

# 大豆 MYB 转录因子的全基因组鉴定及生物信息学分析

成舒飞,端木慧子,陈 超,刘艾林,肖佳雷,朱延明

(东北农业大学 农业生物功能基因重点实验室,黑龙江 哈尔滨 150030)

**摘要:**MYB 转录因子是植物中最大的转录因子家族之一,近年的研究表明其广泛参与多种生物学过程。MYB 转录因子在其 N 端含有其特有的 DNA 结合基序(MYB-binding domain),能够与 DNA 分子大沟结合而调控基因表达。本研究通过生物信息学手段,在全基因组水平上筛选鉴定出了 304 个大豆 MYB 类转录因子,并对其进行了分类和保守结构域分析;通过与拟南芥的 MYB 基因进行系统发生分析,将 304 个大豆 MYB 转录因子分为了 22 个亚类;染色体定位分析表明大豆 MYB 转录因子在所有染色体上都有分布,部分染色体间的 MYB 蛋白序列高度相似,表明此类基因在进化上具有相同的来源;利用野生大豆盐、碱胁迫下转录组数据,分析 MYB 基因的表达模式,发现部分 MYB 转录因子能够响应盐、碱胁迫的诱导,且在盐、碱胁迫下具有不同的表达模式。

**关键词:**大豆;MYB;转录因子;生物信息学

**中图分类号:**S565. 1      **文献标识码:**A      **DOI:**10. 11861/j. issn. 1000-9841. 2016. 01. 0052

## Whole Genome Identification of Soybean MYB Transcription Factors and Bioinformatics Analysis

CHENG Shu-fei, DUANMU Hui-zi, CHEN Chao, LIU Ai-lin, XIAO Jia-lei, ZHU Yan-ming

(Key Laboratory of Agricultural Biological Functional Genes, Northeast Agricultural University, Harbin 150030, China)

**Abstract:** MYB transcription factor family is one of the largest families of transcription factors in plants. Recent research indicated MYB transcription factors participate in variety kinds of biological processes. MYB transcription factors have a unique DNA-binding domain in their N terminal, which can combine with DNA molecules and regulate the expression of downstream genes. In this study we conducted a genome-wide analysis to identify, classified 304 MYB transcription factors of soybean and analyzed their conserved domain. These MYB genes were classified into 22 subgroups by phylogenetic analysis. Chromosomal location analysis was conducted. Further analysis showed that part of the MYB genes among different chromosomal showed highly similarity. Transcriptome analysis of the MYB genes showed that most of them could response to salinity-alkalinity stress.

**Keywords:** Soybean; MYB; Transcription factors; Bioinformatics

转录因子又称反式作用因子,是真核生物细胞内一类能够与特异基因启动子结合,进而调控下游基因的表达的蛋白质分子。植物中常见的转录因子有 WRKY、HD-ZIP、ERF、bZIP、MYB 等家族,而 MYB 转录因子是植物转录因子中数量最多,功能最为多样化的家族之一<sup>[1-2]</sup>。

植物 MYB 转录因子的典型特征是其含有的 MYB-DNA 结合结构域(MYB-binding domain),是一段含有约 51 ~ 52 个氨基酸残基的不完全保守肽段<sup>[3]</sup>。在该结构域中分布有 3 个相对保守的色氨酸残基(互相间隔 18 或 19 个氨基酸残基),起疏水核心的作用,对维持 MYB 转录因子的构象具有重要作用。MYB 转录因子通常含有 1 ~ 4 个串联的 MYB-DNA 结合结构域,因此根据其含有的结构域数目不同可将 MYB 转录因为分为 4 类,1R-MYB、

R2R3-MYB、R1R2R3-MYB 和 4R-MYB<sup>[4]</sup>。其中 R3 结构域的第一个色氨酸残基通常被亮氨酸、异亮氨酸或苯丙氨酸等所取代。

大豆是重要的粮食和经济作物。东北地区是中国大豆的主产区,其在生长发育中易受多种非生物胁迫(尤其是盐碱胁迫)的影响而减产。研究表明,MYB 转录因子在植物非生物胁迫应答过程中起着重要作用<sup>[5]</sup>。本研究拟对栽培大豆中的 MYB 转录因子进行全基因组水平的鉴定、分类及盐碱胁迫下的表达模式分析<sup>[6]</sup>,为后续大豆 MYB 转录因子的研究提供参考。

### 1 材料与方法

#### 1.1 大豆 MYB 蛋白序列的获取

用大豆已知的 MYB 转录因子蛋白序列进行序

收稿日期:2014-12-30  
基金项目:转基因生物新品种培育重大专项(2011ZX08004-002);国家自然科学基金(31171578);黑龙江省高校科技创新团队建设计划(2011TD055);国家基础科学人才培养基金项目(J1210069)。  
第一作者简介:成舒飞(1988-),男,硕士,主要从事生物信息学研究。E-mail:sf\_cheng@163.com。  
通讯作者:朱延明(1955-),男,教授,博导,主要从事植物基因工程与分子生物学研究。E-mail:ymzhu2001@neau.edu.cn。

列比对,对对比结果进行分析处理得到了长度为 95 个氨基酸残基的高度保守的 MYB 转录因子片段。用该片段对植物基因组数据库 Phytozome (<http://www.phytozome.net>)中的大豆基因组进行 BLASTP 搜索<sup>[7]</sup>,得到的序列去除短序列及冗余转录本后利用 PFAM 对每个蛋白序列进行蛋白家族分析<sup>[8-9]</sup>,去除不含 MYB-DNA 结合结构域的序列,同时结合植物转录因子数据库(PlantTDB)对得到的序列进行进一步鉴定<sup>[10]</sup>。

### 1.2 大豆 MYB 蛋白序列的分类与基序分析

对得到 MYB 蛋白序列进行批量 SMART (<http://smart.embl-heidelberg.de/>) 搜索<sup>[11]</sup>,根据蛋白序列所含有的 DNA-binding 结构域的数量对得到的序列进行分类<sup>[1]</sup>。利用 MEGA 5.0 对含量最多的 R2R3-MYB 类转录因子蛋白序列进行序列比对分析<sup>[12]</sup>,并保存比对结果。将得到的比对结果删除不保守区域和 GAP 后,利用 Weblogo (<http://weblogo.berkeley.edu/>)对 MYB 的 DNA 结合结构域进行保守氨基酸基序分析<sup>[13-14]</sup>。

### 1.3 大豆 MYB 基因系统发生分析及染色体定位

为研究大豆 MYB 转录因子家族的进化情况,利用 MEGA 5.0 软件,采用近邻法(Neighbor-Joining)构建了包含拟南芥 22 个已知功能的 MYB 转录因子及大豆 MYB 转录因子的系统发生树<sup>[15]</sup>。

根据 MYB 蛋白序列的 GeneID 在 Phytozome 数据库中搜索得到其对应的 Locus,并在 NCBI 基因组数据库中获得了大豆 20 条染色体的长度信息。利用 MapInspect 软件构建了染色体定位图。以每条 MYB 蛋白序列为查询序列,以得到的所有的 MYB 蛋白序列为数据库进行了本地 Blast,对与每一条 MYB 蛋白相似度最高的 MYB 蛋白序列进行统计分析,以揭示 MYB 基因在染色体间的复制行为。

### 1.4 野生大豆盐碱胁迫下 MYB 基因在根中的表达模式聚类分析

研究表明,MYB 转录因子在植物非生物胁迫中起着重要作用。野生大豆与栽培大豆具有较高的同源性,为揭示大豆 MYB 基因在盐碱胁迫下的表达情况,采用 Matlab 对实验室前期获得的野生大豆盐、碱胁迫下根的表达谱数据中的差异表达 MYB 基因进行表达模式聚类分析<sup>[16]</sup>。

## 2 结果与分析

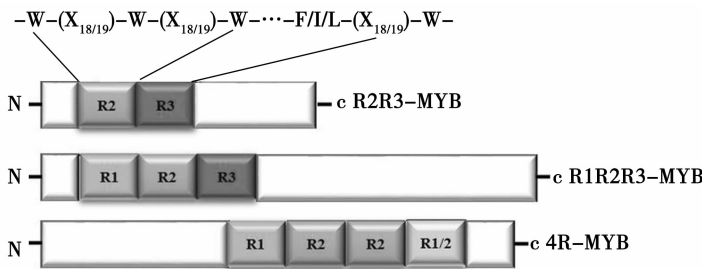
### 2.1 大豆 MYB 蛋白序列的获取

用已知的几个大豆 MYB 蛋白序列进行序列比对,获取了长度为 95 个氨基酸残基的 MYB 蛋白保守结构域。用该序列在 Phytozome 数据库中进行 BLASTP 搜索得到 342 条蛋白序列,去除短序列及冗余转录本后,剩余蛋白序列经 PFAM 数据库筛选去除不含 MYB 结构域的序列,最终得到 MYB 蛋白序列 304 条。

### 2.2 大豆 MYB 蛋白序列的分类与基序分析

对得到的 304 条 MYB 蛋白序列进行批量 SMART 搜索后按照所含有的 DNA binding domain 的数量进行分类(图 1)。含有 2 个 DNA binding domain 的 MYB 为 R2R3-MYB 类转录因子,含有 3 个 DNA binding domain 的 MYB 为 R1R2R3-MYB 类转录因子,同理含有 4 个 DNA 结合结构域的序列为 4R-MYB 类转录因子。共鉴定出 R2R3-MYB 基因 294 个, R1R2R3-MYB 基因 8 个, 4R-MYB 基因 2 个。

将在大豆中含量最为丰富的 R2R3-MYB 转录因子蛋白序列共 294 条,利用 MEGA 5.0 进行序列比对后,去除 DNA binding domain 之外的不保守区域,利用 Weblogo 对 R2R3-MYB 转录因子蛋白的 DNA 结合结构域进行基序分析(图 2)。



W: 色氨酸残基; F: 苯丙氨酸残基; I: 异亮氨酸残基; L: 亮氨酸残基; X: 其它氨基酸残基。  
W: Tryptophane; F: Phenylalanine; I: Isoleucine; L: Leucine; O: Other amino acid.

图 1 大豆 MYB 转录因子分类示意图

Fig. 1 Classification of MYB transcription factors of *Glycine max*

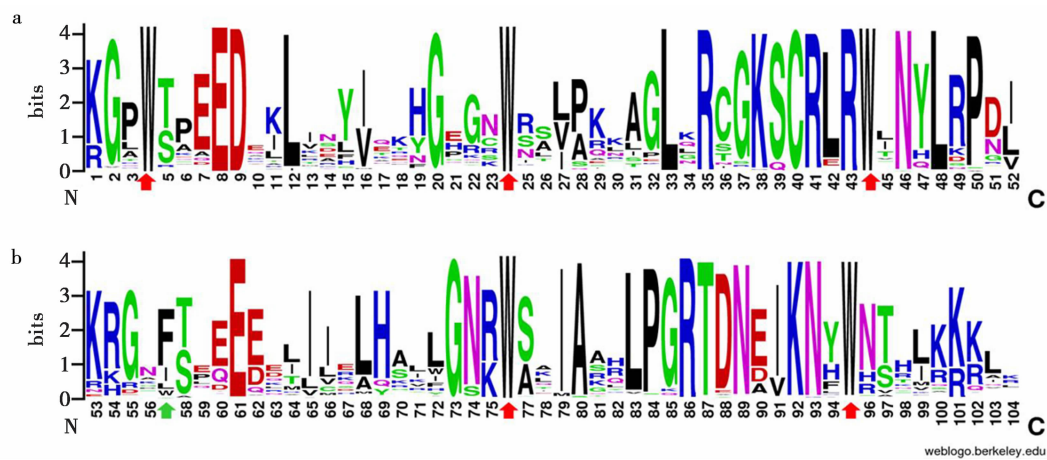


图2 大豆 R2R3-MYB 转录因子高度保守的 DNA binding 结构域

Fig. 2 Highly conserved DNA binding domain of soybean R2R3-MYB transcription factors

图中字母高度表示其在对应的位点上该氨基酸残基的保守程度,字母越高表示其对应氨基酸在该位点上出现概率越大进而保守程度越高。长度约为 52 个氨基酸残基的 R2 结构域(图 2a)中含有 3 个极其保守的色氨酸残基,而 R3 结构域(图 2b)的第一个色氨酸残基一般会被其它疏水氨基酸所取代(图 2,绿色箭头处)。如图 2 所示,替代色氨酸残基的通常为苯丙氨酸(F),也有异亮氨酸(I)以及亮氨酸(L)参与色氨酸残基的替代。R3 结构域的第二和第三个色氨酸也高度保守。除了高度保守的色氨酸残基以外,R2 及 R3 结构域中也存在大量的其它保守氨基酸残基,如结构域起始的赖氨酸(K)精氨酸(R)甘氨酸(G)序列、在 R2 结构域中的甘氨酸(G)赖氨酸(K)丝氨酸(S)半胱氨酸(C)精氨酸(R)亮氨酸(L)精氨酸(R)序列以及 R3 结构中存在的亮氨酸(L)脯氨酸(P)甘氨酸(G)精氨酸(R)苏氨酸(T)天冬氨酸(D)天冬酰胺(N)序列。推测以上保守氨基酸残基与色氨酸残基共同维持 MYB 转录因子的螺旋-转角-螺旋(HTH)结构,与 DNA 分子的大沟结合进而行使功能。

2.3 大豆 MYB 基因系统发生分析及染色体定位

采用近邻法构建了包含有 304 条大豆 MYB 蛋白序列和 22 条已知功能的拟南芥 MYB 蛋白的系统发生树。根据大豆 MYB 蛋白与拟南芥 MYB 蛋白的同源关系,将 304 个大豆 MYB 蛋白分为 22 个亚家族(图 3)。系统发生分析表明,某些大豆 MYB 蛋白与拟南芥中次生代谢相关 MYB 蛋白聚为一个亚家族(S1,S3,S4,S8,S9,S10),推测这类 MYB 蛋白参与了大豆的次生代谢相关生物过程。部分大豆 MYB 蛋白与拟南芥中发育相关的 MYB 转录因子聚为一个亚家族(S2,S13,S15,S18,S19),此类 MYB 转录因子可能在大豆在不同发育时期中起着重要作用。据报道,部分拟南芥 MYB 转录因子参与了植

物对生物及非生物胁迫的响应。拟南芥 AtMYB60 参与了植物对于干旱胁迫的响应,在盐胁迫下 AtMYB2 基因能够启动下游盐胁迫相关基因的表达。在图 3 的系统发生树中有部分大豆 MYB 转录因子基因与拟南芥响应生物及非生物胁迫的相关 MYB 转录因子聚为一个亚家族(S6,S7,S16,S21),该类大豆 MYB 蛋白可能参与了大豆对生物及非生物胁迫的响应。

从 Phytozome 数据库中查询 304 个大豆 MYB 基因的 Locus,并记录其 GeneID 及 Locus 起始位置。从 NCBI 基因组数据库中获取大豆 20 条染色体长度并记录。将以上数据导入 MapInspect 软件构建了大豆 MYB 转录因子基因染色体定位图谱。结果如图 4,大豆的 MYB 基因在所有染色体上均有分布,但大部分 MYB 倾向于分布在染色体的两端,只有极少数的几个 MYB 基因分布在染色体中段。而 MYB 基因在染色体间的分布却有较大差异,第 6、13、19 号染色体分布有 21 个 MYB 基因,数量最多,而第 16 号染色体上分布的 MYB 基因数量最少,只有 9 个。

将 304 条 MYB 蛋白序列作为查询数据库,对 MYB 蛋白序列逐一进行本地 Blast 后发现,部分 MYB 转录因子在染色体间具有高度的同源性。例如,1 号染色体上的 6 个 MYB 基因与 2 号染色体上的 6 个 MYB 基因(黄色框线内)、3 号染色体上的 8 个 MYB 基因分别与 19 号染色体上的 8 个 MYB 基因(红色框线内)、1 号染色体上靠近 3'端的 6 个 MYB 基因与 11 号染色体上靠近 5'端的 6 个 MYB 基因(蓝色框线内),均具有极高的序列相似性。因此推断,MYB 基因可能因大豆基因组的大片段复制或染色体倍增从而产生了庞大的 MYB 转录因子家族。



图 3 基于近邻法构建的大豆 MYB 蛋白系统发生树  
Fig. 3 Phylogenetic tree of soybean MYB protein based on NJ method

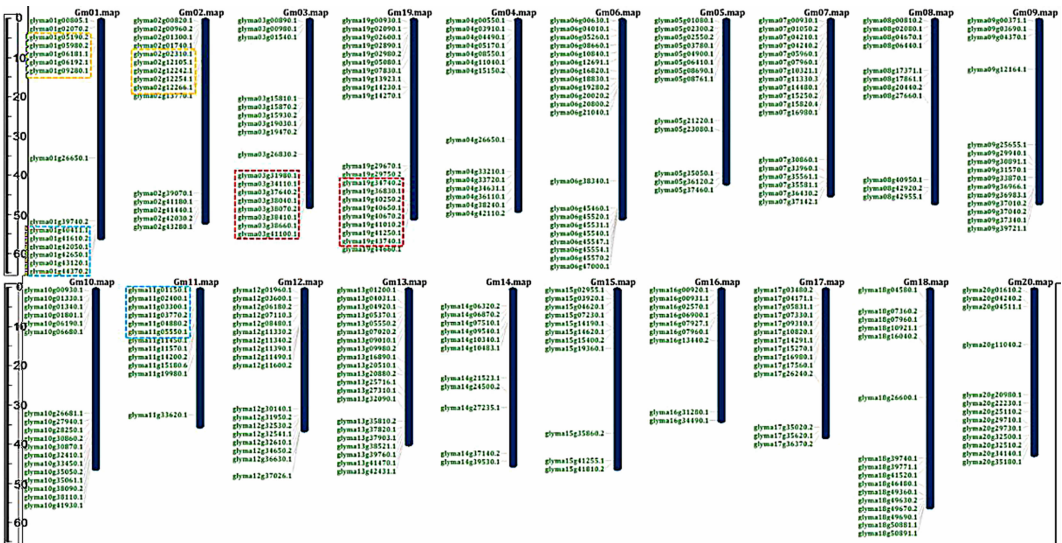


图 4 大豆 MYB 转录因子染色体定位

Fig. 4 Chromosomal locations of soybean MYB transcription factors

2.4 野生大豆盐碱胁迫下 MYB 基因在根中的表达模式聚类分析

由于野生大豆与大豆在进化上的高度同源性,为研究大豆 MYB 转录因子在盐碱胁迫下的响应情况,我们利用了野生大豆盐碱胁迫 6 h 后根的转录组数据构建了野生大豆 MYB 转录因子基因在盐、碱

胁迫下表达模式聚类图。如图 5 所示,在鉴定出来的 304 个大豆 MYB 转录因子中,有 48 个基因响应盐、碱胁迫。在盐、碱胁迫下,大部分差异表达 MYB 基因呈现明显的上调表达趋势,且表达量变化都在 2 倍以上,只有小部分 MYB 基因下调表达。另外,一些 MYB 转录因子在盐、碱胁迫下呈现相反的表达



模式,如 Glyma07g30860、Glyma20g34140 等在盐胁迫下上调表达,而在碱胁迫下呈现下调表达的趋势。已有研究表明 Glyma07g30860 基因的拟南芥同源基因 *AtMYB102* 会受到渗透胁迫的诱导而表达,是拟南芥中公认的盐胁迫响应基因<sup>[17]</sup>。上述结果

表明,部分 MYB 转录因子响应盐碱胁迫,可能是通过表达量的变化,进而启动下游应答基因,增强植物在盐碱胁迫下的耐性,而植物对于盐胁迫和碱胁迫具有不同的响应机制。

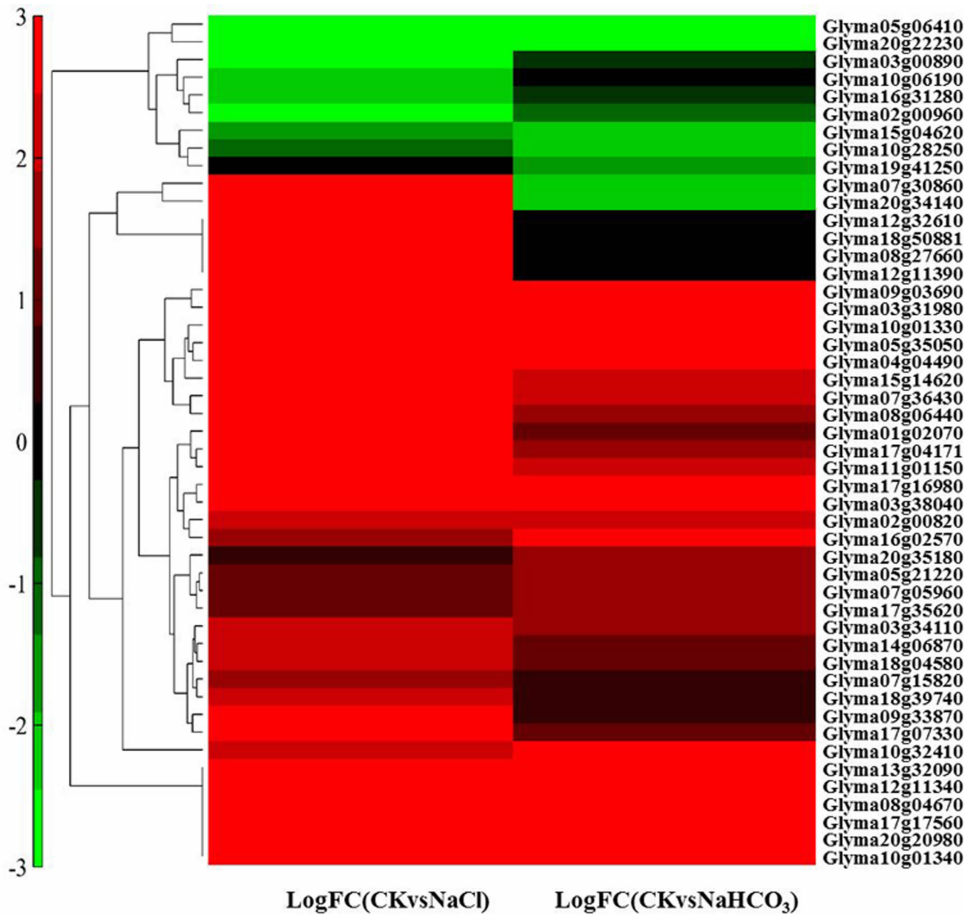


图5 野生大豆根在盐碱胁迫下差异表达 MYB 基因表达模式聚类

Fig. 5 Clustering analysis of differentially expressed MYB genes in wild soybean root under saline-alkali stress

### 3 结论与讨论

MYB 转录因子是植物基因组中数量最大的家族之一,在植物中按照 MYB 蛋白所含有的 DNA-binding 结构域的数量可将 MYB 转录因子分为 4 类,包括 1R-MYB、R2R3-MYB、R1R2R3-MYB 和 4R-MYB。而植物中最常见的 MYB 转录因子属于 R2R3-MYB 类转录因子。本研究在全基因组水平上通过一系列的生物信息学方法,首次在全基因组水平上对大豆 MYB 转录因子进行了筛选鉴定。共鉴定出 304 个 MYB 转录因子,其中 R2R3-MYB 基因 294 个,R1R2R3-MYB 基因 8 个,4R-MYB 基因 2 个。对 294 条大豆 R2R3-MYB 蛋白序列进行保守结构域分析表明,其含有 MYB 转录因子所特有 DNA 结合结构域,R2R3-MYB 转录因子中的 R3 结

构域第一个氨基酸通常被苯丙氨酸取代。将拟南芥 MYB 蛋白序列与鉴定出的 304 个大豆 MYB 蛋白序列进行序列比对表明,大豆的 MYB 基因与拟南芥 MYB 基因在 N 端具有高度同源性,暗示 MYB 转录因子是一个古老的基因家族,其特有的 N 端结构域对于保持 MYB 蛋白功能具有重要作用<sup>[2]</sup>。系统发生分析显示,部分大豆 MYB 转录因子与拟南芥次生代谢相关的 MYB 转录因子进化关系较近,而部分大豆 MYB 蛋白与拟南芥发育相关 MYB 蛋白聚为一类,同时另一部分大豆 MYB 蛋白序列与 *AtMYB2*、*AtMYB102* 等拟南芥响应胁迫的相关蛋白聚为一类。该系统发生分析的结果表明了 MYB 转录因子功能的多样化。盐碱胁迫下的野生大豆中 MYB 基因的表达模式聚类分析显示,大多数 MYB 基因在盐碱胁迫下都明显上调表达,也有少部

分呈现下调表达的趋势,说明这些 MYB 基因参与了大豆对盐碱胁迫的响应。本研究对大豆 MYB 类转录因子进行了初步的鉴定及生物信息学分析,进一步分析显示 MYB 类转录因子在功能上的多样化,至于更细致的功能分析,需要后续实验进行针对性研究。

参考文献

[1] Dubos C, Stracke R, Grotewold E, et al. MYB transcription factors in *Arabidopsis* [J]. Trends in Plant Science, 2010, 15: 573-581.

[2] 刘蕾, 杜海, 唐晓凤, 等. MYB 转录因子在植物抗逆胁迫中的作用及其分子机理[J]. 遗传, 2008(10):1265-1271. (Liu L, Du H, Tang X F, et al. The roles of MYB transcription factors on plant defense responses and its molecular mechanism[J]. Hereditas, 2008(10):1265-1271. )

[3] 杜海, 杨文杰, 刘蕾, 等. 大豆 MYB 转录因子基因 GmMYBJ6 和 GmMYBJ7 的克隆及表达分析[J]. 作物学报, 2008, 34(7): 1179-1187. (Du H, Yang W J, Liu L, et al. Cloning and functional identification of the two MYB transcription factors GmMYBJ6 and GmMYBJ7 in soybean[J]. Acta Agronomica Sinica, 2008, 34(7):1179-1187. )

[4] 陈清, 汤浩茹, 董晓莉, 等. 植物 MYB 转录因子的研究进展[J]. 基因组学与应用生物学, 2009(2):365-372. (Chen Q, Tang H R, Dong X L, et al. Progress in the study of plant MYB transcription factors[J]. Genomics and Applied Biology, 2009(2):365-372. )

[5] 李科友, 朱海兰. 植物非生物逆境胁迫 DREB/CBF 转录因子的研究进展[J]. 林业科学, 2011(1):124-134. (Li K Y, Zhu H L. Research progress of DREB/CBF transcription factor in response to abiotic-stresses in plants[J]. Scientia Silvae Sinicae, 2011(1):124-134. )

[6] Riechmann J L, Heard J, Martin G, et al. Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes [J]. Science, 2000, 290:2105-2110.

[7] Goodstein D M, Shu S, Howson R, et al. Phytozome: A comparative platform for green plant genomics [J]. Nucleic Acids Research, 2012, 40:1178-1186.

[8] Finn R D, Bateman A, Clements J, et al. Pfam: The protein families database [J]. Nucleic Acids Research, 2014, 42:222-230.

[9] Du H, Zhang L, Liu L, et al. Biochemical and molecular characterization of plant MYB transcription factor family, biochemistry [J]. Biokhimiia, 2009, 74:1-11.

[10] 孙晶, 盛碧涵, 韩英鹏, 等. 大豆基因组 WRKY31 同源基因的生物信息学分析[J]. 大豆科学, 2014, 33(5):642-647. (Sun J, Sheng B H, Han Y P, et al. Bioinformatics analysis of WRKY31 homologs in soybean genome [J]. Soybean Science, 2014, 33(5):642-647. )

[11] Schultz J, Milpetz F, Bork P, et al. SMART, a simple modular architecture research tool: Identification of signaling domains[J]. Proceedings of the National Academy of Sciences of the United States of America, 1998, 95:5857-5864.

[12] Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in *Arabidopsis thaliana* [J]. Current Opinion in Plant Biology, 2001, 4:447-456.

[13] Crooks G E, Hon G, Chandonia J M, et al. WebLogo: A sequence logo generator [J]. Genome Research, 2004, 14:1188-1190.

[14] Kranz H D, Denekamp M, Greco R, et al. Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana* [J]. The Plant Journal, 1998, 16(2): 263-276.

[15] 曹忠慧. 苹果 MYB 转录因子家族的生物信息学分析及 MdMYB121 和 MdSIMYB1 在非生物胁迫响应中的作用[D]. 泰安: 山东农业大学, 2013. (Cao Z H. Genome-wide analysis of the apple MYB transcription factor allows the identification of MdMYB121 and MdSIMYB1 gene conferring abiotic stress tolerance [D]. Taian: Shandong Agricultural University, 2013. )

[16] 江董丽, 才华, 端木慧子, 等. 大豆 GST 基因家族全基因组筛选、分类和表达[J]. 分子植物育种, 2013(5):465-475. (Jiang D L, Cai H, DuanMu H Z, et al. Genome-wide filter, classification and expression analysis of GST gene family in soybean [J]. Molecular Plant Breeding, 2013(5):465-475. )

[17] Denekamp M, Smeekens S C. Integration of wounding and osmotic stress signals determines the expression of the AtMYB102 transcription factor gene [J]. Plant Physiology, 2003, 132: 1415-1423.