机制的分析, 加快玉米抗病育种进程奠定理论基础。

1 材料与方法

1.1 植物材料及其基因组 DNA 信息获得

以玉米自交系 B73 全基因组数据为研究对象, 资料来自玉米基因组测序网站 http://www.maizegenome.org/data_portal.html。

1.2 NBS 类型基因的确定

第一步, 在 NCBI 网站的 GenBank 中以“resistance”和“disease”为关键词搜索, 获得以单子叶植物为主的 20 种植物 180 个含 NBS 区域的基因, 将这些基因提交蛋白家族数据库(Pfam, <http://pfam.janelia.org>)以获取抗病基因 NBS 区域。然后, 使用 DNAtools 软件建立玉米 B73 基因组 cDNA 本地数据库, 再以抗病基因的 NBS 区域序列为递交序列, P-value 设为 10^{-6} 进行 BLAST 搜索, 寻找同源序列^[11]。

第二步, 对所有符合要求的序列再通过 Pfam 数据库 (<http://pfam.wustl.edu/hmmsearch.shtml>)、SMART 蛋白基序分析 (Simple Modular Architecture Research Tool, <http://smart.embl-heidelberg.de>)、COILS(<http://www.ch.embnet.org/software>, threshold = 0.9)来确定它们的蛋白基序、区域, 并进行家族分类信息^[12]。

第三步, 对所有已被选取基因的核苷酸序列通过 ClustalW 的方法进行多序列排列, 根据排列的结果, 去除候选基因中的重复序列。

1.3 NBS 类型基因的分类

参照水稻、拟南芥中 NBS 类型基因的分类方法, NBS 区域由 pre-P-Loop 结构开始至 MHDV 结构结束, 一般包含 260~300 个氨基酸。在 pre-P-Loop 结构之前以及 MHDV 结构之后分别是 N-端和 LRR 结构^[13]。根据 NBS 类型蛋白的 N 端结构和 LRR 结构的不同, 对其进行分类。

1.4 NBS 类型基因系统进化树的构建

由于抗病基因 NBS 结构的氨基酸序列具有高度保守性, 因此可以进行多序列的比对和构建进化树, 以用于研究它们的差异以及系统进化关系^[14]。本研究将玉米 B73 基因组 NBS 基因 NBS 结构区域氨基酸序列进行 ClustalW 排序, 利用 DNAMAN 软件对所有序列使用 Kimrua 距离法构建系统进化树^[15], 以分析其抗病基因的进化关系。

1.5 NBS 类型基因的复制

基因复制(gene duplication)是多基因家族中普遍存在的方式, 通过不等交换(unequal crossing over)、逆转录转座(retroposition) 或全基因组重复(whole genome duplication) 等途径产生一个与原基因相似的基因或碱基序列, Bai 等^[16]将基因复制概念归纳为两个特征, 即发生基因重复的两个基因序列之间有大于 70% 的核苷酸序列可以排列; 可排列的区域氨基酸序列的序列相似度大于 70%。通过对基因复制的研究, 能更好地理解基因的起源、基因家族的形成、基因表达水平的多样性^[17]。利用 MEGA3.1 中的 ClustalW 模块将所有的玉米 NBS 类型抗病基因进行排序, 再将排序结果进行基因间序列相似度计算, 根据计算结果统计基因复制事件。

2 结果与分析

2.1 NBS 类型基因的鉴定

在获得的 180 个含 NBS 区域的核苷酸序列中, 经 BLAST 将序列比对后的重叠序列去除, 最终得到 172 个候选基因。再经 Pfam 分析和多序列排序去除无 NBS 结构区域的候选基因, 最终得到 165 个含 NBS 结构的基因。

2.2 NBS 类型基因的分类

在 165 个 NBS 类型抗病基因中, 153 个基因的 NBS 区域的基序高度保守, 具有完整的开放阅读框(ORF), 这类基因被称为标准 NBS 类型基因, 但其他 12 个基因的 NBS 区域呈高度分化状态, 它们的 NBS 区域的结构与上述 153 个基因有很大差别, 往往 N-端序列过短(小于正常的 2/3 长), 被称为非标准 NBS 类型基因(nonregular NBS gene)。这些非标准 NBS 类型基因在 NBS 区域的 TBLASTN 工作中, 显示出非常高的 P 值匹配, 这也说明它们在 NBS 区域与标准的 NBS 类型基因结构上的差距, 但是它们也带有 LRR

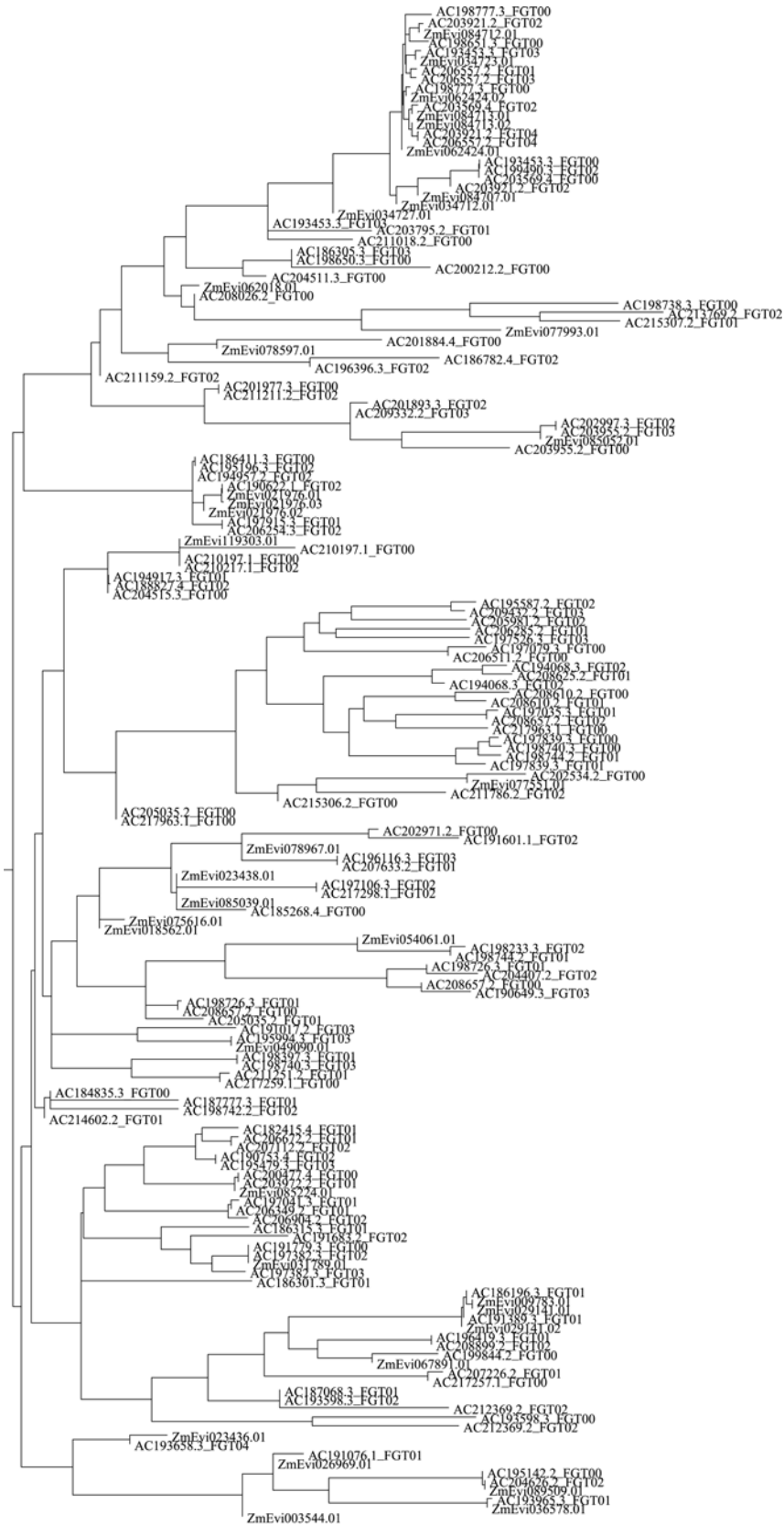


图 1 玉米自交系 B73 中 NBS 基因的系统发生树
Fig. 1 Phylogenetic tree of NBS genes in maize inbred line B73

区域, 还有的非标准 NBS 类型基因序列只有部分保守的 NBS 基序。

在拟南芥中, 两大类 NBS 类型基因的分类依据其 N-端 TIR 结构的存在与否^[18]; 在玉米 B73 基因组中, 标准 NBS 类型基因主要区别在于 N-端结构和 C-端结构的差异, 所以本研究参考拟南芥和水稻的抗病基因分类方法, 通过 Pfam 数据库的分析, 按照 N-端和 LRR 区域基序的区别对这些 NBS 抗病基因进行分类, 共分为 11 类(表 1)。大多数标准 NBS 类型基因的 N-端含有一些未知基序, 都统一以 X 表示其结构。经过 COILS 和 Pfam 分析, 共 59 个基因序列含有 CC 基序, 主要分为 CC-NBS-LRR 和 CC-NBS 两类。通过 Pfam 和 COILS 无法获取 N-端类型的共 91 个, 分为 NBS、NBS-NBS、NBS-LRR、NBS-NBS-LRR 和 NBS-X 6 类, 占有标准 NBS 类型基因的 59%。此外, 有一类在编码蛋白 N 端包含一个 TIR 结构的 NBS 抗病基因, 被称作为 TIR-NBS-LRR 类型, 它在双子叶植物中广为报道, 单子叶植物中未见报道。在拟南芥 149 条

NBS-LRR 基因中占有 94 条(约占 63.1%), 而在玉米所有 NBS-LRR 基因中未发现带有 TIR 结构, 这与同是单子叶禾本科作物的水稻十分相似^[19]。

2.3 NBS 类型基因的系统发生分析

165 个 NBS 类型基因的结构域进化树呈现两个明显分支(图 2), 且分支内抗病基因的数量差异较大。从图 2 还可以看出, 玉米基因组中 NBS 类型基因中含 CC 基序基因的 NBS 区域存在于不同的进化分支中, 不存在 CC 单独分支, 说明这两类基因没有单独的进化模式, 这同与在拟南芥^[20]和水稻^[21]上的分析结果类似。

2.4 NBS 类型基因的复制

在 165 个 NBS 类型基因中共发现 63 个基因发生了复制, 可分为 27 个基因家族, 其中最大家族的成员为 6 个, 平均家族成员数 2.33。通过比对 NBS 类型基因的复制事件, 发现玉米与拟南芥较为相似, 但它们复制的多基因数量、多基因所占的比率都明显少于水稻, 在水稻中一半以上 NBS 抗病基因都发生了复制, 且基因家族多达 93 个^[22]。

3 讨论

利用生物信息学方法, 在玉米自交系 B73 的全基因组中发现 165 个含有 NBS 结构的基因, 约占基因总数的 0.3%, 而在拟南芥中这一比例约为 0.47%^[23], 在水稻中约为 1%^[24]。无论从绝对数量上还是相对的比例上, 玉米基因组中的 NBS 类型基因明显少于同是单子叶禾本科作物的水稻。本研究结果显示玉米中 NBS 结构基因发生的复制事件远远少于水稻, 水稻中一半以上 NBS 类型基因都发生了复制, 也就是说虽然水稻 NBS 类型基因数量是增多了, 但一半以上基因都是由于复制而产生的, 基因间的同源性很高^[25], 这可能是玉米 NBS 类型基因明显少于水稻的一个重要原因之一。

蛋白 N 端包含一个 TIR 结构的 NBS 抗病基因, 在双子叶植物中广为报道, 而本研究在对搜索到的 165 个玉米 NBS 抗病基因进行归类时, 却未能发现带有 TIR 结构的基因, 这同水稻十分相似。目前含有 TIR 结构的 NBS 类型抗病基因在单子叶植物中还未见报道, 但是否所有单子叶植物 NBS 类型抗病基因中都没有 TIR 结构, 有待进一步研究。

表 1 玉米 B73 基因组中 NBS 抗病基因分类及数目
Table 1 Number and category of nucleotide-binding site (NBS) disease-resistance genes in maize

基因类别 Type of gene	简写代号 Code abbreviation	基因数目 Numbers of genes
规则 NBS 基因 Regular NBS gene		153
CC-NBS-LRR	CNL	19
CC-NBS	CN	40
NBS	N	47
NBS-NBS	NN	8
NBS-LRR	NL	29
NBS-NBS-LRR	NNL	3
NBS-X	NX	4
X-NBS	XN	3
非规则 NBS 基因 Nonregular NBS gene		12
NBS	NGN	8
NBS-LRR	NGNL	4
NBS 基因总计 Total NBS genes		165

表 2 玉米、拟南芥和水稻基因组 NBS 类型抗病基因的基因复制比较
Table 2 Comparison of gene duplications of NBS-encoding R-genes in genomes of maize, Arabidopsis, and rice

物种 Species	NBS 基因 NBS gene				基因家族 Gene family		
	总数 Total	单基因 SG	多基因 MG	多基因/单基因比值 SG/MG ratio	数目 No. of families	最大成员数 Maximal member	平均家族成员数 Average member
玉米 Maize	165	102	63	0.62	27	6	2.33
拟南芥 Arabidopsis	174	93	81	0.87	25	7	3.24
水稻 Rice	519	240	279	1.16	93	10	3.00

表中部分数据源自 Yang 等^[25]。

Partial data in the table are from Yang et al.^[25]. SG: single gene; MG: multigene.

在对全基因组内 NBS 类型抗病基因进行系统进化树分析时,发现明显的两大分支,这与水稻的星状发散型分布^[26-27]有很大差异。虽然拟南芥中也存在两个明显的分枝,但那是由于 TIR 结构和 CC 结构的分化造成的^[28],由于玉米中并不含有 TIR 结构,且其两个明显的分枝中都分别含有 NL 和 CNL 等类型基因,并没有其他显著的特点,因此两者也存在较大差异。

References

- [1] Duan M-X(段民孝), Zhao J-R(赵久然), Wang Y-D(王元东). Study progress in starch of corn Kerne. *J Maize Sci* (玉米科学), 2002, 10(1): 29–32 (in Chinese with English abstract)
- [2] Richly E, Kurth J, Leister D. Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. *Mol Biol Evol*, 2002, 19: 76–84
- [3] Staskawicz B J, Ausubel F M, Baker B J, Ellis J G, Jones J D. Molecular genetics of plant disease resistance. *Science*, 1995, 268: 661–667
- [4] Holub E. Arms race is an ancient history in *Arabidopsis*, the wildflower. *Nat Rev Genet*, 2001, 2: 516–527
- [5] Meyers B C, Kaushik S, Nandety R S. Evolving disease resistance genes. *Curr Opin Plant Biol*, 2005, 8: 129–134
- [6] Dangl L J, McDowell J M. Two modes of pathogen recognition by plants. *Proc Natl Acad Sci USA*, 2006, 103: 8575–8576
- [7] Noutoshi Y, Ito T, Seki M, Nakashita H, Yoshida S, Marco Y, Shirasu K, Shinozaki K. A single amino acid insertion in the WRKY domain of the *Arabidopsis* TIR-NBS-LRR-WRKY-type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. *Plant J*, 2005, 43: 873–888
- [8] Montesinos E, Bonaterra A, Badosa E, Francés J, Alemany J, Llorente I, Moragrega C. Plant-microbe interactions and the new biotechnological methods of plant disease control. *Int Microbiol*, 2002, 5: 169–175
- [9] Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science*, 1991, 252: 1162–1164
- [10] Xu J H, Messing J. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc Natl Acad Sci USA*, 2008, 105: 105–118
- [11] Tian D, Traw M B, Chen J Q, Kreitman M, Bergelson J. Fitness cost of *R*-gene mediated resistance in *Arabidopsis thaliana*. *Nature*, 2003, 423: 74–77
- [12] Chen G, Pan D, Zhou Y, Lin S, Ke X. Diversity and evolutionary relationship of nucleotide binding site-encoding disease-resistance gene analogues in sweet potato (*Ipomoea batatas* Lam.). *J Biosci*, 2007, 32: 713–721
- [13] Xu Q, Wen X, Deng X. Isolation of TIR and non-TIR NBS-LRR resistance gene analogues and identification of molecular markers linked to a powdery mildew resistance locus in chestnut rose (*Rosa roxburghii* Tratt). *Theor Appl Genet*, 2005, 111: 819–830
- [14] Plocik A, Layden J, Kesseli R. Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce, and chicory. *Mol Phylogenet Evol*, 2004, 31: 153–163
- [15] Zhou T, Wang Y, Chen J Q, Araki H, Jing Z, Jiang K, Shen J, Tian D C. Genome-wide identification of NBS genes in *japonica* rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics*, 2004, 271: 402–415
- [16] Bai J, Pennill L A, Ning J, Lee S W, Ramalingam J, Webb C A, Zhao B, Sun Q, Nelson J C, Leach J E, Hulbert S H. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res*, 2002, 12: 1871–1884
- [17] Peng G-Z(彭贵子), Chen L-L(陈玲玲), Tian D-C(田大成). Progress in the study of gene duplication. *Hereditas* (遗传), 2006, 28(7): 886–892 (in Chinese with English abstract)
- [18] Meyers B C, Morgante M, Michelmore R W. TIR-X and TIR-NBS proteins: Two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. *Plant J*, 2002, 32: 77–92
- [19] Meyers B C, Kozik A, Griego A, Kuang H, Michelmore R W. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell*, 2003, 15: 809–834
- [20] Noir S, Combes M C, Anthony F, Lashermes P. Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). *Mol Genet Genomic*, 2001, 265: 654–662
- [21] López C E, Zuluaga A P, Cooke R, Delseny M, Tohme J, Verdier V. Isolation of resistance gene candidates (RGCs) and characterization of an RGC cluster in cassava. *Mol Genet Genomic*, 2003, 269: 658–671
- [22] Tan X P, Meyers B C, Kozik A, West M A, Morgante M, St Clair D A, Bent A F, Michelmore R W. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biol*, 2007, 7: 56
- [23] Mondragón-Palomino M, Meyers B C, Michelmore R W, Gaut B S. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res*, 2002, 12: 1305–1315
- [24] Yaish M W, Sáenz de Miera L E, Pérez de la Vega M. Isolation of a family of resistance gene analogue sequences of the nucleotide binding site (NBS) type from Lens species. *Genome*, 2004, 47: 650–659
- [25] Yang S H, Zhang X H, Yue J X, Tian D C, Chen J Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Genet Genomics*, 2008, 280: 187–198
- [26] Kohler A, Rinaldi C, Duplessis S, Bauchner M, Geelen D, Duchaussoy F, Meyers B C, Boerjan W, Martin F. Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Mol Biol*, 2008, 66: 619–636
- [27] Pan Q, Wendel J, Fluhr R. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J Mol Evol*, 2000, 50: 203–213
- [28] Shirano Y, Kachroo P, Shah J, Klessig D F. A gain-of-function mutation in an *Arabidopsis* Toll Interleukin receptor-nucleotide binding site-leucine-rich repeat type R gene triggers defense responses and results in enhanced disease resistance. *Plant Cell*, 2002, 14: 3149–3162