

## *GmAGL15* 基因启动子的克隆及序列分析

郭伟,汪潇淋,王慧,喻德跃

(南京农业大学国家大豆改良中心,作物遗传与种质创新国家重点实验室,江苏 南京 210095)

**摘要:**为探明大豆 *GmAGL15* 基因的表达调控规律,应用电子克隆技术从大豆基因组中克隆了 *GmAGL15* 1000 bp 的启动子序列,用 PLACE 在线启动子预测工具分析表明:该序列含有启动子的一般结构,如 TATA-BOX、CAAT-BOX。另外还含有一些顺式作用元件,如调控 *GmAGL15* 的组织特异的和特定发育阶段的表达,对胁迫的应答,对光的响应,以及反馈调节,推测大豆 *GmAGL15* 基因表达有相应组织特异性,可能受蔗糖、生长素和乙烯等的调控。用 FootPrinter 和 PLACE 在线工具对大豆与拟南芥等其他 4 种植物的同源基因启动子的顺式元件进行比较,发现不同植物的启动子既有保守性,又有多多样性,转录因子结合位点的分布相似,但也有区别,暗示了 *GmAGL15* 基因表达调控的精确性或多样性。

**关键词:**大豆; *GmAGL15*; 启动子; 电子克隆; 序列分析

**中图分类号:**S565.1 **文献标识码:**A **文章编号:**1000-9841(2008)05-0727-05

## Cloning and Sequence Analysis of *GmAGL15* Gene Promoter from Soybean

GUO Wei, WANG Xiao-lin, WANG Hui, YU De-yue

(National Center for Soybean Improvement, National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China)

**Abstract:** In order to study *GmAGL15* gene expression and regulation in soybean, its 1000 bp promoter sequence was cloned from the soybean genome in silico. Promoter sequence analyzed by PLACE showed that it had TATA-box, CAAT-box and some cis-acting elements which could regulate specific expression in different organs and response to stresses, light, and self-feedback. In addition, *GmAGL15* may be regulated by sucrose, auxin and ethylene. Comparaison of the promoter of *GmAGL15* with homologues from other plants by FootPrinter and PLACE suggested that these promoters had conserved and diverse domains, while the distribution of transcription factor-binding sites had similarity and difference. These findings implied the accuracy and diversity of *GmAGL15* gene expression and regulation.

**Key words:** Soybean (*Glycine max* L, Merr); *GmAGL15*; Promoter; In silico cloning; Sequence analysis

MADS-box 基因是一大类广泛存在于真菌、植物和动物中的转录因子,它们都具有一段保守的 MADS 结构域,通过其保守结构域与特定的 DNA 序列结合来调控基因的表达<sup>[1]</sup>。目前在拟南芥和水稻中已经发现 MADS-box 基因家族成员各有 107 和 71 种<sup>[2-3]</sup>。MADS-box 基因参与了植物的花和果实的发育、根的形成和胚的形态建成等过程<sup>[4-7]</sup>。

研究表明, *AGL15* (GenBank 登录号: AAA65653) 是拟南芥中一个控制胚发育的 MADS-box 基因<sup>[8]</sup>, *AGL15* 组成型表达可以促进体细胞胚的形成<sup>[9]</sup>, 它与葡萄的 *BnAGL15-2* (GenBank 登录

号: AAB03807)、矮牵牛的 *PMADS9* (GenBank 登录号: AAQ72497)、白杨的 *PMADS40* 同源性较高<sup>[10-12]</sup>, 但对其功能则了解不多。

*GmAGL15* (GenBank 登录号: AAR17483) 是最近被克隆的 MADS-box 基因, 它是 *AGL15* 在大豆中的同源基因。研究表明 *GmAGL15* 在大豆种子发育的不同时期均有不同程度的表达, 其中在开花后 15 d 的幼胚中表达量较高, 而在开花后 30 d 的幼胚中表达量最高, *GmAGL15* 转入大豆的研究表明它与 *AGL15* 有类似的作用<sup>[14]</sup>。而有关大豆 *GmAGL15* 基因启动子的研究尚未见报道。

收稿日期: 2008-06-25

基金项目: 国家自然科学基金资助项目 (30771362); 国家高技术研究发展计划 (863 计划) 资助项目 (2006AA10Z1C1); 教育部高等学校引智计划 (111 计划) 资助项目 (B08025)。

作者简介: 郭伟 (1981-), 男, 硕士研究生, 研究方向为生物化学及分子生物学。E-mail: thundergw@yahoo.com.cn。

通讯作者: 喻德跃, 教授, 博士生导师。E-mail: dyu@njau.edu.cn。

研究克隆大豆 *GmAGL15* 基因的启动子序列,对潜在的顺式元件进行分析,并和其他 4 个物种同源基因的启动子序列进行比较分析,为 *GmAGL15* 基因应用于大豆种子的品质改良提供依据。

## 1 材料与方法

### 1.1 序列来源

大豆基因组序列来自 Phytozome (<http://www.phytozome.net/>),白杨序列来自 JGI Genome Portal Site ([http://genome.jgi-psf.org/Poptr1\\_1/](http://genome.jgi-psf.org/Poptr1_1/)),拟南芥基因组序列来自 TAIR (<http://www.arabidopsis.org>)。

### 1.2 启动子的获得及分析

利用已报道的白杨 *PMADS40* 在白杨基因组中搜索获得 *PMADS40* 对应片段。利用已登录的大豆 *GmAGL15* 核苷酸序列,在已公布的大豆基因组中搜寻其对应区域。利用已登录的拟南芥 *AGL15* 核苷酸序列,在已公布的拟南芥基因组中搜寻其对应区域。

在拟南芥、白杨和大豆基因组中分别查找 *AGL15*、*PMADS40*、*GmAGL15* 起始密码子上游 1000 bp 的序列,这部分序列作为启动子序列。在 GenBank 数据库中查找 *BnAGL15-2*、*PMADS9* 基因及其启动子,将起始密码子上游 1000 bp 和 500 bp 的序列作为启动子序列。

应用在线植物转录元件分析工具 PLACE (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>) 对 *GmAGL15* 及其同源基因启动子序列进行分析<sup>[15]</sup>,寻找转录因子结合位点(transcription factor-binding sites,TFBS)。根据上面的分析结果,对不同物种 *GmAGL15* 同源基因启动子的 TFBS 进行统计,寻找物种间保守的以及 *GmAGL15* 特异的 TFBS。应用 FootPrinter<sup>[16]</sup> 对不同物种的 *GmAGL15* 同源基因启动子序列进行比较,结合 PLACE 分析结果预测获得的保守元件的作用。

## 2 结果与分析

### 2.1 大豆 *GmAGL15* 的启动子的获得和顺式元件分析

利用已登录的 *GmAGL15* 核苷酸序列,在已公布的大豆基因组 Phytozome 中搜寻其对应区域,发现在 scaffold\_80 上有一段与之最为匹配(图 1)。在

大豆基因组 Phytozome 中搜索 *GmAGL15* 起始密码子上游 1000 bp 的序列,这部分序列作为启动子序列。

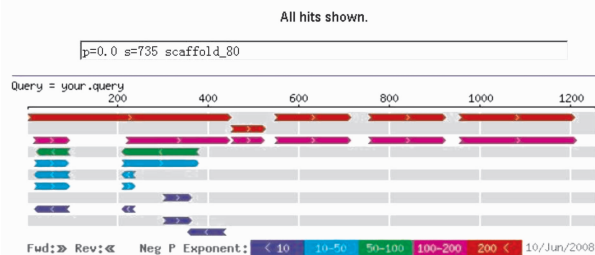


图 1 在 Phytozome 数据库检索对 *GmAGL15* 比对结果

Fig. 1 The result of blast hits on *GmAGL15* in the Phytozome

启动子区域经 PLACE 在线分析,预测的转录因子结合位点见表 1。在翻译起始密码子上游-360 处包含一个 TATA-box,即 RNA 聚合酶结合位点,保证转录的精确起始;上游还包括多个 CAAT-box,主要控制着转录起始的频率。此外在序列中发现多个潜在的顺式作用元件,它们控制 *GmAGL15* 的组织器官特异表达(CACTFTPPCA1,EBOX-BNNAPA,CATGCA,GTGANTG10,L1BOXATPDF1,NODCON2GM,TAAAGSTKST1,PROLAMINBOXOS-GLUB1),控制大豆胁迫下的应答(CURECORECR,MYB1AT,MYB2CONSENSUSAT,SURE1STPAT21,SURECOREATSULTR11,MYBCORE,MYBCOREAT-CYCB1,MYBPZM),受激素的调节(ARR1AT,ERELEE4),光调节相关(CIACADIANLELHC,EVENINGAT,GT1CONSENSUS,REALPHALGLH-CB21,TBOXATGAPB),还包括一个自身反馈调节位点 CARGCW8GAT。说明大豆 *GmAGL15* 基因的表达可能具有高度的组织器官特异性,并受胁迫、激素和光等的调节,以及受自身的反馈调控。

### 2.2 不同物种 *GmAGL15* 同源基因启动子序列的比较

应用 FootPrinter 在线分析工具对克隆的大豆 *GmAGL15* 启动子序列与其它 4 种植物同源基因启动子序列进行比较,结果如图 2 所示。从图 2 中可以看出,不同物种 *GmAGL15* 同源基因的启动子区域中存在一系列保守元件,包括 TATA-BOX,CAAT-box、光响应元件、种子特异表达元件,E2F 结合位点,乙烯应答元件。

表 1 启动子区顺式作用元件分析

Table 1 Cis-acting regulatory elements analysis of promoter sequences

应用 PLACE 在线预测 TFBS TFBS analyzed by PLACE			由 FootPrinter 分析获得的保 守 TFBS Conserved TFBS analyzed by FootPrinter	统计的保守 TFBS Conserved TFBS by statistic anal- ysis	大豆特有 的 TFBS Characteris- tic TFBS of soybean
位点名称 Site name	信号序列 Signal sequence	位点功能 Function of site			
ACGTATERD1	ACGT	黄化诱导表达的元件序列			
AMYBOX2	TATCCAT	淀粉酶元件			√
ARR1AT	NGATT	细胞分裂素调节基因 ARR1 结合位点			
BOXIINTPATPB	ATAGAA	发现于烟草质体 atpB 基因启动子		√	
CAAT - box	CCAAT	启动子、增强子区域普通顺式作用元件	√	√	
CACTFTPPCA1	YACT	涉及玉米叶肉特异表达的顺式元件			
CARGCW8GAT	CWWWWWWWWG	AGL15 结合位点		√	
CGCGBOXAT	VCGCGB	AtSR1 -6 识别元件			
CIACADIANLELHC	CAANNNNATC	涉及基因昼夜节律的区域			
CURECORECR	GTAC	铜应答元件			
DOFCOREZM	AAAG	Dof 蛋白结合位点		√	
E2FCONSENSUS	WTSSCSS	E2F 结合位点	√		
EBOXBNNAPE	CANNTG	储藏蛋白 napA 的 E - box		√	
EECCRCAH1	GANTTNC	Myb 转录因子 LCR1 结合位点			
ERELEE4	AWTTCAAA	乙烯应答元件	√		
EVENINGAT	AAAATATCT	周期蛋白启动子的晚间元件			√
GATABOX	GATA	ASF -2 结合位点			
GT1CONSENSUS	GRWAAW	光调节基因结合位点	√	√	
GT1CORE	GGTTAA	对 GT - 1 结合起关键作用			
GTGANTG10	GTGA	花粉晚期基因启动子的元件			
INRNTPSADB	YTCANTYY	Inr 元件		√	
LIBOXATPDF1	TAAATGYA	涉及基因茎端分生组织 L1 层特异表达			√
LECPLEACS2	TAAAATAT	LeCp 结合的顺式元件			
MYB1AT	WAACCA	与干旱诱导有关的 MYB 结合位点			
MYB2CONSENSUSAT	YAACKG	与脱水响应相关的 MYB 识别位点			
MYBCORE	CNGTTR	MYB 蛋白结合位点			
MYBCOREATCYCB1	AACGG	Myb 核心 xulie			
MYBPZM	CCWACC	与干旱诱导有关的 MYB 结合位点			
MYBST1	GGATA	MybSt1 结合位点			
MYCCONSENSUSAT	CANNTG	MYC 识别位点或 ICE1 结合位点		√	
NODCON2GM	CTCTT	涉及基因在根瘤的表达		√	
NTBBF1ARROLB	ACTTTA	Dof 蛋白结合位点			
OSE2ROOTNODULE	CTCTT	感染的根瘤细胞激活的启动子顺式元件		√	
POLLEN1LELAT52	AGAAA	花粉特异表达的顺式作用元件		√	
PROLAMINBOXOSGLUB1	TGCAAAG	内胚乳特异表达必须的顺式元件			
PYRIMIDINEBOXOSRAMY1A	CCTTTT	BPBF 蛋白结合位点			
RAV1AAT	CAACA	RAV1 蛋白结合位点		√	
REALPHALGLHCB21	AACCAA	光敏色素调节元件			
ROOTMOTIFTAPOX1	ATATT	rolD 的启动子元件		√	
RYREPEATBNNAPA	CATGCA	涉及基因在种子特异表达	√	√	
SEF1MOTIF	ATATTTAAWW	SEF1 结合元件			√
SEF4MOTIFGM7S	RTTTTTR	SEF4 结合位点			
SURE1STPAT21	AATAGAAAA	蔗糖应答元件			√
SURECOREATSULTR11	GAGAC	硫应答元件			
TAAAGSTKST1	TAAAG	涉及基因在保卫细胞特异表达		√	
TATA - box	TATA	转录起始 - 30 核心启动子元件	√	√	
TATCCAOSAMY	TATCCA	MYB 蛋白结合位点			
TBOXATGAPB	ACTTTG	涉及基因转录的光诱导激活			
WRKY71OS	TGAC	W - box 的核心区域		√	



个物种的启动子序列上都出现了,并且出现的位置也基本位于翻译起始密码子上游-250 处左右,PLACE 预测表明它调控了基因在种子特异的表达,*GmAGL15* 同源基因在种子的发育中可能都起了重要作用。在大豆和拟南芥、葡萄中的实验发现 *GmAGL15* 与其同源基因 *AGL15*、*BnAGL15-2* 在种子中表达量较高<sup>[14,19]</sup>,推测其他物种 *GmAGL15* 同源基因也可能在种子中特异表达。

花粉特异表达的顺式作用元件 POLLEN1LELAT52 元件在大豆启动子序列上重复出现了 5 次,并且在对不同物种的 *GmAGL15* 同源基因启动子序列分析的时候发现 POLLEN1LELAT52 都存在并且重复出现多次。拟南芥中的实验表明 *AGL15* 在花粉中有表达<sup>[20]</sup>,暗示了 *GmAGL15* 在花粉中表达的可能以及与对花粉形成的重要作用。

顺式作用元件的预测为最终了解 *GmAGL15* 基因的表达调控提供了依据。在后续的工作中,可以根据 *GmAGL15* 的启动子构建多个表达载体(包括对 *GmAGL15* 启动子区的关键顺式元件进行点突变),转化大豆,在不同条件下检测报告基因的表达,以对启动子分析的结果进行验证。

利用 FOOTPRINTER 和公共的 TFBS 数据库如 PLACE 对启动子的顺式元件进行分析时,发现这种方法确实可以发现已知的和新的元件。但是,公共的 TFBS 数据库还很不完全,导致分析的时候可能存在潜在的有活性的顺式元件被忽略了;此外,对于这些元件是否代表有活性的 TFBS,转录因子是否能够结合在它们上以实现基因的调节作用,这些都是以后需要进一步研究的内容。

## 参考文献

- [1] Riechmann J L, Meyerowitz E M. MADS domain proteins in plant development [J]. *Biological Chemistry*, 1997, 378 (10): 1079-1101.
- [2] Parenicova L, Folter S de, Kieffer M, et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis; New openings to the MADS world [J]. *Plant Cell*, 2003, 15(7): 1538-1551.
- [3] Nam J, Kim J, Lee S, et al. Type I MADS-box genes have experienced faster birth- and- death evolution than type II MADS-box genes in angiosperms [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101 (7): 1910-1915.
- [4] Causier B, Kieffer M, Davies B. Plant biology: MADS- Box genes reach maturity [J]. *Science*, 2002, 296 (5566): 275-276.
- [5] Irish V F. The evolution of floral homeotic gene function [J]. *Bioessays*, 2003, 25(7): 637-646.
- [6] Zik M, Irish V F. Flower development: Initiation, differentiation, and diversification [J]. *Annual Reviews in Cell and Developmental Biology*, 2003, 19: 119-140.
- [7] Perry S E, Lehti M D, Fernandez D E. The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins [J]. *Plant Physiology*, 1999, 120(1): 121-129.
- [8] Zhu C, Perry S E. Control of expression and autoregulation of AGL15, a member of the MADS- box family [J]. *Plant Journal*, 2005, 41(4): 583-594.
- [9] Harding E W, Tang W, Nichols K W, et al. Expression and maintenance of embryogenic potential is enhanced through constitutive expression of AGAMOUS-like 15 [J]. *Plant Physiology*, 2003, 133(2): 653-663.
- [10] Heck G R, Perry S E, Nichols K W, et al. AGL15, a MADS domain protein expressed in developing embryos [J]. *Plant Cell*, 1995, 7(8): 1271-1282.
- [11] Vandenbussche M, Zethof J, Souer E, et al. Toward the analysis of the petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require SEPALLATA-like MADS box genes in Petunia [J]. *The Plant Cell*, 2003, 15(11): 2680-2693.
- [12] Leseberg C H, Li A L, Kang H, et al. Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa* [J]. *Gene*, 2006, 378: 84-94.
- [13] Thakare D, Tang W, Hill K, et al. The MADS-domain transcriptional regulator AGAMOUS-like 15 promotes somatic embryo development in arabidopsis and soybean [J]. *Plant Physiology*, 2008, 146(4): 1663-1672.
- [14] 汪潇琳, 陈艳萍, 喻德跃. MADS-box 基因 *GmAGL15* 在大豆种子发育过程中的表达 [J]. *作物学报*, 2008, 34(2): 330-332. (Wang X L, Chen Y P, Yu D Y. Expression of the MADS-box gene *GmAGL15* in seed development of soybean [J]. *Acta Agronomica Sinica*, 2008, 34(2): 330-332.)
- [15] Higo K, Ugawa Y, Iwamoto M, et al. Plant cis- acting regulatory DNA elements (PLACE) database [J]. *Nucleic Acids Research*, 1999, 27(1): 297-300.
- [16] Blanchette M, Tompa M. FootPrinter; a program designed for phylogenetic footprinting [J]. *Nucleic Acids Research*, 2003, 31(13): 3840-3842.
- [17] Körbes A P, Droste A. Carbon sources and polyethylene glycol on soybean somatic embryo conversion [J]. *Pesquisa Agropecuária Brasileira*, 2005, 40(3): 211-216.
- [18] Adamczyk B J, Lehti-Shiu M D, Fernandez D E. The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in Arabidopsis [J]. *Plant Journal*, 2007, 50(6): 1007-1019.
- [19] Rounsley S D, Ditta G S, Yanofsky M F. Diverse roles for MADS box genes in arabidopsis development [J]. *Plant Cell*, 1995, 7(8): 1259-1269.
- [20] Alvarez-Buylla E R, Liljegren S J, Pelaz S, et al. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes [J]. *Plant Journal*, 2000, 24(4): 457-466.