



大豆 TOE 基因的进化分析及开花调控功能解析

刘俊, 李海洋, 贺米兰, 张婷, 刘宝辉, 赵晓晖

(广州大学 生命科学学院, 广东 广州 510006)

摘要: 为了进一步解析大豆中重要植物开花和花器官发育调控转录因子 AP2 的编码基因 TOE 的进化规律及其对开花功能的调控作用, 为大豆 TOE 基因的功能解析和大豆纬度适应性研究提供基础, 本研究利用生物信息学手段对大豆 TOE 基因进行聚类分析、序列特征分析、染色体区段共线性分析和组织特异性表达分析, 预测关键开花基因启动子区段 AP2 结合位点, 并验证不同单倍型大豆开花时间。结果显示: 从 PlantTFDB 数据库检索到 12 个大豆 TOE 基因, *GmTOE6b* (Glyma.02G087400) 为新发现的大豆 TOE 基因。*GmTOE6a* 和 *GmTOE6b* 均只有 1 个 AP2 结构域, 其余大豆 TOE 基因均有两个 AP2 结构域。6 个大豆 TOE 基因与拟南芥 *TOE1* 基因聚为一类; 2 个与拟南芥 *TOE2* 基因聚为一类; 4 个与拟南芥 *TOE3*、*AP2* 聚为一类。12 个大豆 TOE 基因都有且仅有 1 个 *miR172* 靶位点, 且该靶位点序列与拟南芥 TOE 的 *miR172* 靶位点序列高度一致。染色体区段共线性分析显示, 大豆 12 个 TOE 基因按起源方式可以分为 3 类, 6 个随大豆基因组复制而产生; 4 个起源于大豆物种形成之前且与拟南芥 TOE 基因有共同祖先; 2 个起源于大豆物种形成之前且与拟南芥 TOE 基因无共同祖先。大豆开花关键基因 *GmFT2a* 和 *GmFT5a* 启动子序列中均含有多个 AP2 结合位点, *GmTOE4b* 和 *GmTOE5b* 两个基因均可影响大豆的开花时间。研究结果说明大豆中 12 个 TOE 基因极有可能均是 *miR172* 的靶基因, 虽然其编码蛋白的氨基酸组成非常相似, 但它们的进化规律和组织特异性表达规律存在不同, 它们在进化过程中可能存在功能分化。*GmTOE4b* 和 *GmTOE5b* 可能通过结合 *GmFT2a* 和 *GmFT5a* 启动子上的顺式元件来调控其基因转录, 从而调控开花。

关键词: 大豆; 开花调控; TOE 基因; 生物信息学分析; *GmTOE6b*; 共线性分析; 单倍型

Molecular Evolutionary and Flowering Regulation Function Analysis of TOE Genes in Soybean

LIU Jun, LI Hai-yang, HE Mi-lan, ZHANG Ting, LIU Bao-hui, ZHAO Xiao-hui

(School of Life Sciences, Guangzhou University, Guangzhou 510006, China)

Abstract: In order to analyze the evolution of *GmTOEs* and its regulation on flowering function in soybean, to provide the basis for functional analysis of *GmTOEs* and research on latitude adaptability of soybean, we used bioinformatics methods to perform cluster analysis, sequence characterization analysis, chromosomal segment collinearity analysis and tissue specific expression analysis of *GmTOEs*, predicted AP2 binding sites in the promoter regions of important flowering genes and verified flowering times in different haplotypes of soybean. The results showed that 12 *GmTOEs* were retrieved from PlantTFDB database, and *GmTOE6b* (Glyma.02G087400) was a newly discovered TOE gene in soybean. *GmTOE6a* and *GmTOE6b* had only one AP2 domain, while the other *GmTOEs* had two AP2 domains. Six *GmTOEs* were clustered into the same group as the *Arabidopsis TOE1*, two were clustered with *Arabidopsis TOE2*, four were clustered with *Arabidopsis TOE3* and *AP2*. Each *GmTOEs* had one *miR172* target site, and the target site sequence was highly consistent with *AtTOEs*. Chromosome segment collinearity analysis showed that *GmTOEs* could be divided into three categories according to their origin, and six of them were generated with the replication of soybean genome. Four of them originated before the speciation of soybean and shared a common ancestor with *AtTOEs*. Two of them originated before the speciation of soybean and had no common ancestor with *AtTOEs*. The important flowering regulation genes *GmFT2a* and *GmFT5a* have multiple AP2 binding sites in their promoter, and both *GmTOE4b* and *GmTOE5b* can affect the flowering time of soybean. The results indicated that the 12 *GmTOEs* were the most likely to be the target of *miR172*. Although the amino acid composition of the protein encoded by them was very similar, their evolutionary rules and tissue-specific expression were different, and they may have functional differentiation during the evolutionary process. *GmTOE4b* and *GmTOE5b* may regulate the transcription of *GmFT2a* and *GmFT5a* by binding to their *cis* acting elements on the promoters, to regulate flowering.

Keywords: Soybean; Flowering regulation; TOE genes; Bioinformatics analysis; *GmTOE6b*; Collinearity analysis; Haplotype

大豆 [*Glycine max* (L.) Merr.] 是世界上最主要的植物油和蛋白质供给作物之一, 并且贡献了世界上超过四分之一的食用蛋白和饲用蛋白^[1]。随着国民生活水平的提升, 人们对肉蛋奶的消费快速

收稿日期: 2021-03-09

基金项目: 国家自然科学基金(31801383); 广东省自然科学基金(2018A030313199)。

第一作者: 刘俊(1995—), 男, 硕士, 主要从事大豆开花调控分子机理研究。E-mail: 2111814034@e.gzhu.edu.cn。

通讯作者: 赵晓晖(1985—), 女, 博士, 副教授, 主要从事大豆分子遗传学研究。E-mail: zhaoxiaohui_1234@163.com。

增长,大豆作为主要的动物饲料来源导致我国对大豆的需求量逐年递增,目前大豆来源高度依赖进口,主要原因是大豆单产水平未得到大幅度提升导致种植收益低,种植面积减少^[2]。因此,提高大豆单产是缓解国产大豆供应不足的关键^[3]。

大豆是典型的短日照植物,对光周期反应极其敏感,光周期是影响大豆开花期和生育期的重要因素,也是大豆单一品种只能适应较小纬度范围的主要限制因素^[4]。我国幅员辽阔,南北纬度跨度大,需要具有不同程度的光周期不敏感性的品种以适应不同地区的大豆生产需求,而大豆光周期不敏感性又是由生育期基因及其调控基因控制,因此对大豆生育期基因的研究将有利于大豆品种布局、栽培区划及分子育种^[5],进而提高大豆单产。

TOE(target of early activation tagged)基因编码 AP2 转录因子,在调控植物开花和花器官发育方面起着重要作用。通过经典遗传学研究发现 12 个大豆生育期基因,包括命名从 *E1* 到 *E11* 的 11 个基因和控制长童期性状的 *J* 基因,其中 *E6*、*E9*、*E11* 和 *J* 具有促进大豆开花的功能,而 *E1*、*E2*、*E3*、*E4*、*E7*、*E8* 和 *E10* 基因都具有延迟大豆开花和成熟的功能^[6-17]。此外,参考拟南芥开花调控基因,通过反向遗传学方法也寻找到许多大豆开花调控基因^[18-19]。在大豆中,FT 是主要的开花整合因子,是调控开花时间的重要基因^[20-22]。FT 蛋白在叶片中合成,通过韧皮部运输到大豆茎顶端诱导花分生组织的发育^[23]。大豆共有 12 个 FT 基因 *GmFT1a*、*GmFT1b*、*GmFT2a*、*GmFT2b*、*GmFT2c*、*GmFT2d*、*GmFT3a*、*GmFT3b*、*GmFT4*、*GmFT5a*、*GmFT5b* 和 *GmFT6*。其中,*GmFT2a* 和 *GmFT5a* 具有明显的组织特异性,在大豆叶片中高表达,是功能最强的关键大豆开花促进因子^[23-24]。对大豆 FT 基因的表达调控将极大地影响开花时间,从而调节生育期,提高大豆产量。

在拟南芥的开花调控途径中,FT 基因的表达受到两类 AP2 超家族基因编码的转录因子的抑制,拟南芥 AP2 转录因子通过与 FT 启动子结合来抑制 FT 转录,从而抑制开花。这两类 AP2 超家族基因分别是 RAV 家族的 2 个基因(*TEM1* 和 *TEM2*)和 AP2 亚家族的 6 个基因(*AP2*、*TOE1*、*TOE2*、*TOE3*、*SMZ* 和 *SNZ*)^[25]。AP2 亚家族的这 6 个基因同时也是 *miR172* 的靶基因^[26-27],在拟南芥的生长发育过程中,随着 *miR172* 表达量逐渐升高其靶基因的表达量逐渐降低,当低于临界值时,对拟南芥开花的抑制解除,从而导致拟南芥开花^[27]。*miR172* 通过抑制 AP2 类转录因子的表达诱导开花的调控模式在水稻、拟南芥、马铃薯等多种植物中都有报道^[28-30]。

大豆 *miR172* 的靶基因 *GmTOE4a*(Glyma. 15G044400) 和 *GmTOE1a* (Glyma. 03G17750) 也 通 过 抑 制 *GmFT2a* 和 *GmFT5a* 的表达延迟开花,这两个基因均属于 AP2 亚家族^[31-32],且与拟南芥抑制 FT 基因表达的 6 个 AP2 亚家族基因同源。在大豆中过表达 *GmTOE4a* 基因会表现晚花表型,并且 *GmFT2a* 和 *GmFT5a* 的转录水平下调^[31],但 *GmTOE4a* 具体如何调控 *GmFT2a* 和 *GmFT5a* 的转录并未见报道。因此,大豆 TOE 基因更加深入的研究对大豆开花调控基因网络的研究具有巨大的促进作用。

前期研究以拟南芥 *AtTOE1* 为参考序列在 Phytozome 中比对到 11 个大豆 TOE 序列,其中 6 个与 *AtTOE1* 聚为一类^[31],本研究在 PlantTFDB 数据库寻找大豆中与拟南芥调控 FT 基因表达的全部 AP2 亚家族基因 (*AP2TOE1*、*TOE2*、*TOE3*、*AP2*、*SMZ*、*SNZ*) 同源的基因,对其序列进行比对分析,并比对拟南芥和大豆的 *miR172* 序列及其靶点。由于大豆经历过两次基因组扩增事件,第一次是在菜豆族分化前,第二次发生在大豆属分化时^[33],因此大部分基因有多个拷贝^[18]。为了理清大豆各 TOE 间的进化关系,本研究将大豆基因组与拟南芥基因组以及大豆基因组自身进行染色体区段共线性分析。由于 *GmFT2a* 和 *GmFT5a* 在大豆叶片中高表达,因此调控其表达的基因也极有可能在叶片中表达。本研究将 *GmFT2a* 和 *GmFT5a* 的启动子区段输入 PlantTFDB 数据库中,预测其 AP2 结合位点,以在叶片中高表达的 TOE 基因为目标基因,从已发表的 424 个大豆自然群体材料的重测序数据中^[34]选取目标基因的单倍型基因个体,统计主要单倍型的开花期表型,验证目标基因是否具有调控开花的功能。本研究有助于进一步明确 AP2 转录因子的功能,为大豆开花调控机制研究提供新方向,为大豆生育期和产量育种和引种提供新的基因资源和遗传材料。

1 材料与方法

1.1 大豆 TOE 基因的鉴定及聚类分析

从 PlantTFDB 数据库 (<http://plantfdb.gao-lab.org/>) 中获取大豆和拟南芥所有 AP2 家族基因的序列号,再从 Phytozome 数据库 (<https://phytozome.jgi.doe.gov/>) 中下载所有基因最长编码的氨基酸序列。将氨基酸序列导入 MEGA-X 软件中,用 MUSCLE 进行序列比对,再用 MODELS 预测最优模型,使用最大似然法基于最优模型构建进化树,Bootstrap 值设置为 1 000。研究与拟南芥 TOE 基因聚在同一分支的大豆基因。

1.2 大豆 TOE 基因的序列特征分析

利用在线工具 Expassy (<https://web.expasy.org>) 计算大豆各 TOE 蛋白的分子量与等电点。将大豆所有 TOE 与拟南芥 *AP2*、*TOE1*、*TOE2*、*TOE3*、*SMZ* 和 *SNZ* 编码的氨基酸序列输入 NCBI (<https://www.ncbi.nlm.nih.gov/>) 中,分析比较其保守结构域,并用 TBtools 软件进行可视化^[35]。利用 DNAMAN 比对上述各基因的氨基酸序列,比较其 *AP2* 结构域序列的相似性。

1.3 拟南芥和大豆 *miR172* 序列比对及靶位点搜索

在 psRNATarget 数据库 (<http://plantgrn.noble.org>) 中搜索大豆所有 TOE 的 miRNA 靶位点及对应的 miRNA,并用 DNAMAN 比对其序列相似性。

1.4 TOE 基因的染色体区段共线性分析

在 Phytozome 数据库下载拟南芥和大豆的基因组序列及注释信息,其中拟南芥基因组序列版本为 Athaliana_167_TAIR9,注释信息版本为 Athaliana_167_TAIR10,大豆基因组序列版本为 Gmax_275_v2.0,注释信息版本为 Gmax_275_Wm82.a2.v1.gene。用 TBtools 软件分析拟南芥和大豆染色体及大豆染色体间的共线性关系,并突出显示物种内和物种间 TOE 基因的共线性关系。

1.5 大豆 TOE 基因的组织特异性表达分析

在 Phytozome 数据库中获取大豆 TOE 基因在根、茎、叶、顶端分生组织、花、荚和种子中的表达量数据,利用 TBtools 软件绘制热图并按照基因表达的组织特异性差异进行聚类。

1.6 *GmFT2a* 和 *GmFT5a* 启动子区 *AP2* 结合位点

将 *GmFT2a*、*GmFT5a* 起始密码子前2 000 bp 作为启动子区,在 PlantTFDB 数据库中预测其转录因子结合位点,将其中 *AP2* 结合位点的位置输入

TBtools 软件进行可视化分析。

1.7 大豆 TOE 基因调控开花的功能验证

将由 85 个野生豆品种,153 个农家品种和 186 个育成品种组成的 424 份大豆自然群体材料于 2018 年在合肥 (32°N) 和郑州 (34°N) 种植,每份种植 30 株,取这 30 株大豆开花期的平均值作为该品种的开花期(R1)^[34]。这 424 份大豆自然群体材料的重测序数据已发表^[34]。利用 Vcftools 在重测序数据中提取大豆 TOE 基因的变异信息和对应的品种编号,再用 Excel 2019 提取非同义变异类型及对应的品种花期,最后利用 GraphPad 8.0 统计各单倍型花期并分析作图。

2 结果与分析

2.1 大豆 TOE 基因的鉴定及聚类分析

在 PlantTFDB 数据库中共有拟南芥 *AP2* 蛋白 30 个,由 18 个基因编码;大豆 *AP2* 蛋白 99 个,由 50 个基因编码。这 68 个 *AP2* 基因各自编码的最长氨基酸序列的进化树如图 1 所示,拟南芥 *TOE1*、*TOE2*、*TOE3*、*AP2*、*SMZ*、*SNZ* 与大豆的 12 个基因聚在同一分支(图 1)。这 12 个大豆基因大致都成对聚为 1 个分支,除 11 个已被报导为大豆 TOE 基因^[31] 外,存在 1 个新的大豆 TOE 基因 (*Glyma.02G087400*),由于其与之前命名的 *GmTOE6* 聚类在同一分支,所以将原来的 *GmTOE6* 重新命名为 *GmTOE6a*,新发现的 TOE 基因命名为 *GmTOE6b* (图 1)。其中 *GmTOE1a*、*GmTOE1b*、*GmTOE4a*、*GmTOE4b*、*GmTOE5a* 和 *GmTOE5b* 这 6 个基因与拟南芥 *TOE1* 基因聚为一类,*GmTOE6a* 和 *GmTOE6b* 与拟南芥 *TOE2* 基因聚为一类,*GmTOE2a*、*GmTOE2b*、*GmTOE3a* 和 *GmTOE3b* 4 这个基因与拟南芥 *TOE3*、*AP2* 聚为一类。

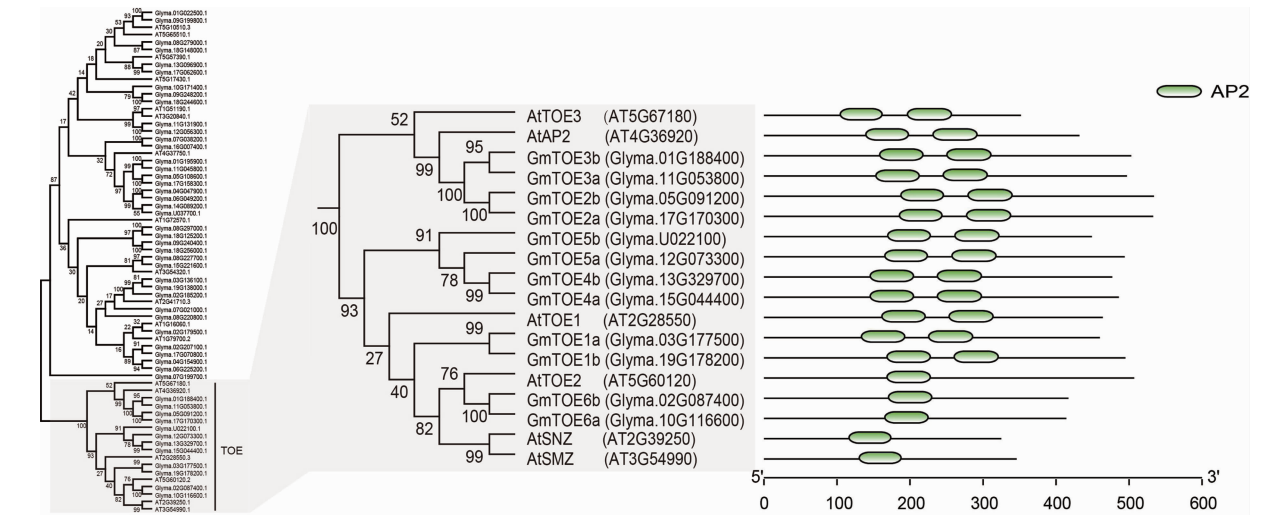


图 1 大豆与拟南芥 TOE 蛋白序列聚类及结构域分布

Fig. 1 The phylogenetic tree and domain distribution of TOE homologs protein between soybean and Arabidopsis

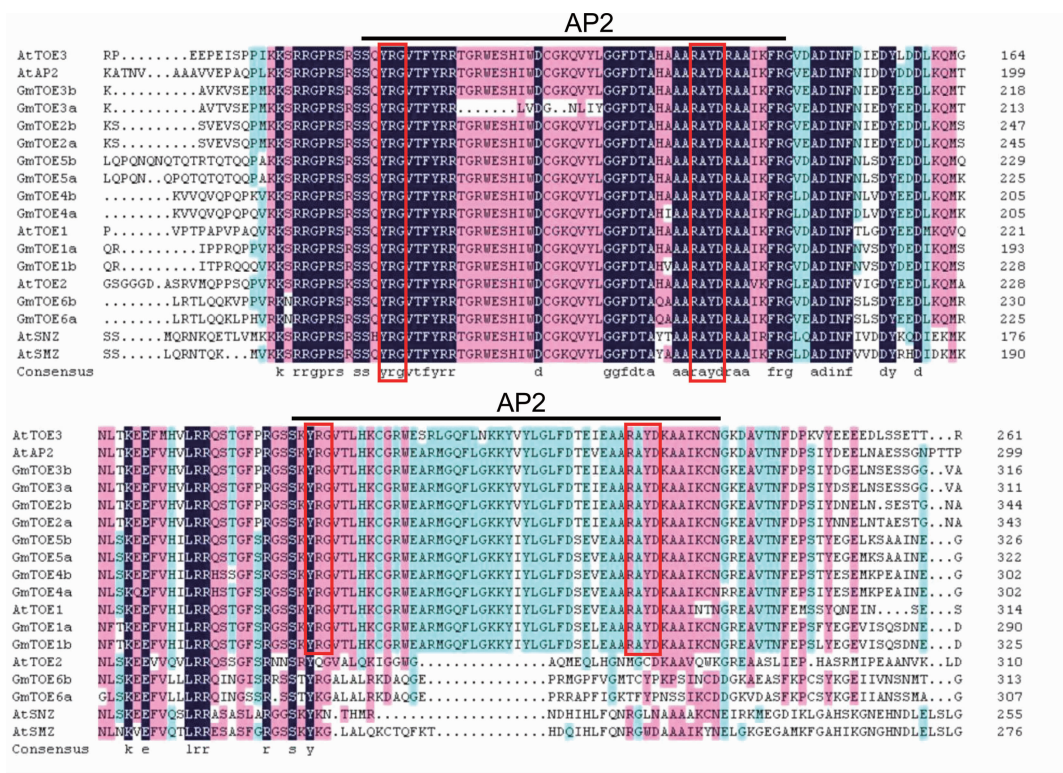
2.2 大豆 TOE 基因序列特征分析

大豆 12 个 TOE 基因编码的氨基酸数目为 413 ~ 533 aa, AP2 结构域都位于氨基酸氨基酸序列的 20% ~ 70% 区域内, *GmTOE6a* 和 *GmTOE6b* 与拟南芥 *TOE2* 一样, 都只有 1 个 AP2 结构域, 其余基因都有 2 个 AP2 结构域(表 1、图 1、图 2)。这 12 个基因的 CDS 长度为 1 242 ~ 1 602 bp, *miRNA172* 靶位点都位于 CDS 区的后 20% 区域内(表 1)。大豆 TOE 蛋白的分子量差异较小, 为 45 539.20 ~ 59 057.72 Da。等电点为 5.89 ~ 9.11(表 1)。

表 1 大豆 TOE 基因基本信息
Table 1 The essential information of *GmTOEs*

基因名称 Gene name	基因 ID Gene ID	CDS 长度 CDS length /bp	<i>miRNA172</i> 靶位点 <i>miRNA172</i> target site /bp	氨基酸数目 Amino acid number/aa	等电点 pI	蛋白分子量 Molecular weight/Da	AP2 结构域位置 AP2 domain site /aa
<i>GmTOE1a</i>	Glyma.03G177500	1380	1229 ~ 1252	459	6.27	50271.67	124 ~ 173 216 ~ 266
<i>GmTOE1b</i>	Glyma.19G178200	1485	1334 ~ 1357	494	7.73	54330.34	159 ~ 208 251 ~ 301
<i>GmTOE2a</i>	Glyma.17G170300	1599	1502 ~ 1525	532	6.07	58862.60	176 ~ 225 268 ~ 318
<i>GmTOE2b</i>	Glyma.05G091200	1602	1505 ~ 1528	533	6.05	59057.72	178 ~ 227 270 ~ 320
<i>GmTOE3a</i>	Glyma.11G053800	1491	1397 ~ 1420	496	7.77	54211.62	153 ~ 193 236 ~ 286
<i>GmTOE3b</i>	Glyma.01G188400	1509	1415 ~ 1438	502	8.11	55637.49	149 ~ 198 241 ~ 291
<i>GmTOE4a</i>	Glyma.15G044400	1458	1232 ~ 1255	485	8.53	53752.01	136 ~ 185 228 ~ 278
<i>GmTOE4b</i>	Glyma.13G329700	1431	1250 ~ 1273	476	7.20	52685.74	136 ~ 185 228 ~ 278
<i>GmTOE5a</i>	Glyma.12G073300	1482	1307 ~ 1330	493	7.09	54101.75	156 ~ 205 248 ~ 298
<i>GmTOE5b</i>	Glyma.U022100	1347	1199 ~ 1222	448	5.89	49126.99	160 ~ 209 252 ~ 302
<i>GmTOE6a</i>	Glyma.10G116600	1242	1166 ~ 1189	413	9.11	45539.20	156 ~ 205
<i>GmTOE6b</i>	Glyma.02G087400	1251	1175 ~ 1198	416	8.51	46347.22	161 ~ 210

大豆与拟南芥 TOE 氨基酸序列的对比结果如图 2 所示, 第一个 AP2 结构域内除 *GmTOE3a* 有非保守区域的氨基酸缺失外, 其余序列相似性达到 98%, 具有第二个 AP2 结构域的 TOE 蛋白其结构域相似性也达到了 96%。结果表明 AP2 结构域在不同物种的 TOE 中都有极高的序列保守性, 说明该结构域对于 TOE 蛋白行使功能极为重要。但并非所有 TOE 蛋白都有第二个 AP2 结构域, 因此, 第一个 AP2 结构域的功能在不同物种中更加趋于保守。



红色方框内为 AP2 结构域的保守基序 YRG 和 RAYD。
The red boxes denote the completely conserved sequence motifs (YRG and RAYD motifs) in AP2 domain.

图 2 大豆与拟南芥 TOE 蛋白序列 AP2 结构域相似性比较

Fig. 2 Multiple alignment of protein sequences comparing the AP2 domains of TOE homologs between soybean and Arabidopsis

2.3 拟南芥和大豆 miR172 序列比对及靶点分析

拟南芥和大豆的 *miR172* 序列比对结果显示, 它们的序列相似性非常高, 仅有第一位和最后 3 位碱基不一致 (图 3A)。这 12 个大豆 TOE 基因都有且仅有 1 个 *miR172* 的靶位点, 且这 12 个基因的靶

位点序列与拟南芥 TOE 的 *miR172* 靶位点序列高度一致 (图 3B), 说明大豆 TOE 基因与拟南芥 TOE 基因在序列结构上极为保守, 且大豆 TOE 基因极有可能都是 *miR172* 的靶基因。

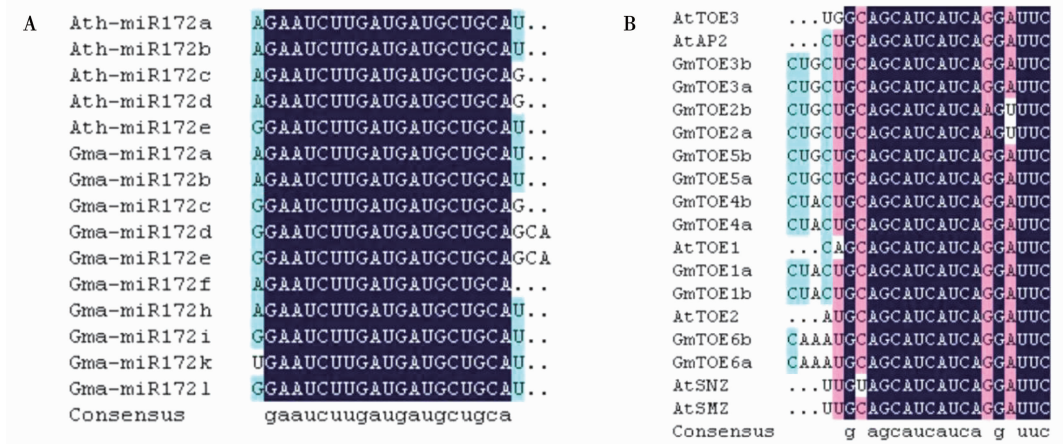


图 3 大豆与拟南芥 *miR172* (A) 及 TOE (B) 靶位点相似性比较
Fig. 3 Sequence comparison of *miR172* (A) and TOE (B) target site between soybean and Arabidopsis

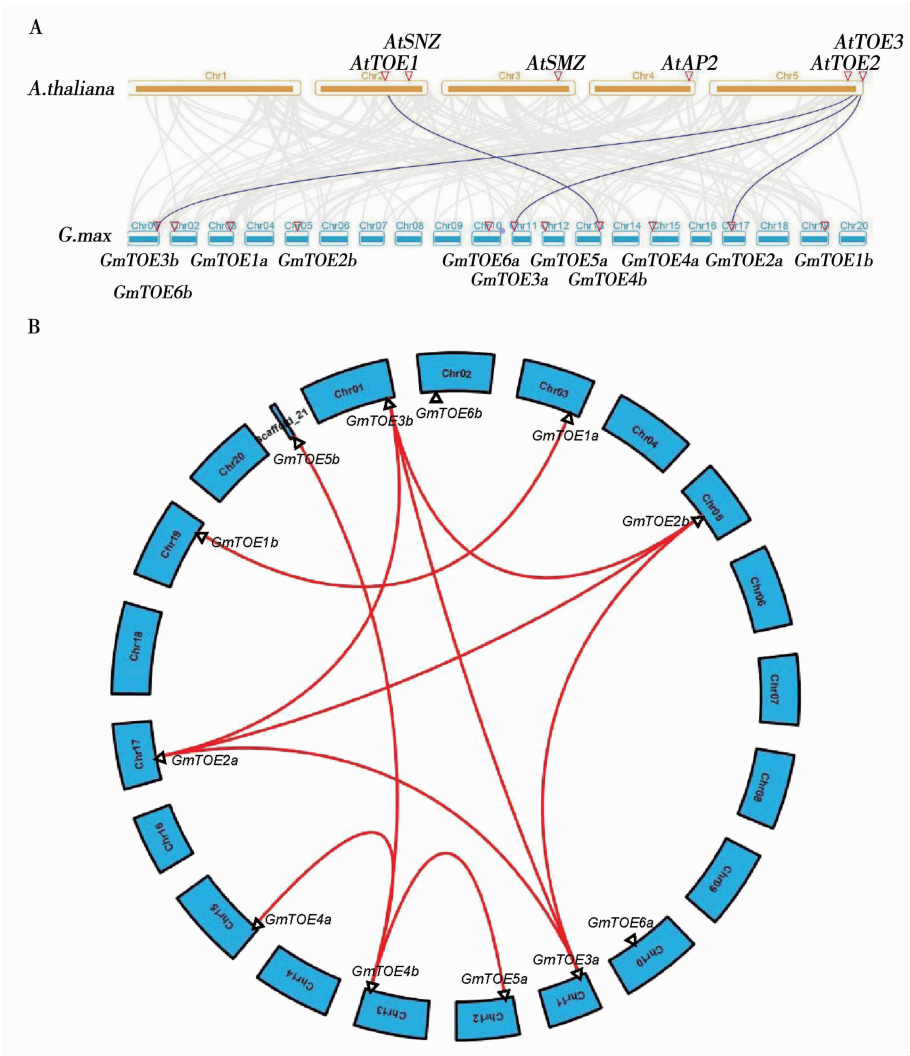
2.4 大豆 TOE 基因的染色体区段共线性分析

共线性分析结果如图 4 所示, 在 *GmTOE2a*、*GmTOE2b*、*GmTOE3a* 和 *GmTOE3b* 这 4 个基因中,

GmTOE2a、*GmTOE3a* 和 *GmTOE3b* 这 3 个基因与拟南芥 *TOE3* 具有共线性 (图 4A), 而 *GmTOE2b* 所在染色体区段与拟南芥染色体无共线性, 结合进化树

(图1)中 *GmTOE2b* 与 *GmTOE2a* 聚在同一分支分析可知, *GmTOE2b* 是在大豆基因组扩增时由 *GmTOE2a* 复制而来。 *GmTOE4a*、 *GmTOE4b*、 *GmTOE5a* 和 *GmTOE5b* 这4个基因中只有 *GmTOE4b* 与拟南芥 *TOE1* 具有共线性(图4A), 且在大豆基因组中 *GmTOE4b* 与 *GmTOE4a*、 *GmTOE5a* 和 *GmTOE5b* 这3个基因都有共线性关系, *GmTOE4a*、 *GmTOE5a* 和 *GmTOE5b* 这3个基因之间没有共线性关系(图4B), 因此 *GmTOE4a*、 *GmTOE5a* 和 *GmTOE5b* 这3个基因是由 *GmTOE4b* 在基因组扩增时复制而来的, 进化树中这4个基因并无两两配对的情况也印证了这一结果(图1)。 *GmTOE1a*、 *GmTOE1b*、 *GmTOE6a* 和 *GmTOE6b* 所在染色体区段

与拟南芥染色体并无共线性关系(图4A), 说明这4个基因的祖先基因分化时间较早。基因组内部共线性分析发现 *GmTOE1a* 与 *GmTOE1b* 具有共线性关系(图4B), 表明这两个基因是在大豆基因组复制过程中由1个基因复制而来。 *GmTOE6a* 与 *GmTOE6b* 所在染色体区段没有共线性(图4B)。因此, 大豆12个TOE基因按起源方式可以分为3类: *GmTOE1a*、 *GmTOE1b*、 *GmTOE2b*、 *GmTOE4a*、 *GmTOE5a* 和 *GmTOE5b* 随大豆基因组复制而产生; *GmTOE2a*、 *GmTOE3a*、 *GmTOE3b* 和 *GmTOE4b* 起源于大豆物种形成之前, 且与拟南芥TOE基因有共同祖先; *GmTOE6a* 和 *GmTOE6b* 起源于大豆物种形成之前, 且与拟南芥TOE基因无共同祖先。



A. 大豆与拟南芥 TOE 基因的共线性关系, 灰色线表示大豆与拟南芥基因组的共线性区块, 蓝色线表示大豆与拟南芥 TOE 基因的共线性; B. 大豆基因组各 TOE 基因间的共线性关系。

A. Collinearity analysis between soybean and *Arabidopsis*. The gray line represents the collinearity of soybean and *Arabidopsis* genome, while the blue line represents the collinearity of soybean and *Arabidopsis* TOE genes. B. Collinearity analysis of *GmTOEs* in soybean genome.

图4 大豆 TOE 基因的共线性分析
Fig. 4 Collinearity analysis of *GmTOEs*

2.5 大豆 TOE 基因的组织特异性表达分析

将 Phytozome 数据库中下载的大豆 12 个 TOE 基因的组织表达量绘制成热图并依据表达模式差异聚类,大部分成对的基因表达模式都相近,仅有 *GmTOE1b* 的表达模式与 *GmTOE1a*、*GmTOE6a* 和 *GmTOE6b* 不一致。相对而言,*GmTOE4a* 基因在所有组织中表达量均较高,而 *GmTOE6a* 整体较低。在大豆根组织中这 12 个基因相对表达量均较高,尤其 *GmTOE4a*、*GmTOE4b*、*GmTOE5a* 和 *GmTOE5b* 这 4 个基因在根中的表达量明显高于其他基因。*GmTOE1b* 在茎中的表达量远高于其他基因;叶中表达量最高的是 *GmTOE4b* 和 *GmTOE5b*;在生长点表达量最高的是 *GmTOE3b*; *GmTOE3a* 基因在花、荚和种子中的表达量在这 12 个基因中都是最高的(图 5)。大豆 12 个 TOE 基因表达模式的多样化表明它们在进化过程中可能存在功能上的分化。

2.6 *GmFT2a* 和 *GmFT5a* 启动子区 AP2 结合位点

对 *GmFT2a* 和 *GmFT5a* 启动子区段 AP2 结合位点的预测结果显示, *GmFT2a* 启动子区有 14 个 AP2 结合位点,大都集中在起始密码子前 10 ~ 50 bp 和 1 200 ~ 1 500 bp; *GmFT5a* 启动子区有 3 个 AP2 结合位点,分布较为分散(图 6,表 2)。*GmFT2a* 和 *GmFT5a* 的启动子区均有 AP2 结合位点,而大豆 TOE 基因均具有 AP2 结构域,因此,大豆 TOE 基因可能是通过结合 *GmFT2a* 和 *GmFT5a* 启动子上的顺式元件来调控其基因转录,从而调控开花。

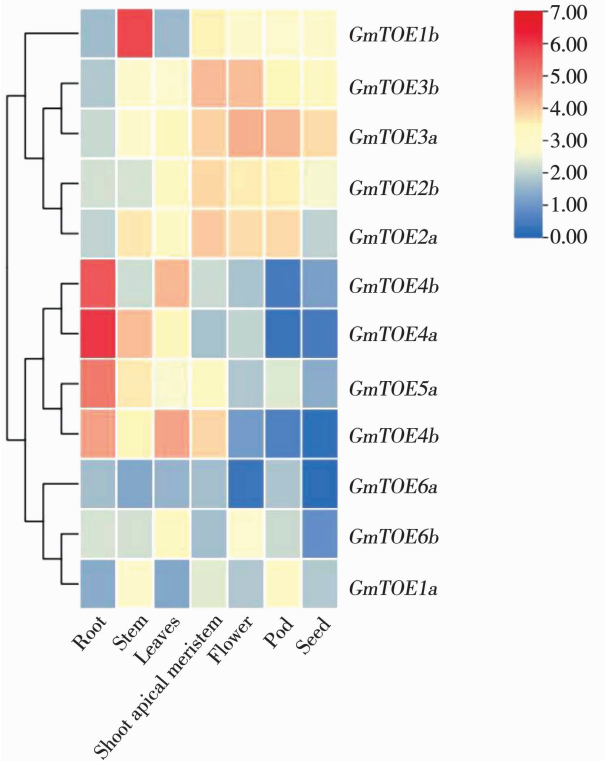


图 5 大豆 12 个 TOE 基因在不同组织中的差异表达
Fig.5 Expression of GmTOEs in different tissues

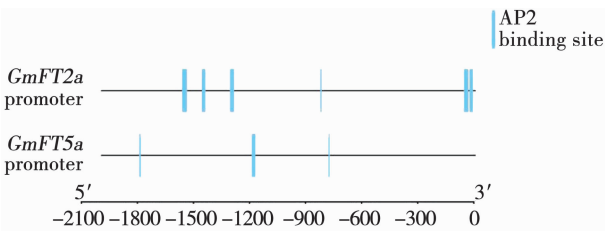


图 6 *GmFT2a* 和 *GmFT5a* 启动子区 AP2 结合域
Fig.6 AP2 binding sites in *GmFT2a* and *GmFT5a* promoter

表 2 *GmFT2a* 和 *GmFT5a* 启动子区 AP2 结合序列预测
Table 2 AP2 binding site rediction in *GmFT2a* and *GmFT5a* promoter

基因名称 Gene name	AP2 结合位置 AP2 binding site	正负链 Strand	p 值 p-value	q 值 q-value	结合序列 Matched sequence
<i>GmFT2a</i>	-12 ~ -31	+	0.0000323	3.23E-06	AGGAACATAGAGAAAAGAAG
<i>GmFT2a</i>	-37 ~ -56	-	0.00009270	9.27E-05	AAAAAAAAAAAAAAAAACTCT
<i>GmFT2a</i>	-39 ~ -58	-	0.00002150	2.15E-05	AGAAAAAAAAAAAAAAAAAACT
<i>GmFT2a</i>	-40 ~ -59	-	0.00002830	2.83E-05	TAGAAAAAAAAAAAAAAAAAAC
<i>GmFT2a</i>	-41 ~ -60	-	0.00001830	1.83E-05	TTAGAAAAAAAAAAAAAAAAAA
<i>GmFT2a</i>	-42 ~ -61	-	0.00007250	7.25E-05	TTTAGAAAAAAAAAAAAAAAAA
<i>GmFT2a</i>	-822 ~ -829	-	0.00004980	4.98E-05	TCTCGTCC
<i>GmFT2a</i>	-1291 ~ -1310	-	0.00001580	1.58E-05	AGAAAAAAAAAGACAAGAAC
<i>GmFT2a</i>	-1293 ~ -1312	-	0.00004200	4.20E-05	AGAGAAAAAAAAAGACAAGA

续表 2

基因名称 Gene name	AP2 结合位置 AP2 binding site	正负链 Strand	p 值 p-value	q 值 q-value	结合序列 Matched sequence
<i>GmFT2a</i>	- 1294 ~ - 1313	-	0. 00000303	3. 03E-06	CAGAGAAAAAAAAGACAAG
<i>GmFT2a</i>	- 1444 ~ - 1463	+	0. 00005850	5. 85E-05	AGTAACGTGAAAAAAATAA
<i>GmFT2a</i>	- 1542 ~ - 1561	+	0. 00003740	3. 74E-05	GAGGAGCAATGGAAGAGAGA
<i>GmFT2a</i>	- 1544 ~ - 1563	+	0. 00000167	1. 67E-06	GGAGCAATGGAAGAGAGAAA
<i>GmFT2a</i>	- 1550 ~ - 1569	+	0. 00002350	2. 35E-05	ATGGAAGAGAGAAAATGAGA
<i>GmFT5a</i>	- 778 ~ - 785	+	0. 00006320	6. 32E-05	TGTCGTCC
<i>GmFT5a</i>	- 1177 ~ - 1196	+	0. 00000455	4. 55E-06	TCTAAAGAAAGAAAGAGACG
<i>GmFT5a</i>	- 1790 ~ - 1799	+	0. 00003420	3. 42E-05	TCATCGTACA

2.7 *GmTOE4b* 和 *GmTOE5b* 大豆开花的功能验证

2.7.1 *GmTOE4b* 单倍型大豆开花期分析 从 424 份大豆自然群体材料的重测序数据^[34]中提取出 10 种 *GmTOE4b* 的有效单倍型,其中 H1 型与参考基因组 Wm82 一致,其余 9 种单倍型都为单个碱基对的变异导致氨基酸置换(图 7)。其中 H7 和 H8 两种单倍型是除了 H1 型以外较多的单倍型。H7 型是

GmTOE4b 基因 CDS 第 715 位的 G 变异成了 A 导致第 239 位氨基酸由甘氨酸(G)变为了丝氨酸(S),这一变异正好处于 *TOE4b* 的第二个 AP2 结构域内,极有可能影响 *GmTOE4b* 的功能。H8 型是 CDS 第 283 位的 T 变异成 A,第 328 位的 C 变异成 A,导致氨基酸序列第 95 位由苯丙氨酸(F)变为异亮氨酸(I),第 110 位氨基酸由组氨酸(H)变为了天冬酰胺(N)。

TOE4b	单倍型 Haplotype												数量 Number
Wm82	H1	T	C	A	C	C	G	C	G	T	G	T	341
	H2	C	C	A	C	C	G	C	G	T	G	T	1
	H3	T	A	A	C	C	G	C	G	T	G	T	4
	H4	T	C	C	C	C	G	C	G	T	G	A	1
	H5	T	C	A	A	C	G	C	G	T	G	T	1
	H6	T	C	A	C	A	G	C	G	T	G	T	1
	H7	T	C	A	C	C	A	C	G	T	G	T	27
	H8	T	C	A	C	C	G	A	G	A	G	T	29
	H9	T	C	A	C	C	G	A	G	T	C	T	4
	H10	T	C	A	C	C	G	C	A	T	G	T	1
W82 CDS 碱基数 CDS nt # in W82		1234	1086	958	848	841	715	328	306	283	245	139	
W82氨基酸数 AA # in W82		412	362	320	283	281	239	110	102	95	82	47	
W82氨基酸 AA identified in W82		S	F	N	A	P	G	H	M	F	S	C	
氨基酸变化 AA change		P	L	H	D	T	S	N	I	I	T	S	

图 7 *GmTOE4b* 在自然群体中的单倍型统计
Fig. 7 Haplotypes of *GmTOE4b* in the natural populations

对主要单倍型 H1(341 个品种)、H7(27 个品种)和 H8(29 个品种)在郑州(34°N)和合肥(32°N)两地开花期的调查结果显示,在郑州种植时始花时间比在合肥种植时长(图 8)。其原因可能是大豆作为光周期敏感的短日照植物,在生长季随着种植区域纬度升高,日照时长增加,出现开花期延后的现象。虽然两地日照时长不一致,但两个地点不同单倍型

的开花期变化趋势一致,即 H7 型开花最早,平均开花时间为郑州 40. 04 d,合肥 35. 88 d;H1 型次之,平均开花时间为郑州 46. 36 d,合肥 42. 78 d;H8 型开花最晚,平均开花时间为郑州 56. 03 d,合肥 52. 07 d(图 8)。3 种单倍型之间在两个地点的开花时间均有显著性差异,表明 *GmTOE4b* 具有调控大豆开花期的功能。

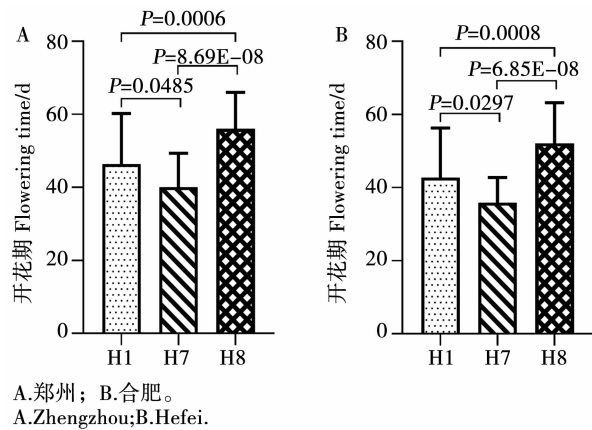


图 8 *GmTOE4b* 主要单倍型开花期统计

Fig. 8 Flowering time variations of main alleles in the natural populations of *GmTOE4b*

2. 7. 2 *GmTOE5b* 单倍型大豆开花期分析 本研究从 424 份材料的重测序数据中共提取出 5 种 *GmTOE5b* 的有效单倍型,其中 H1 型与参考基因组 Wm82 一致,其余 4 种单倍型都是发生了单个氨基酸变异的类型(图 9)。该基因在本研究的样本中较为保守,大部分品种均为 H1 型。

对 H1 型(402 个品种)与除 H1 型外数量较多的 H2 型(10 个品种)在郑州和合肥的开花期的调查结果显示,郑州 H1 型的始花时间平均为 47.77 d,H2 型的始花时间平均为 41.1 d;合肥 H1 型的始花时间平均为 44.37 d,H2 型的始花时间平均为 36.9 d (图 10)。两地 H2 型的始花时间均早于 H1 型,两单倍型的始花时间在两个地点均有显著性差异,说

明 *GmTOE5b* 也具有调控开花的功能。

TOE5b	单倍型 Haplotype					数量 Number
Wm82	H1	G	A	G	C	402
	H2	A	A	G	C	10
	H3	G	C	G	C	2
	H4	G	A	T	C	1
	H5	G	A	G	T	2
W82 CDS 碱基数 CDS nt # in W82		86	99	256	371	
W82氨基酸数 AA # in W82		29	33	86	124	
W82氨基酸 AA identified in W82		G	Q	V	T	
氨基酸变化 AA change		D	H	L	I	

图 9 *GmTOE5b* 在自然群体中的单倍型统计

Fig. 9 Haplotypes of *GmTOE5b* in the natural populations

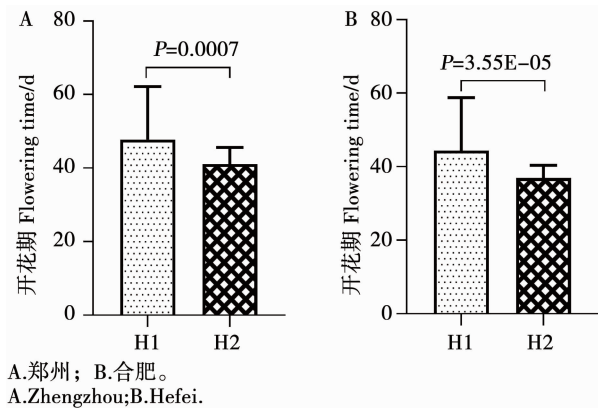


图 10 *GmTOE5b* 主要单倍型开花期统计

Fig. 10 Flowering time variations of main alleles in the natural populations of *GmTOE5b*

3 讨 论

AP2 类转录因子是植物特有的重要转录因子,首个 AP2 基因是在拟南芥中克隆得到的,它具有两个 AP2 结构域,属于控制花发育 ABC 模型中的 A 类基因^[36],同时 AP2 还通过抑制 FT 表达来延迟拟

南芥开花^[37],*miR172* 通过调控其 6 个 AP2 类靶基因(*TOE1*、*TOE2*、*TOE3*、*AP2*、*SMZ* 和 *SNZ*)的蛋白质合成从而调控开花时间是拟南芥重要的开花调控通路之一^[27]。在大豆 AP2 家族中,前人通过拟南芥 TOE 的氨基酸序列在大豆中搜索到 11 个大豆 TOE 基因,其中除 *GmTOE6a* 外,聚类结果与本研究一

致^[31]。本研究由于使用的基因组版本较新,共搜索到12个大豆 TOE 基因,其中 *GmTOE6b* 为本研究新发现的大豆 TOE 基因,将其与拟南芥 TOE 基因聚类分析后发现 *GmTOE6a*、*GmTOE6b* 与拟南芥 *TOE2* 聚为一类。本研究中的12个 TOE 基因在氨基酸序列及结构上是与拟南芥 TOE 最接近的基因,预测这12个 TOE 基因都有1个 *miR172* 的靶位点。虽然这12个大豆 TOE 基因的结构与拟南芥 TOE 高度相似,但大豆与拟南芥的 TOE 基因并不都具有共线性,*GmTOE1a*、*GmTOE1b*、*GmTOE6a* 和 *GmTOE6b* 与拟南芥 TOE 无共线性,可能是由于功能冗余的基因在进化过程中丢失,也可能是由于其染色体共线性区段太小而未能检测到。

这12个大豆 TOE 基因结构非常保守,依据结构的相似性将其分成从 *GmTOE1a* / *GmTOE1b* 到 *GmTOE6a* / *GmTOE6b* 的6对,成对的 TOE 基因表达模式大都相近,说明大豆 TOE 基因具有加倍的现象,并且在功能上存在冗余。但非成对的基因表达模式却有很大的差异,说明它们存在功能分化。

通过群体分析发现 *GmTOE4b* 和 *GmTOE5b* 不同单倍型的开花期不一致,说明它们具有调控大豆开花的功能,且 *GmTOE4b* 的 H7 型单倍型在 AP2 结构域发生变异,使其比 AP2 结构域未变异的类型早花,说明 *GmTOE4b* 的功能是抑制大豆开花。由于在 *GmFT2a* 和 *GmFT5a* 的启动子区检测到多个 AP2 结合位点,并且 *GmTOE4b* 在 AP2 结构域的变异导致开花提前,表明大豆 TOE 基因极有可能与拟南芥一样,通过结合 FT 启动子上的顺式元件抑制 FT 基因转录,从而调控开花期延迟,这也很好地解释了 *GmTOE1a* 能回补拟南芥 *toe1* 的早花表型^[32] 和过表达 *GmTOE4a* 的转基因大豆表现出晚花表型^[31] 的分子机制,但仍需要进一步试验验证。

4 结 论

通过将大豆和拟南芥 AP2 家族的氨基酸序列联合构建进化树,鉴定到12个大豆 TOE 基因,分别为 *GmTOE1a*、*GmTOE1b*、*GmTOE2a*、*GmTOE2b*、*GmTOE3a*、*GmTOE3b*、*GmTOE4a*、*GmTOE4b*、*GmTOE5a*、*GmTOE5b*、*GmTOE6a* 和 *GmTOE6b*,这12个基因的 AP2 结构域及 *miR172* 靶位点序列与拟南芥 TOE 高度相似。*GmTOE2a*、*GmTOE2b*、*GmTOE3a* 和 *GmTOE3b* 与拟南芥 *TOE3* 来源于同一祖先基因。*GmTOE4a*、*GmTOE4b*、*GmTOE5a* 和 *GmTOE5b* 与拟南芥 *TOE1* 来源于同一祖先基因。*GmTOE1a* 和

GmTOE1b 来源于同一祖先基因,在大豆进化过程中随染色体扩增时复制而来,与拟南芥 TOE 基因无共同来源。*GmTOE6a* 和 *GmTOE6b* 在大豆染色体扩增前就已经存在,并且它们是独立起源的。通过基因表达模式分析发现 *GmTOE4b* 和 *GmTOE5b* 在叶中表达量最高, *GmFT2a* 和 *GmFT5a* 启动子均具有 AP2 结合域。对424份大豆材料的 *GmTOE4b* 和 *GmTOE5b* 的单倍型和开花期表型的分析发现这两个基因不同单倍型具有不同的开花时间,说明 *GmTOE4b* 和 *GmTOE5b* 具有调控大豆开花的功能,同时 *GmTOE4b* 和 *GmTOE5b* 基因的不同单倍型也将为定向选育不同地域适应不同光周期条件的高产优质大豆新品种提供新的基因资源。

参考文献

[1] Graham P H, Vance C P. Legumes: Importance and constraints to greater use[J]. Plant Physiology, 2003, 131(3): 872-877.

[2] 姚林. 中美贸易摩擦下的中国大豆产业现状与发展趋势[J]. 中国油脂, 2020, 45(2): 10-14. (Yao L. Current situation and development trend of China's soybean industry under Sino-US trade friction[J]. China Oils and Fat, 2020, 45(2): 10-14.)

[3] 查霆, 钟宣伯, 周启政, 等. 我国大豆产业发展现状及振兴策略[J]. 大豆科学, 2018, 37(3): 458-463. (Zha T, Zhong X B, Zhou Q Z, et al. Development status of China's soybean industry and strategies of revitalizing[J]. Soybean Science, 2018, 37(3): 458-463.)

[4] Watanabe S, Harada K, Abe J. Genetic and molecular bases of photoperiod responses of flowering in soybean[J]. Breeding Science, 2012, 61(5): 531-543.

[5] 夏正俊. 大豆光周期反应与生育期基因研究进展[J]. 作物学报, 2013, 39(4): 571-579. (Xia Z J. Research progresses on photoperiodic flowering and maturity genes in soybean (Glycine max Merr.) [J]. Acta Agronomica Sinica, 2013, 39(4): 571-579.)

[6] Xia Z J, Watanabe S, Yamada T, et al. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering[J]. Proceedings of the National Academy of Sciences, 2012, 109(32): E2155-E2164.

[7] Watanabe S, Xia Z J, Hideshima R, et al. A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering [J]. Genetics Society of America, 2011, 188(2): 395-407.

[8] Watanabe S, Hideshima R, Xia Z J, et al. Map-based cloning of the gene associated with the soybean maturity locus *E3* [J]. Genetics Society of America, 2009, 182(4): 1251-1262.

[9] Liu B H, Kanazawa A, Matsumura H, et al. Genetic redundancy in soybean photoresponses associated with duplication of the

phytochrome A gene[J]. Genetics Society of America, 2008, 180 (2): 995-1007.

[10] Dissanayaka A, Rodriguez T O, Di S, et al. Quantitative trait locus mapping of soybean maturity gene *E5* [J]. Breeding Science, 2016, 66(3): 407-415.

[11] Cober E R. Long juvenile soybean flowering responses under very short photoperiods[J]. Crop Science, 2011, 51(1): 140-145.

[12] Cober E R, Voldeng H D. Low R:FR light quality delays flowering of *E7E7* soybean lines[J]. Crop Science, 2001, 41 (6): 1823-1826.

[13] Cober E R, Molnar S J, Charette M, et al. A new locus for early maturity in soybean[J]. Crop Science, 2010, 50(2): 524-527.

[14] Kong F J, Nan H Y, Cao D, et al. A new dominant gene *E9* conditions early flowering and maturity in soybean [J]. Crop Science, 2014, 54(6): 2529-2535.

[15] Samanfar B, Molnar S J, Charette M, et al. Mapping and identification of a potential candidate gene for a novel maturity locus, *E10*, in soybean[J]. Theoretical and Applied Genetics, 2017, 130(2): 377-390.

[16] Wang F F, Nan H Y, Chen L Y, et al. A new dominant locus, *E11*, controls early flowering time and maturity in soybean[J]. Molecular Breeding, 2019, 39(5): 70.

[17] Lu S J, Zhao X H, Hu Y L, et al. Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield[J]. Nature Genetics, 2017, 49(5): 773-779.

[18] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean [J]. Nature, 2010, 463 (7278): 178-183.

[19] Cao D, Li Y, Lu S J, et al. *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long - day conditions[J]. Plant and Cell Physiology, 2015, 56(12): 2409-2422.

[20] Takeshima R, Hayashi T, Zhu J, et al. A soybean quantitative trait locus that promotes flowering under long days is identified as *FT5a*, a *FLOWERING LOCUS T* ortholog [J]. Journal of Experimental Botany, 2016, 67(17): 5247-5258.

[21] Jiang B J, Zhang S W, Song W W, et al. Natural variations of *FT* family genes in soybean varieties covering a wide range of maturity groups[J]. BMC Genomics, 2019, 20(1): 230.

[22] Sun F, Xu M L, Park C, et al. Characterization and quantitative trait locus mapping of late - flowering from a Thai soybean cultivar introduced into a photoperiod - insensitive genetic background[J]. PLoS One, 2019, 14(12): e0226116.

[23] Kong F J, Liu B H, Xia Z J, et al. Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean[J]. Plant Physiology, 2010, 154(3): 1220-1231.

[24] Guo G Y, Xu K, Zhang X M, et al. Extensive analysis of *GmFTL* and *GmCOL* expression in northern soybean cultivars in field conditions[J]. PLoS One, 2015, 10(9): e0136601.

[25] Kim S, Soltis P S, Wall K, et al. Phylogeny and domain evolution in the *APETALA2-like* gene family [J]. Molecular Biology and Evolution, 2006, 23(1): 107-120.

[26] Schmid M, Uhlenhaut N H, Godard F, et al. Dissection of floral induction pathways using global expression analysis [J]. Development, 2003, 130(24): 6001-6012.

[27] Aukerman M J, Sakai H. Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2 - like* target genes [J]. The Plant Cell, 2003, 15(11): 2730-2741.

[28] Lee Y S, Lee D Y, Cho L H, et al. Rice *miR172* induces flowering by suppressing *OsIDS1* and *SNB*, two *AP2* genes that negatively regulate expression of *Ehd1* and florigens[J]. Rice, 2014, 7(1), 31.

[29] Wu G, Park M Y, Conway S R, et al. The sequential action of *miR156* and *miR172* regulates developmental timing in *Arabidopsis* [J]. Cell, 2009, 138(4): 750-759.

[30] Martin A, Adam H, Díaz-Mendoza M, et al. Graft - transmissible induction of potato tuberization by the microRNA *miR172* [J]. Development, 2009, 136(17): 2873-2881.

[31] Zhao X H, Cao D, Huang Z J, et al. Dual functions of *GmTOE4a* in the regulation of photoperiod - mediated flowering and plant morphology in soybean[J]. Plant Molecular Biology, 2015, 88 (4): 343-355.

[32] Wang T, Sun M Y, Wang X S, et al. Over-expression of *GmGla - regulated* soybean *miR172a* confers early flowering in transgenic *Arabidopsis thaliana* [J]. International Journal of Molecular Sciences, 2016, 17(5): 645.

[33] Kim K D, El Baidouri M, Abernathy B, et al. A comparative epigenomic analysis of polyploidy - derived genes in soybean and common bean[J]. Plant Physiology, 2015, 168(4): 1433-1447.

[34] Lu S, Dong L D, Kong F J, et al. Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication [J]. Nature Genetics, 2020, 52 (4): 428-436.

[35] Chen C J, Chen H, Zhang Y, et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data[J]. Molecular Plant, 2020, 13(8): 1194-1202.

[36] Jofuku K D, den Boer B G, van Montagu M, et al. Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*[J]. The Plant Cell, 1994, 6(9): 1211-1225.

[37] Castillejo C, Pelaz S. The balance between *CONSTANS* and *TEMPRANILLO* activities determines *FT* expression to trigger flowering[J]. Current Biology, 2008, 18(17): 1338-1343.