

大豆 *Glyma08g02580* 及其同源蛋白的生物信息学分析

刘冬冬,王 洋,柏 锡

(东北农业大学 生命科学学院,黑龙江 哈尔滨 150030)

摘要: WRKY 蛋白是一个植物特有的转录因子,最古老的 WRKY 转录因子具有 2 个高度保守的 WRKY 结构域,大概起源于 15 亿~20 亿年前的真核生物。虽然所有的 WRKY 蛋白质主要通过特异地结合靶基因启动子区域的 W 盒序列而调控其表达,但各家族成员的生物学功能存在着各自的特异性。通过生物信息学方法,结合大豆基因组数据库、高等植物基因组注释数据库和 NCBI 等相关数据库,识别、筛选并获得了大豆 WRKY 基因家族的蛋白序列,以实验室前期克隆得到大豆基因 *Glyma08g02580.1* 为目的基因,分析其在拟南芥、水稻等高等植物中的同源蛋白,并对其进行了系统进化分析;同时利用 GEO 的芯片数据对大豆在胁迫下的根和叶基因表达谱进行了分析,通过相关分析,获得了大豆基因组中全部的 WRKY 蛋白及对应的分类结果,由表达谱分析结果进一步证明了 WRKY 基因表达的组织特异性。

关键词: WRKY;转录因子;生物信息学;系统进化分析;基因表达谱

中图分类号: S565.1

文献标识码: A

DOI: 10.11861/j.issn.1000-9841.2014.05.0648

Bioinformatics Analysis on Transcription Factor *Glyma08g02580* and Its Homology Proteins

LIU Dong-dong, WANG Yang, BAI Xi

(Life Science College of Northeast Agricultural University, Harbin 150030, China)

Abstract: Genes containing the WRKY domain are plant-specific transcriptional regulators and are expressed in various developmental stages and tissues. The most archaic WRKY transcription factors, which may originate some 1.5 billion to 2 billion years ago in eukaryotes, contain two highly conserved WRKY domain. Although all WRKY proteins regulate mainly the expression of target genes by specifically binding to the (T)(T)GACC(A/T)(W box) sequences of their promoter, the biological function of each member is specific. In the study, a bioinformatics analysis on WRKY family genes in *Glycine max* was performed. Based on bioinformatics methods, taking advantage of soybean database combined with public database (NCBI, TIGR, TAIR), 196 WRKY protein from soybean genome and download gene expression profile from GEO database were identified. Meanwhile, 197 homology proteins of *Glyma08g02580.1* in other plants were identified, such as *Arabidopsis* and *Oryza sativa*. The phylogenetic of 197 WRKY family protein sequence and domain characteristics were analyzed. The results of the investigation could be definitely provide a significant foundation for further research on function analysis of *Glycine max* WRKY gene family.

Key words: WRKY; Transcription factor; Bioinformatics; Phylogenetic analysis; Gene expression profile

WRKY 转录因子在调节植物生长发育、抵御外界环境胁迫等方面具有重要作用^[1-2],与应答植物生物胁迫和非生物胁迫密切相关^[3-4]。WRKY 是一类植物特有的转录调控因子,保守的 WRKY 结构域能特异性识别(T)(T)TGAC(C/T)序列,即 W-box,进而发挥作用^[5];而 W-box 主要位于抗病相关的启动子区域^[6-8],在植物损伤^[9-10]以及一些衰老相关基因^[11-12]、水杨酸诱导的基因的启动子区域^[13]也普遍存在。如 Yoda 等^[14]从烟草中克隆得到的 TIZZ 基因就是一个转录因子,实验证明转该基因的烟草受 TMV 病毒诱导后表达,但并不受 SA 的诱导,说明 TIZZ 是一个不依赖 SA 的防御性 WRKY 转录因子;Mare 等^[15]从大麦中分离得到的 Hv-WRKY38 转录

因子证明了 WRKY 与植物的冷胁迫和干旱胁迫应答相关,在非生物胁迫调节方面具有重要作用。此外,在植物的衰老、发育代谢、抗病低温及赤霉素信号传导等重要生理生化活动中也发现了 WRKY 基因的表达,表明 WRKY 转录因子普遍参与植物的生命活动^[16-18]。

本研究以实验室前期克隆得到的大豆 *Glyma08g02580.1* 基因^[19]为目的基因,通过生物信息学方法,在大豆基因组数据库及公共数据库(NCBI、TIGR、TAIR)基础上,对大豆中 *Glyma08g02580.1* 的同源蛋白及其他高等植物中的近缘蛋白基因进行了分析,对其序列结构、分子进化及基因表达等方面进行了分析与总结,以期为该基因的后续功能分

收稿日期:2014-01-05

基金项目:教育部高等学校博士学科点专项科研基金(20102325120002);国家自然科学基金(31371544);教育部科学技术研究重点项目(212049)。

第一作者简介:刘冬冬(1986-),男,硕士,主要从事植物分子遗传学研究。E-mail:liudongdong0524521@hotmail.com。

通讯作者:柏锡(1975-),男,博士,副教授,主要从事植物分子生物学与逆境生理学研究。E-mail:baixi@neau.edu.cn。

析和应用提供参考。

1 材料与方法

1.1 数据库搜索与序列获取

利用 Glyma08g02580.1 保守结构域的氨基酸序列在 GenBank 数据库 (<http://www.ncbi.nlm.nih.gov/genbank/>) 进行序列相似性搜索,同时,通过 BLASTp 在大豆基因组注释数据库 (<http://www.phytozome.net/>) 进行搜索。然后,利用 PFAM 工具对候选蛋白质进行分析,若存在 WRKY 保守结构域,则保留该蛋白。搜索大豆 Glyma08g02580.1 在不同物种中的同源蛋白步骤同上,同时涉及到的数据库包括 TAIR 和 TIGR。

1.2 大豆 WRKY 蛋白序列分类分析

针对得到的大豆中与 Glyma08g02580.1 具有相同结构域的 WRKY 蛋白,借助 ORF (finder<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) 进行 ORF 预测,将预测到的每个 ORF 对应的氨基酸序列用 Pfam (<http://pfam.sanger.ac.uk/>) 和 CDD 软件 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) 进行验证,并最终通过 WRKY DNA-binding 结构域对大豆中的 WRKY 蛋白序列进行细化分类。

1.3 大豆 Glyma08g02580.1 同源序列比对与进化分析

在大豆基因组数据库中搜索与 Glyma08g02580.1 基因具有共同结构域的同源蛋白序列,通过 Clustal X 软件进行序列比对和保守结构域分析,借助 MEGA 5.0 软件对其进行系统进化树构建。进化树生成算法采用邻接法 (Neighbor-Joining, NJ),缺口设置为“complete Deletion”,校验参数 Bootstrap 重复值设为 1 000。大豆 Glyma08g02580.1 基因的 WRKY DNA-binding 保守结构域及锌指结构分析通过 Clustal X 软件和 CDD 软件进行在线预测分析。

1.4 大豆 WRKY 基因的组织表达特异性分析

筛选保留与目的基因 Glyma08g02580.1 具有同类型结构模式的基因,从 GEO 中下载大豆在 NaHCO₃ 胁迫下根和叶的表达数据,对其进行表达谱分析,包括表达模式分析和层次聚类分析,所用软件为 MeV4.0,从而以此为例说明大豆 WRKY 基因具有组织表达特异性。

2 结果与分析

2.1 数据库搜索与序列获取

针对 Glyma08g02580.1 保守结构域蛋白序列,在大豆基因组注释数据库 (Phytozome v10.9) 及 NCBI 等公共数据库 (TIGR、TAIR) 中进行相似序列搜索,去冗余后共保留 WRKY 基因序列 230 条,对应蛋白 196 条。

2.2 大豆 WRKY 蛋白序列分类分析

针对得到的大豆中与 Glyma08g02580.1 具有相同结构域的 WRKY 蛋白,通过 ORF 和 CDD 保守结构域预测分析结果进行分类,根据保守结构域数量及保守结构域前后序列特点共分成 8 类 (图 1)。

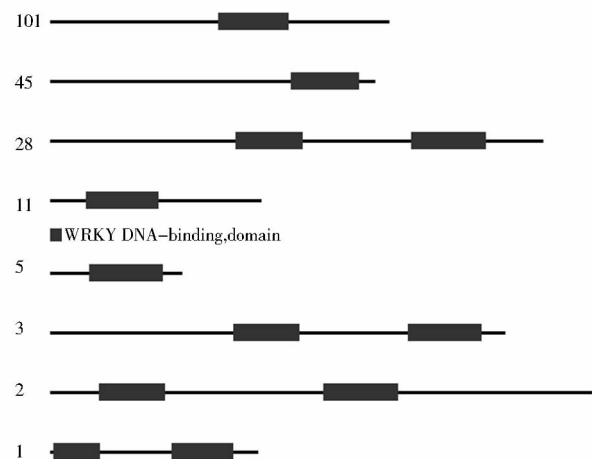


图 1 大豆 WRKY 蛋白序列分类

Fig. 1 The classification of WRKY proteins of *Glycine max*

由分类结果可以看出,大豆 WRKY 蛋白共被分成了 8 个子类,每个子类都包含了 1 ~ 2 个 WRKY DNA-binding 保守结构域,深色长方形代表其保守结构域。其中第一种结构共包含了 101 条蛋白序列,其序列模式为 gap-domain-gap 的结构域,第二种结构包含了 45 条序列,其序列模式为 gap-domain,第四种和第五种结构分别包含了 11 和 5 条蛋白,对应的序列模式为 domain-gap 和 domain,这 4 种结构序列都只包含了 1 个保守结构域;第三种、第六种、第七种和第八种结构都包含了 2 个保守结构域,对应的序列模式为 gap-domain-gap-domain-gap、gap-domain-gap-domain、domain-gap-domain-gap 和 domain-gap-domain。

2.3 Glyma08g02580.1 同源序列比对与进化分析

本实验室从大豆中克隆得到的基因 Glyma08g02580.1 所包含的结构域为第一种形式,结构域的所在区间为 128 ~ 191 碱基,氨基酸序列长度为 359 bp;在大豆基因组数据库 (phytozome) 中的同源蛋白为 Glyma05g36970.1,二者的覆盖度达到 99.4%,序列相似性为 89.4%,由此推测这两个基因可能是旁系同源序列,并且 Glyma05g36970.1 所具有的 WRKY DNA-binding 结构域所在区间为 134 ~ 194 碱基。同时,以 Glyma08g02580 为目的基因为植物基因组数据库中搜索拟南芥、水稻、苜蓿等其高等植物的蛋白同源序列,通过物种过滤保留搜索得到的 196 个同源蛋白基因。

Glyma08g02580.1 基因的特征是有一个 DNA-binding 结构域,本研究根据预测得到的 WRKY 域的氨基酸序列对 Glyma08g02580.1 及其

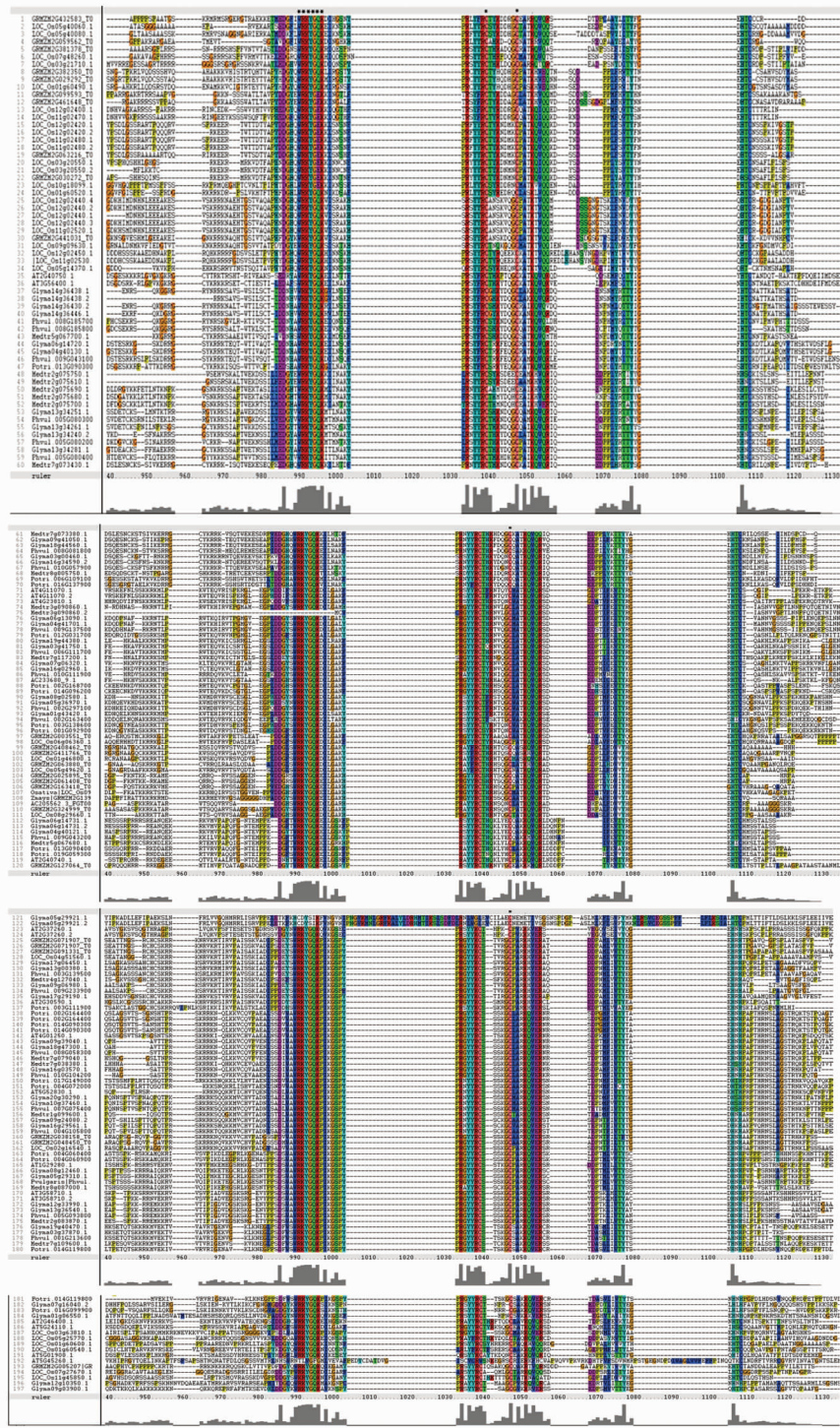


图3 *Glyma08g02580.1* 及其同源蛋白序列分析

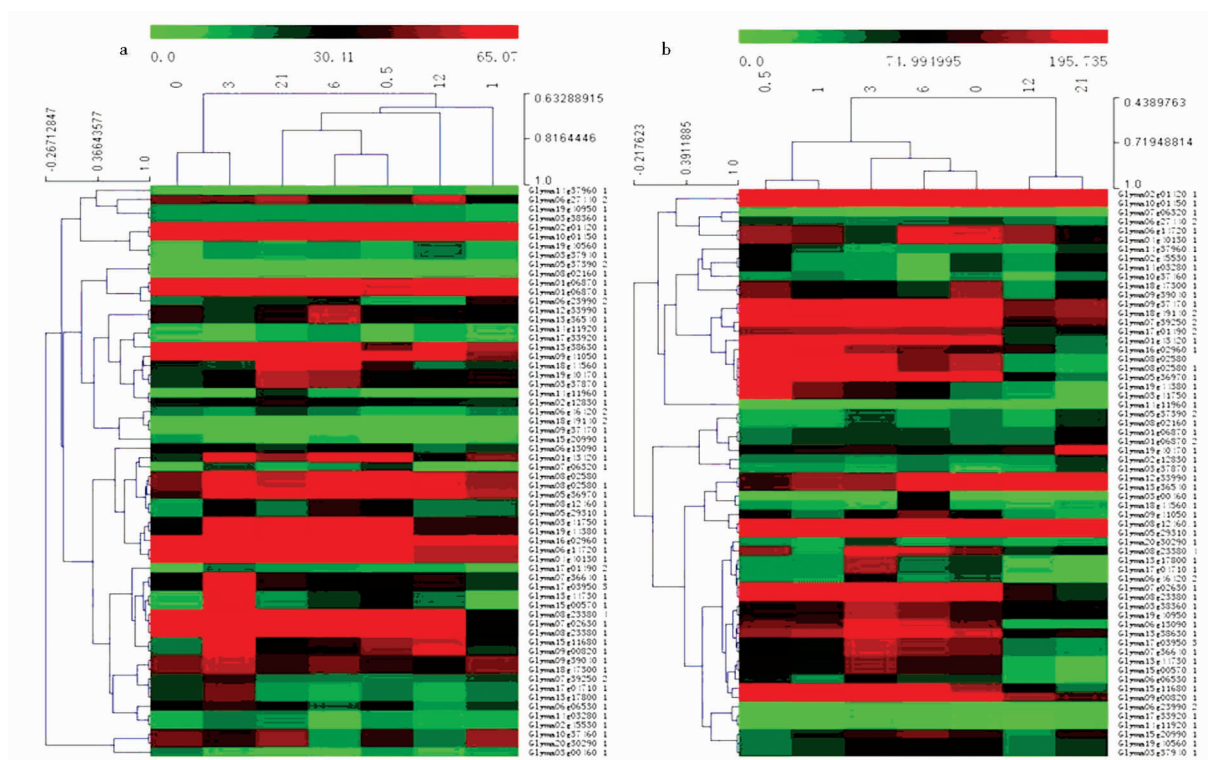
Fig. 3 Amino acid sequence alignment of *Glyma08g02580.1* and its homology proteins

2.4 大豆 *WRKY* 基因的组织表达特异性分析

WRKY 基因的表达具有组织特异性, 有些 *WRKY* 基因在不同胁迫下、不同器官中的表达情况也不相同^[22-24], 如 *Glyma08g02580.1* 这个 *WRKY* 基因在植物的根、叶、子叶、花及上胚轴中都有表达, 但是表达量不同。有些基因则可能只在 1 或者 2 种组织中表达, 并且即便是相同基因在不同胁迫条件下表达情况也会有明显差别, 这表明, 大豆中不同的 *WRKY* 基因表达部位不同, 表达量也各异, 同样,

在大豆防御过程中不同的 *WRKY* 基因也发挥着不同的作用。

本研究借用了从 NCBI 中下载的 NaHCO_3 胁迫下的大豆叶和根中基因表达数据 (GSE20323、GSE17883), 分析 *Glyma08g02580.1* 及其具有同类结构域分布模式的 101 条蛋白, 发现有 62 个基因在 NaCHO_3 胁迫下有表达, 分别对其在叶中和根中的基因表达情况进行层次聚类分析 (图 4) 及表达模式分析 (图 5)。

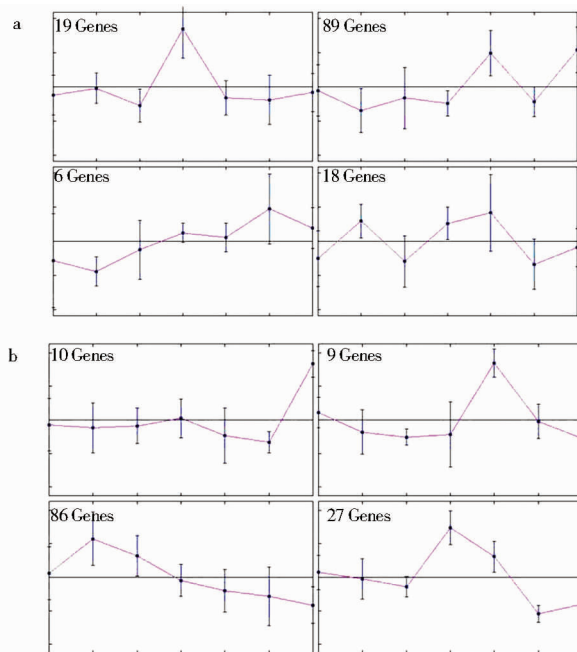


a: 叶; b: 根。

a: Leaf; b: Root.

图4 大豆 WRKY 蛋白表达模式层次聚类分析

Fig.4 Hierarchical cluster analysis of WRKY proteins of *Glycine max*



a: 叶; b: 根。

a: Leaf; b: Root.

图5 大豆 WRKY 蛋白表达模式聚类分析

Fig.5 Gene expression profile of WRKY proteins of *Glycine max*

从层次聚类分析表达谱(图4)可以看出,相同的WRKY基因在不同胁迫时间点以及组织中的表达是不同的。通过对比叶(图4a)和根(图4b)各时间点的基因表达很容易发现,目的基因 *Glyma08g02580.1* 在叶中各时间点的表达量普遍低于

在根中的表达量,分析其他WRKY基因的表达模式也可以发现类似规律,即相同基因在叶和根中的表达量不同,从而再次说明了WRKY基因在不同组织中的表达特异性。对根和叶的全部时间点中WRKY基因进行模式聚类分析,可以发现叶(图5a)与根

(图 5b)中所聚类的 4 种表达模式图截然不同,叶中在 3 h 时有部分基因的表达量达到了峰值,远远高于其他时间点的表达量,而根中对应的达到峰值的时间为 6 h,并且这两个时间点对应的基因也不同。从而进一步验证了,不同的 *WRKY* 基因在防御过程中所行使的功能是不同的。

3 讨 论

从分析结果可知,这 197 个 *WRKY* 转录因子在不同高等植物中所发挥不同功能的原因之一就是其对应不同的锌指结构。在植物中,蛋白功能的多样性的获得主要是通过表达的差异而不是蛋白质序列的差异,因此可以推测,锌指结构增加了 *WRKY* 蛋白的表达的多样性,从而可能使 *WRKY* 蛋白在植物防御过程中起到更加重要的作用。

大豆 Glyma08g02580 及其在大豆中的同源蛋白属于旁系同样关系,可以认为是通过基因复制而产生主要的差异,它们具有相似的编码区序列,所编码的蛋白质具有相同功能,但由于非编码区调控序列的不同,因此可能在不同组织、不同发育阶段或者不同环境条件下表达,从而具有组织表达特异性。基于从 NCBI GEO 数据库中下载的大豆在 NaHCO₃胁迫下不同时间点的芯片数据分析,对大豆 *WRKY* 基因在不同组织的表达情况有了大概了解,至于具体的表达特性及生物功能的识别还需后续有针对性地采用分子试验方法进行验证。

参考文献

- [1] Chen L, Song Y, Li S, et al. The role of WRKY transcription factors in plant abiotic stresses[J]. *Biochimica et Biophysica Acta*, 2012, 1819(2): 120-128.
- [2] Eulgem T, Rushton P J, Schmelzer E, et al. Early nuclear events in plant defence signaling: Rapid gene activation by WRKY transcription factors[J]. *The EMBO Journal*, 1999, 18: 4689-4699.
- [3] 贾翠玲. 葡萄 *WRKY* 基因超家族的进化分析[D]. 大连: 辽宁师范大学, 2010. (Jia C L. Phylogenetic analyses of vitis vinifera *WRKY* gene superfamily[D]. Dalian: Liaoning Normal University, 2010.)
- [4] Ciolkowski I, Wanke D, Birkenbihl R P, et al. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function[J]. *Plant Molecular Biology*, 2008, 68(1/2): 81-92.
- [5] Yokotani N, Sato Y, Tanabe S, et al. WRKY76 is a rice transcriptional repressor playing opposite roles in blast disease resistance and cold stress tolerance[J]. *Journal of Experimental Botany*, 2013, 64(16): 5085-5097.
- [6] Lan A, Huang J, Zhao W, et al. A salicylic acid-induced rice (*Oryza sativa* L.) transcription factor OsWRKY77 is involved in disease resistance of *Arabidopsis thaliana* [J]. *Plant Biology*, 2013, 15(3): 452-461.
- [7] Yu F, Huaxia Y, Lu W, et al. GhWRKY15, a member of the WRKY transcription factor family identified from cotton (*Gossypium hirsutum* L.), is involved in disease resistance and plant development[J]. *BMC Plant Biology*, 2012, 12: 12: 144.
- [8] Ulker B, Somssich I E. WRKY transcription factors: from DNA binding towards biological function[J]. *Current Opinion in Plant Biology*, 2004, 7(5): 491-498.
- [9] 宋钰, 荆绍娟, 余迪求. 水稻 *WRKY* 转录调控因子基因功能研究进展[J]. *中国水稻科学*, 2009, 23(5): 447-455. (Song Y, Jing S J, Yu D Q. Research progress on function analysis of rice *WRKY* genes[J]. *Chinese Journal of Rice Science*, 2009, 23(5): 447-455.)
- [10] Nakayama A, Fukushima S, Goto S, et al. Genome-wide identification of WRKY45-regulated genes that mediate benzothiadiazole-induced defense responses in rice[J]. *BMC Plant Biology*, 2013, 13: 150 doi:10.1186/1471-2229-13-150.
- [11] Dong J, Chen C, Chen Z. Expression profiles of the *Arabidopsis* *WRKY* gene superfamily during plant defense response[J]. *Plant Molecular Biology*, 2002, 51: 21-37.
- [12] Lan A, Huang J, Zhao W, et al. A salicylic acid-induced rice (*Oryza sativa* L.) transcription factor OsWRKY77 in disease resistance of *Arabidopsis thaliana* [J]. *Plant Biology*, 2013, 15(3): 452-461.
- [13] Hwang S H, Yie S W, Hwang D J. Heterologous expression of *OsWRKY6* gene in *Arabidopsis* activates the expression of defense related genes and enhances resistance to pathogens[J]. *Plant Science*, 2011, 181(3): 316-323.
- [14] Yoda H, Ogawa M, Yamaguchi Y, et al. Identification of early-responsive genes associated with the hypersensitive response to tobacco mosaic virus and characterization of a WRKY-type transcription factor in tobacco plants[J]. *Molecular Genet Genomics*, 2002, 267(2): 154-161.
- [15] Mare C, Mazzucotelli E, Crosatti C, et al. Hv-WRKY38: A new transcription factor involved in cold and drought response in barley [J]. *Plant Molecular Biology*, 2004, 55(3): 399-416.
- [16] 田云, 卢向阳, 彭丽莎, 等. 植物 *WRKY* 转录因子结构特点及其生物学功能[J]. *遗传*, 2006, 28(12): 1607-1612. (Tian Y, Lu X Y, Peng L S, et al. The structure and function of plant WRKY transcription factors[J]. *Hereditas*, 2006, 28(12): 1607-1612.)
- [17] Yu D Q, Chen L G, Zhang L P. Transcription factors WRKY superfamily: Origin, structure and function[J]. *Acta Botanica Yunnanica*, 2006, 28(1): 69-77.
- [18] Yu D, Chen C, Chen Z. Evidence for an important role of WRKY DNA binding protein in the regulation of NPR1 gene expression [J]. *The Plant Cell*, 2001, 13: 1527-1539.
- [19] Luo X, Bai X, Zhu D, et al. Expression of wild soybean *WRKY20* in *Arabidopsis* enhances drought tolerance and regulates ABA signaling [J]. *Journal of Experimental Botany*, 2013, 64(8): 2155-2169.
- [20] Yamasaki K, Kigawa T, Seki M, et al. DNA-binding domains of plant-specific transcription factors: structure, function, and evolution[J]. *Trends Plant Science*, 2013, 18(5): 267-276.
- [21] Cheng Y, Zhou Y, Yang Y, et al. Structural and functional analysis of VQ motif-containing proteins in *Arabidopsis* as interacting proteins of WRKY transcription factors[J]. *Plant Physiology*, 2012, 159(2): 810-825.
- [22] 李蕾, 谢丙炎, 戴小枫, 等. *WRKY* 转录因子及其在植物防御反应中的作用[J]. *分子植物育种*, 2005, 3(3): 401-408. (Li L, Xie B Y, Dai X F, et al. WRKY transcription factors and their roles in plant defense response[J]. *Molecular Plant Breeding*, 2005, 3(3): 401-408.)
- [23] 张颖, 蒋卫杰, 凌健, 等. *WRKY* 转录因子表达谱的研究进展[J]. *基因组学与应用生物学*, 2009, 28(4): 803-808. (Zhang Y, Jiang W J, Ling J, et al. Advance on expression profile of transcription factor WRKY[J]. *Genomics and Applied Biology*, 2009, 28(4): 803-808.)
- [24] Xiong W, Xu X, Zhang L, et al. Genome-wide analysis of the *WRKY* gene family in physic nut (*Jatropha curcas* L.) [J]. *Gene*, 2013, 524(2): 124-132.