



# 大豆 ELF3 的生物信息学分析及 *GmELF3Ld* 功能初探

李佳雯<sup>1</sup>, 张帅锋<sup>2</sup>, 吴庚雪<sup>1</sup>, 吴红霞<sup>1</sup>, 杨 龙<sup>1</sup>, 岳岩磊<sup>1</sup>

(1. 河南农业大学 生命科学院, 河南 郑州 450002; 2. 河南农业大学 资源与环境学院, 河南 郑州 450002)

**摘要:** ELF3 是生物节律钟重要的组成部分, 在植物的生长发育中发挥重要作用。大豆是光周期敏感的短日作物, 含有 5 个 ELF3 同源基因, 其中 *GmELF3a/J* 是引起大豆绿色革命的明星基因, 主要调控大豆童期的发育。为分析其他 4 个 *GmELF3* 同源基因的功能, 本研究利用 MEGA-X、Multalin SOPMA 和 ExPASy-Prot Param 等工具对它们的进化关系、蛋白功能、特性、亚细胞定位及基因表达情况进行生物信息学分析, 同时通过转化拟南芥初步研究 *GmELF3Ld* 的功能。结果表明: 大豆中 5 个 ELF3 基因可以分为两类, 其中 *GmELF3a/J*、*GmELF3b* 和 *GmELF3c* 为 A 类, *GmELF3Ld* 和 *GmELF3Le* 为 B 类。 *GmELF3Le* 的  $\alpha$  螺旋含量及不稳定性指数均为最高。另外, 这 5 个蛋白均为亲水性非跨膜蛋白, 含有多个丝氨酸和酪氨酸磷酸化位点, 为蛋白活性调控提供条件。B 类基因表达量远高于 A 类且表现出明显的营养器官与生殖器官分化, 即 *GmELF3b* 和 *GmELF3c* 基因在花和种子中高表达, *GmELF3Ld* 和 *GmELF3Le* 则在叶片中高表达。转基因拟南芥研究结果表明, *GmELF3Ld* 抑制拟南芥开花关键基因 *FT* 和 *CO* 的表达, 延迟拟南芥的花期。本研究结果表明两类 *GmELF3* 同源蛋白具有不同的生物学功能, *GmELF3Ld* 延迟拟南芥的花期。

**关键词:** 大豆; ELF3; 生物信息学; *GmELF3Ld*; 拟南芥; 转基因

开放科学(资源服务)标识码(OSID):



## Bioinformatics Analysis of *GmELF3* and Preliminary Study on the Function of *GmELF3Ld*

LI Jia-wen<sup>1</sup>, ZHANG Shuai-feng<sup>2</sup>, WU Geng-xue<sup>1</sup>, WU Hong-xia<sup>1</sup>, YANG Long<sup>1</sup>, YUE Yan-lei<sup>1</sup>

(1. College of Life Sciences, Henan Agricultural University, Zhengzhou 450002, China; 2. College of Resources and Environment, Henan Agricultural University, Zhengzhou 450002, China)

**Abstract:** ELF3, an indispensable component of the circadian clock, plays an important role in plant growth and development. Soybean, a photoperiod-sensitive short-day crop, contains five ELF3 homolog genes. Of them, *GmELF3a/J* is a key gene controlling the juvenility and leading to the green revolution in soybean. In order to analyze the function of the other four *GmELF3* homologs, and to study the functional differences and genetic molecular mechanism of *GmELF3* in depth, we preformed bioinformatical analysis with MEGA-X, Multalin SOPMA, ExPASy-Prot Param and so on to predict the phylogenetics, protein function, protein property, subcellular location and gene experssion of *GmELF3* homologs (*GmELF3s*). Meanwhile, the function of *GmELF3Ld* was further preliminarerily analyzed in transgenic *Arabidopsis*. The results showed that these five homologs can be grouped into two groups: *GmELF3a/J*, *GmELF3b* and *GmELF3c* are in group A, and *GmELF3Ld* and *GmELF3Le* are in group B. At the same time, *GmELF3Le* contains the most  $\alpha$ -helix and has the highest instability index. These five homologs are all hydrophilic non-trans-membrane proteins. They contained multiple Ser- and Tyr- phosphorylation sites, which allow the precise regulation of their activities. The expression level of B-group genes was much higher than that of A-group genes, and showed obvious differentiation between vegetative and reproductive organs. Such as *GmELF3Ld* and *GmELF3Le* are highly expressed in leaves, while *GmELF3b* and *GmELF3c* are highly expressed in flower and seed. The study of transgenic *Arabidopsis* showed that *GmELF3Ld* overexpression can inhibit the expression of *FT* and *CO* and delay the flowering time. These results demonstrated that these two types of homologous proteins have the different biological functions and *GmELF3Ld* can delay flowering of *Arabidopsis Thaliana*.

**Keywords:** soybean; ELF3; bioinformatics; *GmELF3Ld*; *Arabidopsis*; transgenesis

ELF3 (Early flowering 3) 是植物特有的一类高度保守核蛋白, 在生物节律钟调控系统中不可缺少。它的突变使光信号门控效应消失, 从而导致生物节律失调<sup>[1-3]</sup>。在拟南芥生物节律钟系统中, ELF3 是生物节律钟系统调控网络中的核心枢纽, 它

可以整合上下游信号, 并适时传递给下游代谢途径, 进而调节植物体发育的进程使其适应外界环境变化。ELF3 与 LUX 和 ELF4 蛋白直接结合形成三蛋白体复合物 EC (Evening Complex)<sup>[4]</sup>。EC 复合物各组分相互调控, 其中 ELF3 的核定位是 EC 复合

收稿日期: 2021-06-13

基金项目: 国家自然科学基金(32001573); 河南农业大学青年英才项目(30500613)。

第一作者: 李佳雯(2001—), 女, 硕士研究生, 主要从事基因功能挖掘研究。E-mail: lijiawen8391@126.com。

通讯作者: 岳岩磊(1984—), 女, 博士, 副教授, 主要从事大豆分子育种研究。E-mail: yueyanlei@henau.edu.cn。

物进入细胞核的基础。生物钟调控系统中,ELF3 调节生物节律钟系统的组成,同时 ELF3 还受到生物节律钟系统的反向调控。如拟南芥的 ELF3 与 LUX 共同作用,结合 PRR9 启动子,进而激活 LHY 和 CCA1<sup>[5-6]</sup>;而 CCA1 又可结合 ELF3 的启动子,进一步抑制 ELF3 的表达<sup>[7]</sup>,由此生物节律钟基因在植物体内形成互锁的转录—翻译反馈环路。

研究表明 ELF3 在植物适应环境中起着关键性作用,是植物对光及温度响应调控的重要因子,在生物节律钟中可能起到钟摆的作用。首先,ELF3 可以整合光信号,协调光形态建成。在拟南芥中,ELF3 突变导致下胚轴变长,花期缩短,同时生物节律钟失调<sup>[3,8]</sup>。ELF3 参与蓝光和紫外光信号转导,起到门控调节作用<sup>[8-9]</sup>。ELF3 与光敏色素互作影响生物节律钟,并与 E3 泛素连接酶 CONSTITUTIVELY PHOTOMORPHOGENIC 1(COP1)介导的 ELF3 蛋白质降解有关<sup>[9-11]</sup>。在光信号途径中,ELF3 的动态稳定性确定了植物生物节律,进而协调其营养生长及生殖生长的进程。其次,ELF3 还与植物响应温度有关。拟南芥突变体 elf3-1 具有组成型温度反应,ELF3 是传导温度信号的关键节点<sup>[12]</sup>。当环境温度升高时,夜晚复合物 EC 活性降低,上调 GI、LUX 和 PRR7,与此同时 ELF3 结合 PRR9、LUX 和 PIF4 启动子的能力随着温度升高显著下降<sup>[13]</sup>。

作物中 ELF3 有多个同源基因,其功能有的相似,有的则发生了较大的分化,如水稻的 *OsELF3-1* 可调节生物节律钟系统,但 *OsELF3-2* 不具有该功能<sup>[14]</sup>。而拟南芥 ELF3(*AT2G25930*)基因是生物节律钟不可缺少的一个组分,对植物光形态建成、温度反应以及生殖生长等均起重要作用<sup>[8]</sup>。大豆是对光温反应极为敏感的短日作物,包含 5 个 ELF3 同源基因,其中 *GmELF3a/J* 是引起大豆绿色革命的明星基因,与拟南芥 ELF3 功能不同,*GmELF3a/J* 在高温短光照条件下童期显著延长<sup>[15]</sup>。而另外 4 个 *GmELF3s* 的功能研究尚属空白。以往研究证实大豆同源基因间功能分化明显,甚至功能相反,如开花整合因子 *GmFT1a* 与 *GmFT2a*<sup>[16]</sup>。本研究利用生物信息学方法,首先对 *GmELF3s* 进行进化分析和序列特征分析,其次分析蛋白二级结构、三维结构和蛋白特性,最后对其在细胞中的定位及组织表达特性进行比较分析;同时,通过转基因方法研究 *GmELF3Ld* 在模式植物拟南芥中的功能,旨在为深入探讨大豆 ELF3 家族基因功能和分子遗传机制奠定基础。

1 材料与方法

1.1 材料

pBI111L-GmELF3Ld 载体;拟南芥野生型 Col-0。

1.2 方法

1.2.1 序列获取 以拟南芥 ELF3(*AT2G25930*)序列为参考序列,在 Phytozome V12.1(<https://phytozome.jgi.doe.gov/>)数据库中查找 *GmELF3a/J* 及另外 4 个大豆同源基因的序列信息。

1.2.2 蛋白进化分析 在 MEGA-X 软件中采用邻接法(Neighbor-joining)进行分析,利用 Bootstrap 自助抽样法对其检测,并以泊松模型(Poisson model)为替代模型<sup>[17]</sup>,最终得出 *GmELF3s* 氨基酸序列的系统进化树,对比分析 *GmELF3s* 蛋白之间的亲缘性。

1.2.3 基因组序列分析 利用 Exon-Intron Graphic Maker 绘图工具(<http://www.wormweb.org/exonintron>)绘制 5 个基因编码序列结构图,同时将 5 个蛋白的氨基酸序列以 MultAlin-Fasta 格式输入 MultAlin 网站(<http://multalin.toulouse.inra.fr/multalin/>),选取默认参数对其进行多序列同源对比分析<sup>[18]</sup>。

1.2.4 蛋白二级结构分析 利用 NPS@:SOPMA secondary structure prediction(<https://npsa-prabi.ibcp.fr/>)在线网站,设置螺旋、折叠、转角、卷曲 4 个构象状态数,以 8 为相似度阈值,预测和分析 *GmELF3s* 蛋白的二级结构<sup>[19]</sup>。

1.2.5 蛋白三维结构分析 利用 I-TASSER 服务器(<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>)预测 *GmELF3s* 蛋白的三维结构,构建蛋白三级结构模型<sup>[20]</sup>,结合进化树对 5 个蛋白进行对比分析。

1.2.6 蛋白质理化特性分析 利用 ExPASy-ProtParam Tool(<https://web.expasy.org/>)计算各氨基酸序列的分子量、理论 pI、消光系数、半衰期、不稳定性指数、脂肪族指数和亲水性的总平均值等<sup>[21]</sup>;再利用 ExPASy-ProtScale(<https://web.expasy.org/protscale/>)在线工具,选择氨基酸等级 Hphob./Kyte & Doolittle,预测各位点的亲疏水性数值<sup>[21]</sup>。

1.2.7 在细胞中的功能预测 基于蛋白质序列和结构,利用 NetPhos 3.1 Server(<http://www.cbs.dtu.dk/services/>)对 *GmELF3s* 蛋白中所有丝氨酸、苏氨酸、酪氨酸进行磷酸化位点预测<sup>[22]</sup>;基于 HMM 法,在 TMHMM Server v. 2.0(<http://www.cbs.dtu.dk/>)网站输入 *GmELF3s* 蛋白 FASTA 格式的氨基酸序列,分析预测其跨膜结构域;基于原始的 PSORT 原理,通过 PSORT: Protein Subcellular Localization Prediction Tool(<https://www.genscript.com/>)预测

GmELF3s 蛋白的亚细胞定位<sup>[23]</sup>。

1.2.8 不同组织基因表达量预测分析 利用在线网站 Phytozome V12.1 (<https://phytozome.jgi.doe.gov/pz/portal.html>) 在 *Glycine max* Wm82. a2. v1 大豆基因组的预测序列中查找 5 个 *GmELF3s* 基因在大豆花、茎、叶、根、荚以及种子等器官中的表达量并进行对比分析。

1.2.9 转基因拟南芥的表型和表达分析 利用引物 J07-F (3'-GACTCTAGAGGATCCCGTACATGATG-AATAGGATGAGTGGGG-5') 和 J07-R (3'-GGAAAT-

TCGAGCTCACTAGTTTACAAATGTTTCCTTTCTTC-5'), 以大豆栽培品种中黄 39 的 cDNA 为模板扩增 *GmELF3Ld*, 并用同源重组酶将目的片段连接到 XhoI 线性化的 pBI111L 载体上, 得到过表达载体 pBI111L-*GmELF3Ld*。将 pBI111L-*GmELF3Ld* 基因过表达载体转入模式植物拟南芥, 获得过表达转基因拟南芥, 分析开花时间和莲座叶数目表型差异, 并利用实时荧光定量技术分析关键基因的表达量, 引物序列详见表 1<sup>[15]</sup>。

表 1 实时荧光定量 PCR 引物序列

Table 1 The primer sequences for real-time fluorescence quantitative PCR

引物名称 Primer name	引物序列(5'-3') Primer sequence(5'-3')
ACTIN-RT-250FP	CTTAACCCAAAGGCCAACAGA
ACTIN-RT-250RP	GCAAGGTCAAGACGGAGGAT
GmELF3b-RT-FP	ATGGCACTCTATGAACAATTACAG
GmELF3b-RT-RP	CATGAATTTGGTTAAGTTAGTCTTTTC
RT-FT-FP	ATTCTGCAACTACAGGGGC
RT-FT-RP	AAGTACATGGCCGCTACTGG
RT-CO-FP	TGCACCTAGCCCAGGAAATC
RT-CO-RP	CCATCATCGGTCGTGGACTC

1.3 数据分析

利用 Excel 2021 对获得的数据进行分析。

2 结果与分析

2.1 *GmELF3s* 基因序列与编码蛋白结构分析

2.1.1 基因进化树分析 以拟南芥 ELF3 序列为参考序列, 在 Phytozome 数据库中检索到 *GmELF3a/J* 及另外 4 个大豆 ELF3 的序列信息: *GmELF3a/J* (Glyma. 04G050200)、*GmELF3b* (Glyma. 14G091900)、*GmELF3c* (Glyma. 17G231600)、*GmELF3Ld* (Glyma. 07G013500) 和 *GmELF3Le* (Glyma. 08G197500)。

进化树分析结果表明, *GmELF3s* 分为 2 个聚

类。A 类中包含 3 个基因, 分别是位于 4 号染色体上的 *GmELF3a/J*, 位于 14 号染色体上的 *GmELF3b* 和 17 号染色体上的 *GmELF3c*, 其中 *GmELF3b* 与 *GmELF3c* 亲缘关系较近, 二者的遗传距离为 0.2, 同时与 *GmELF3a/J* 的遗传距离均为 0.5。因此 A 类还可细分为两类, *GmELF3a/J* 为 A1 类, *GmELF3b* 和 *GmELF3c* 为 A2 类。B 类中包含 2 个基因, 分别是位于 7 号染色体的 *GmELF3Ld* 和位于 8 号染色体的 *GmELF3Le*, 二者的遗传距离为 0.1 (图 1)。结果显示在这些基因当中, 遗传距离最近的是 *GmELF3Ld* 和 *GmELF3Le*, 遗传距离最远的是 *GmELF3Le* 与 A2 类的两个基因。

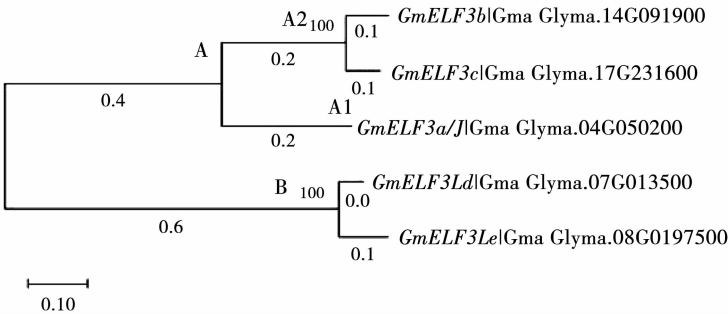
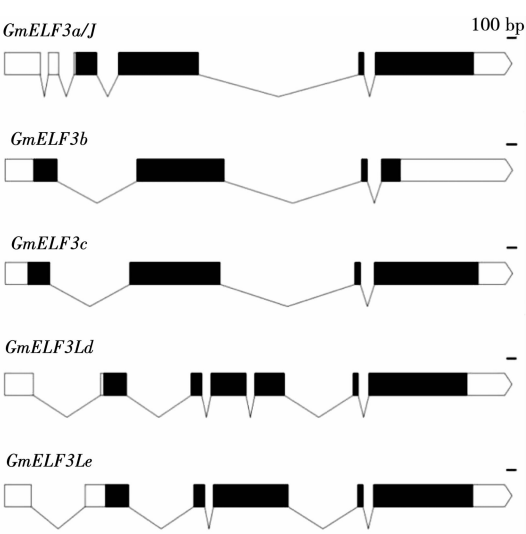


图 1 *GmELF3s* 系统进化树

Fig. 1 The phylogenetic tree of soybean *GmELF3s*

2.1.2 蛋白序列一级结构分析 蛋白一级结构分析结果如图 2 所示, *GmELF3a/J*、*GmELF3b*、*GmELF3c*、*GmELF3Ld* 和 *GmELF3Le* 基因的全长分别为 5 360,4 754,4 917,5 166 和 5 069 bp,编码蛋白长度分别为 714,418,550,679 和 707 个氨基酸,且 5 个基因均为断裂基因,基因编码区分别有 4,4,4,6 和 5 个内显子。说明在大豆进化与选择过程中,*GmELF3s* 基因不仅碱基长度发生了变异,而且在基因断裂区域也发生了极大的变异。

氨基酸序列多重对比结果显示,5 个 *GmELF3s* 蛋白的长度和序列变异均较大(图 3a)。根据进化树结果进一步比对显示, A2 类的 *GmELF3b* 与 *GmELF3c* 相似度约为 90%(图 3b),同时 *GmELF3b* 和 *GmELF3c* 与 *GmELF3a/J* 基因结构差异较大(图 2);B 类的 *GmELF3Ld* 与 *GmELF3Le* 相似度大于 90%(图 3c);A 类和 B 类间的氨基酸序列差异小于 50%。

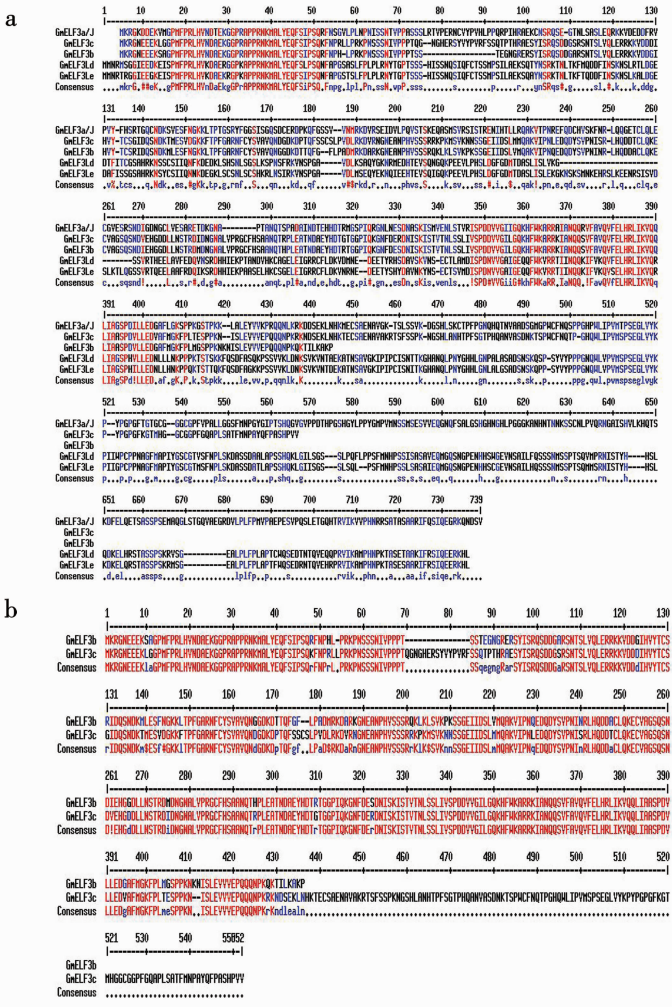


注:白色方框为非编码区,黑色方框为外显子,黑色折线为内含子。

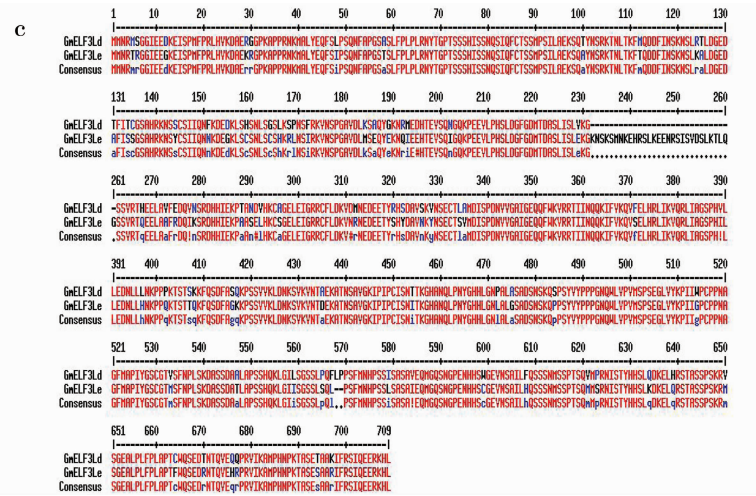
Note:The white box represents UTR, the black box represents the exon, and the black line represents the intron.

图 2 *GmELF3s* 基因的结构

Fig. 2 The schematic diagram of *GmELF3s* gene structure







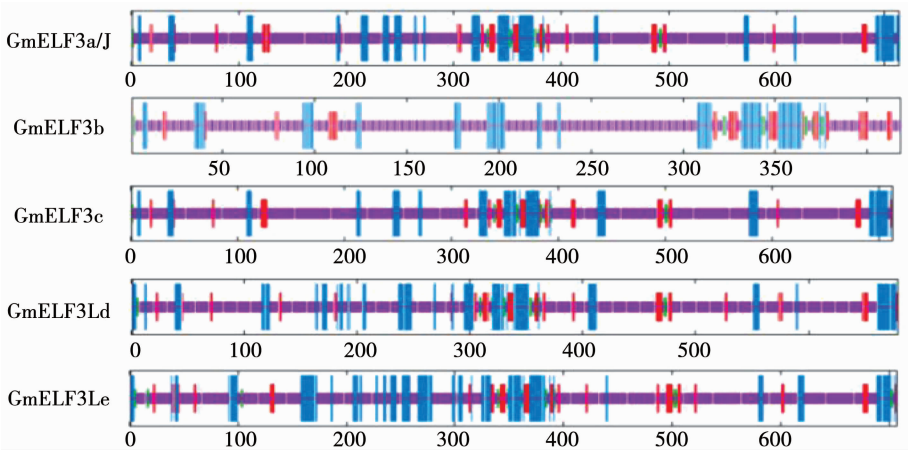
注:红色表示一致性为 100% 的序列;a. 5 个 GmELF3s 氨基酸多序列比对;  
b. GmELF3b和 GmELF3c 蛋白序列比对;c. GmELF3Ld 和 GmELF3Le 蛋白序列  
比对。

Note:The red sequence represents the alignment results with 100% consistency;  
a. Sequence comparison between the five Gm-ELF3s;b. Sequence comparison between  
GmELF3b and GmELF3c; c. Sequence comparison between GmELF3Ld and  
GmELF3Le.

图 3 GmELF3s 蛋白多序列比对  
Fig. 3 The sequence alignment of GmELF3 proteins

2.1.3 蛋白序列二级结构分析 蛋白二级结构分析结果表明, $\alpha$ 螺旋和无规则卷曲是 GmELF3s 蛋白的主要成分,3 类蛋白在二级结构上的分化并不明显(图 4 和表 2),这应该是由遗传的同源性决定的。GmELF3Le 蛋白中的  $\alpha$  螺旋含量最高,为 20.37% ;GmELF3c 蛋白的无规则卷曲含量最高,

为 77.42% ;GmELF3b 蛋白的  $\beta$  折叠和延长链含量最高,分别为 2.63% 和 8.61%。在蛋白质的 4 种二级结构中, $\alpha$  螺旋因其螺旋结构以及特殊的连接方式而成为蛋白二级结构中最稳定的形式。因此如忽略其他因素影响,GmELF3Le 蛋白的结构最为稳定。



蓝色代表  $\alpha$  螺旋;红色代表延长链;绿色代表  $\beta$  折叠;紫色代表无规则卷曲。  
The blue area represents the  $\alpha$  helix; The red area represents the extended chain; The green area  
represents the  $\beta$  angle; The purple area represents the irregular curl.

图 4 GmELF3s 蛋白的二级结构分析  
Fig.4 The secondary structure prediction diagram of GmELF3 proteins

表 2 GmELF3s 蛋白二级结构类型和含量

Table 2 The types and contents of secondary structure in GmELF3 proteins

单位: %

蛋白 Protein	$\alpha$ 螺旋 $\alpha$ helix	$\beta$ 折叠 $\beta$ angle	无规则卷曲 Irregular curl	延长链 Extended chain
GmELF3a/J	15.13	1.82	76.19	6.86
GmELF3b	17.46	2.63	71.29	8.61
GmELF3c	13.94	1.82	77.42	7.01
GmELF3Ld	17.67	2.06	73.05	7.22
GmELF3Le	20.37	2.55	69.31	7.78

2.1.4 蛋白三维结构分析 如图 5 所示,5 个蛋白都为球形蛋白,蛋白表面均有明显的凹槽。5 个蛋白的三维结构特性与进化树分类结果相符,即 B 类蛋白 GmELF3Ld 和 GmELF3Le 结构高度同源,A2 类蛋白 GmELF3b 和 GmELF3c 略有差异但整体结构相

似,而 GmELF3a/J 与 A2 类两个蛋白具有一定的同源性。结果进一步说明 GmELF3s 在进化过程中分别向 3 个不同方向完成了不同程度的分化,产生了分子结构差异。

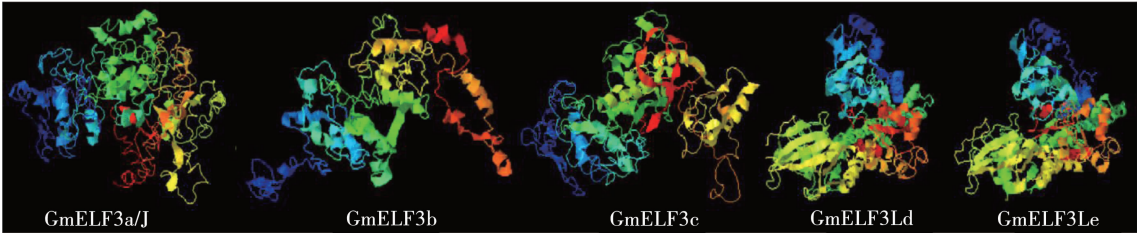


图 5 GmELF3s 蛋白三级结构模型

Fig. 5 The tertiary structure model of GmELF3 proteins

2.2 GmELF3s 蛋白理化特性分析

蛋白的生理生化指数分析结果如表 3 所示,除半衰期外,GmELF3s 蛋白的其他理化特性指标均存在差异。5 个蛋白中分子量最大的是 GmELF3Le,最小的是 GmELF3b,相差 31 834.85 Da。蛋白分子 pI 值从高到低依次为: GmELF3Le、GmELF3b、GmELF3a/J、GmELF3Ld 和 GmELF3c,分别为 9.02, 9.00, 8.84, 8.36 和 8.25。脂肪指数最大的是 GmELF3b。不稳定指数最大和最小的分别是 GmELF3Le 和 GmELF3b,说明 GmELF3s 分子的降解途径具有一致性,但影响蛋白稳定性的因素中,除

了二级结构含量外,一级结构的特性也起着决定性作用。在 GmELF3s 蛋白分子中含量最多的氨基酸为 Ser、Pro、Asn、Leu、Lys 和 Gly,其中 Ser 是 5 种蛋白分子中含量最高的氨基酸。不稳定指数分析结果表明,如果蛋白分子中所有半胱氨酸的巯基基团均形成二硫键,在水中 280 nm 波长的消光系数最大和最小的分别是 GmELF3Ld 和 GmELF3b。5 种蛋白带电荷的相对残基总数相差较大,但每种蛋白残基中都是带正电荷比带负电荷的数量多一些。结果说明 5 个 GmELF3 蛋白的基本特性存在差异。

表 3 GmELF3s 蛋白理化特性分析

Table 3 The physical and chemical indexes of GmELF3 proteins

蛋白 Protein	氨基酸数量 Number of amino acids	分子量 Molecular weight/Da	理论 pI Theoretical pI	最丰富的氨基酸含量 The content of the most abundant amino acid/%					
				Ser	Pro	Asn	Leu	Lys	Gly
GmELF3a/J	714	77622	8.84	10.6	8.3	6.0	6.2	5.6	8.8
GmELF3b	418	46246	9.00	9.3	6.9	7.2	6.7	7.7	6.7
GmELF3c	550	60638	8.25	10.2	8.4	7.1	5.8	6.4	6.7
GmELF3Ld	679	74564	8.36	13.7	7.8	6.3	7.2	6.8	5.2
GmELF3Le	707	78082	9.02	14.3	6.8	6.5	7.2	7.5	5.4

续表 3

消光系数 Extinction coefficients	半衰期 Half-life	不稳定指数 Instability index	脂肪指数 Aliphatic index	亲水性平均值 Grand average of hydropathicity	带电荷的残基总数 Total number of negatively and positively charged residues	
					charged residues	
					正电荷 ( + ) Arg + Lys	负电荷 ( - ) Asp + Glu
0.416	> 10 h	49.23	63.31	-0.675	80	70
0.318	> 10 h	44.28	68.80	-0.769	56	49
0.551	> 10 h	49.17	62.36	-0.744	63	60
0.597	> 10 h	53.87	67.94	-0.597	71	67
0.486	> 10 h	61.24	67.75	-0.698	84	71

氨基酸序列亲疏水性分析结果如图 6 所示, 5 个序列得分均为正数值多于负数值。GmELF3a/J 蛋白的最高值为 1.867, 在第 361 个氨基酸处, 疏水性最强; 最低值为 -3.011, 在第 709 个氨基酸处, 亲水性最强。GmELF3b 蛋白的最高值为 1.867, 在第 352 个氨基酸处, 疏水性最强; 最低值为 -3.167, 在第 407 个氨基酸处, 亲水性最强。GmELF3c 蛋白的最高值为 1.867, 在第 370 个氨基酸处, 疏水性最

强; 最低值为 -3.489, 在第 424 及 425 个氨基酸处, 亲水性最强。GmELF3Ld 蛋白的最高值为 1.444, 在第 628 个氨基酸处, 疏水性最强; 最低值为 -3.022, 在第 287 个氨基酸处, 亲水性最强。GmELF3Le 蛋白的最高值为 1.489, 在第 560 个氨基酸处, 疏水性最强; 最低值为 -3.056, 在第 315 个氨基酸处, 亲水性最强。说明 5 个 GmELF3 蛋白均为亲水性蛋白, 但具体亲水性还需盐析试验进一步确定。

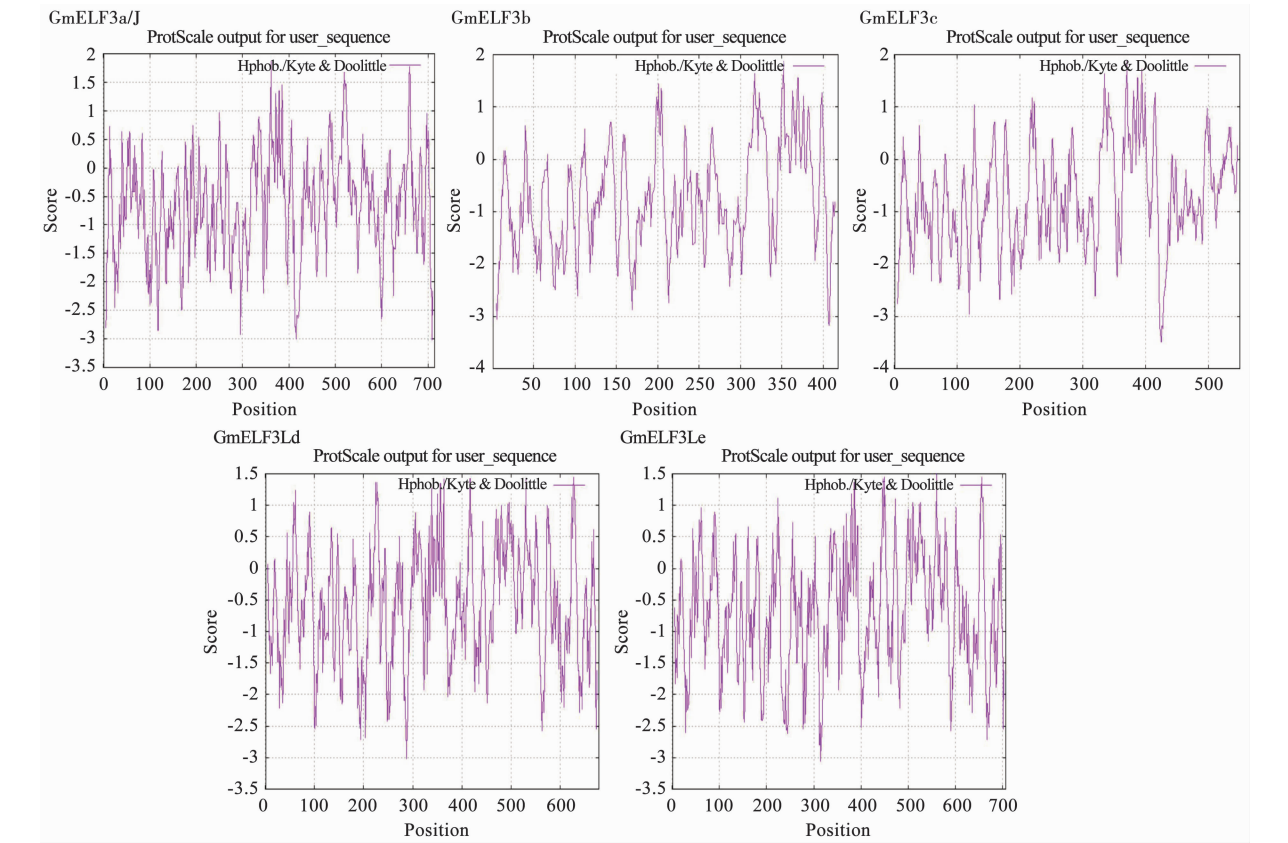


图 6 GmELF3s 蛋白的亲疏水性预测  
Fig. 6 The hydrophobic prediction of GmELF3 proteins

2.3 GmELF3s 在细胞中的功能预测

2.3.1 蛋白定位分析 跨膜结构域预测分析结果如图 7 所示, 5 个 GmELF3 蛋白均没有跨膜结构

域。说明这些蛋白即不是分泌蛋白也不是膜蛋白, 它们可能定位于细胞核、细胞质基质或细胞器基质中。

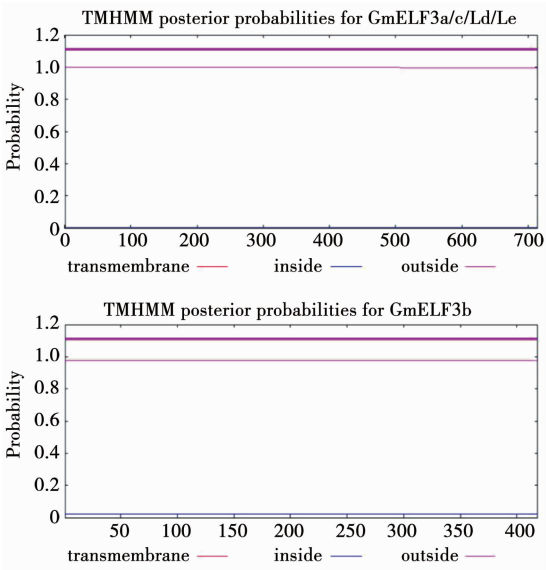
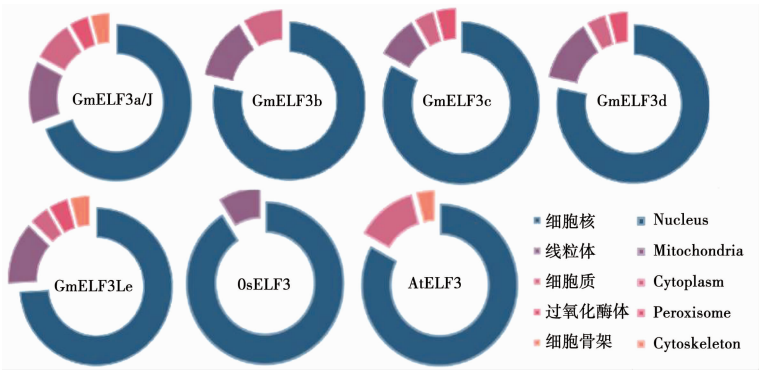


图7 GmELF3s 蛋白跨膜结构域分析  
Fig.7 The analysis of the transmembrane domains in GmELF3 proteins

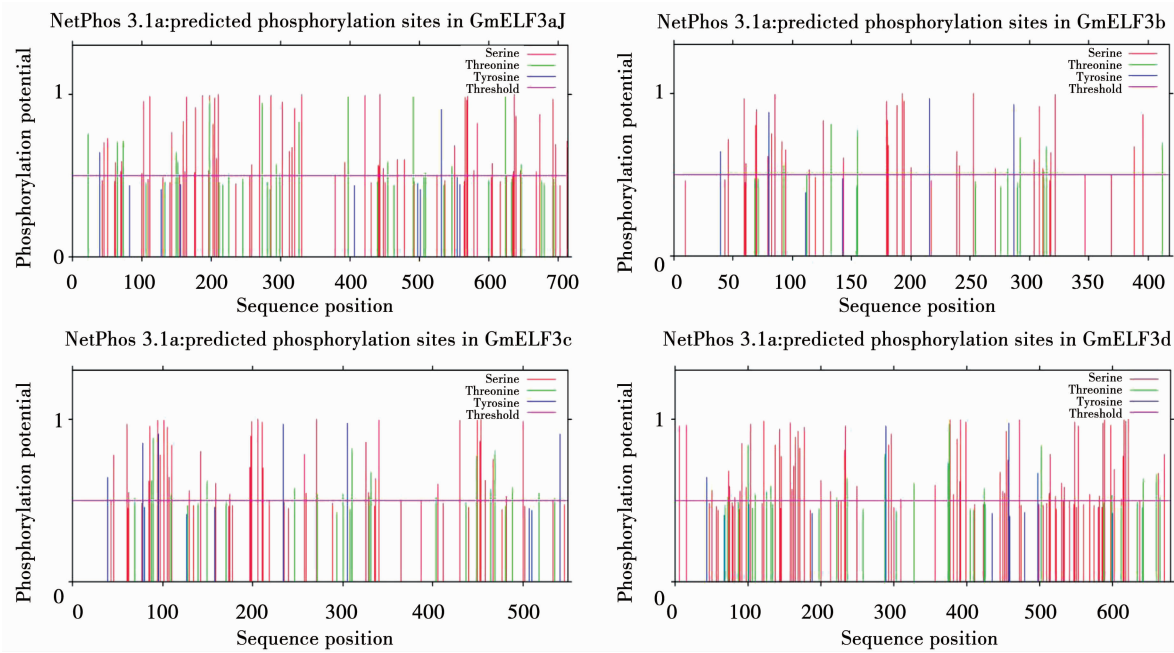
亚细胞定位分析结果如图8所示,5个蛋白的定位概率、位置及种类虽存在差异,但都是在细胞核出现的可能性最大,其次是线粒体基质和细胞质基质(图8),该结果与拟南芥和水稻 ELF3 的定位结果相似,推测这5个蛋白很可能为转录因子。

2.3.2 蛋白磷酸化位点预测分析 蛋白磷酸化位点预测结果如图9和表4所示,GmELF3a/J、GmELF3b、GmELF3c、GmELF3Ld和GmELF3Le含有丝氨酸、苏氨酸和酪氨酸磷酸化位点的总数分别为78,43,62,86和97个。可以明显看出5个蛋白中的磷酸化位点分布具有一定的相似性,但都是丝氨酸的磷酸化最多且最为保守,苏氨酸次之,酪氨酸最少,且磷酸化数值高的位点集中在丝氨酸上。磷酸化位点的具体分布详见附表1(OSID)。结果说明GmELF3s极有可能是与其他蛋白结合发挥作用,该结果与拟南芥的ELF3为EC蛋白复合物的脚手架的结果具有一致性<sup>[11]</sup>。



注:OsELF3 为水稻 ELF3 同源基因;AtELF3 为拟南芥 ELF3 基因。  
Note:OsELF3 and AtELF3 is the homologous gene of rice and Arabidopsis respectively.

图8 GmELF3s 蛋白的亚细胞定位预测  
Fig.8 The subcellular localization prediction of GmELF3 proteins





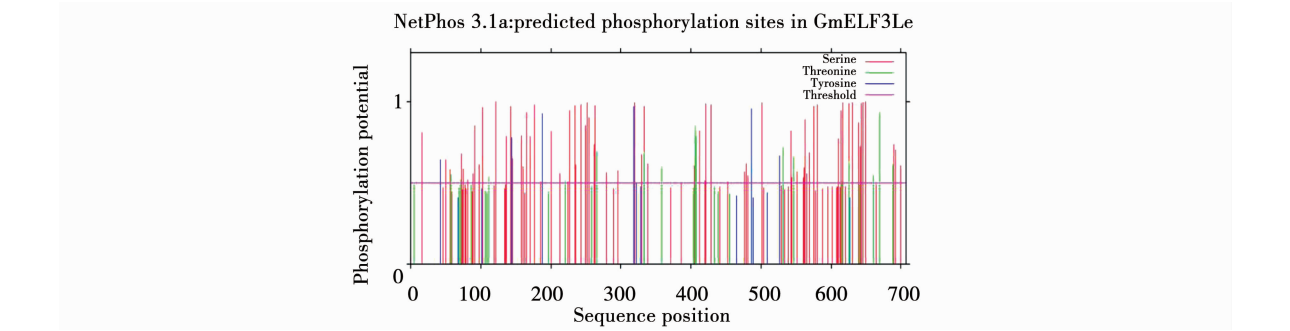


图 9 *GmELF3s* 蛋白磷酸化位点预测

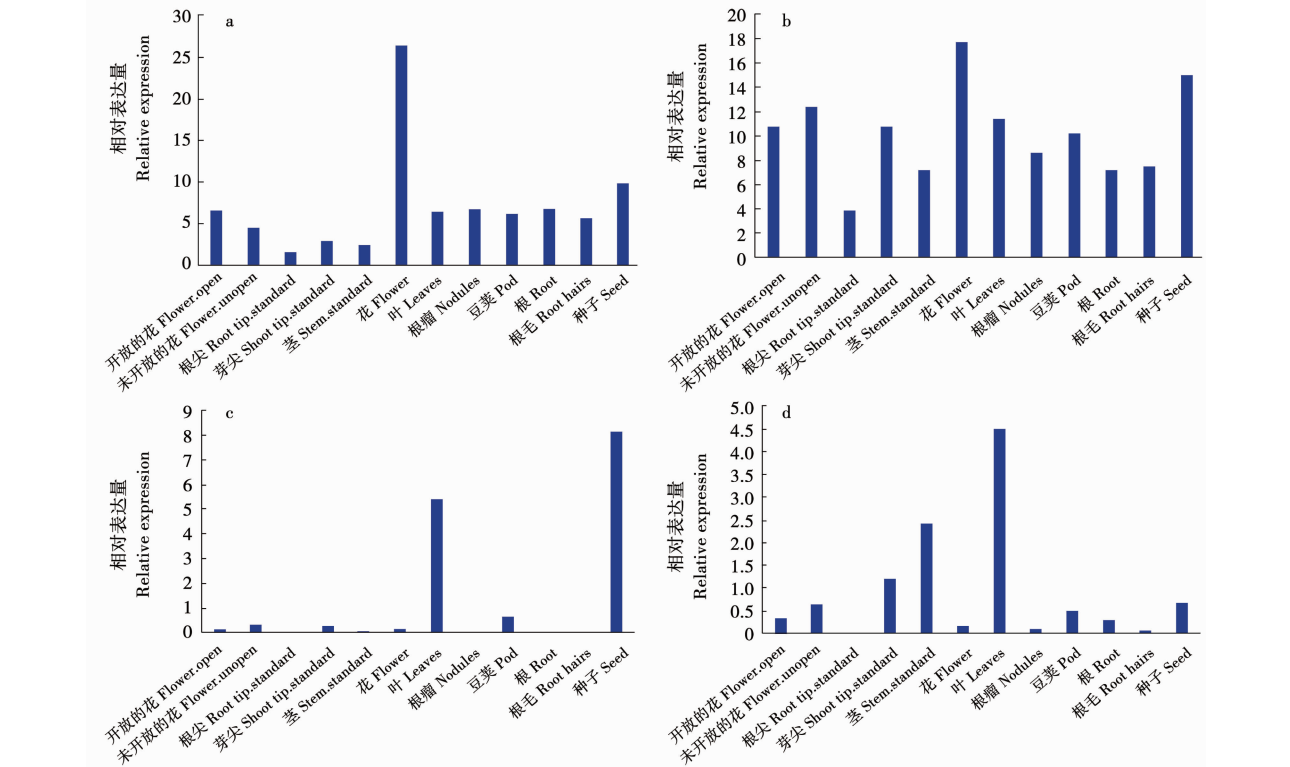
Fig. 9 The phosphorylation sites prediction of *GmELF3* proteins

表 4 <i>GmELF3s</i> 磷酸化位点数量					单位:个
Table 4 The phosphorylation sites number of <i>GmELF3</i> proteins					
蛋白 Protein	丝氨酸磷酸化位点 Serine phosphorylation sites	苏氨酸磷酸化位点 Threonine phosphorylation sites	酪氨酸磷酸化位点 Tyrosine phosphorylation sites	磷酸化数值为 0.99 左右的位点 The phosphorylation value about 0.99 sites	
<i>GmELF3a/J</i>	58	18	2	9	
<i>GmELF3b</i>	31	8	4	3	
<i>GmELF3c</i>	38	18	6	7	
<i>GmELF3Ld</i>	64	18	4	6	
<i>GmELF3Le</i>	74	17	6	7	

2.4 大豆 *GmELF3s* 基因的表达分析

组织特异性表达分析结果如图 10 所示, *GmELF3Ld* 基因在叶片和种子中表达量比较高, *GmELF3Le* 基因在叶片和茎中表达量比较高, *GmELF3b* 基因在花中表达量比较高, *GmELF3c* 基因在花和种子中表达量比较高。整体来看, *GmELF3s* 基因在不同部位表达差异较大, 结合系统进化树分析, B 类基因在营

养器官中高表达, 而 A2 类基因在生殖器官中高表达, 前人研究表明 *GmELF3a/J* 基因不仅在叶中有较高的表达量, 在花和种子中也有较高的表达量, 同时在根瘤和根毛里也有表达量<sup>[15]</sup>, 结合蛋白结构分析结果(图 2、4 和 5), 推测 *GmELF3s* 在进化中朝着 3 个方向进行, 而 *GmELF3a/J* 很有可能在进化中由另外 2 个 *GmELF3s* 融合而来。



注: a. *GmELF3b*; b. *GmELF3c*; c. *GmELF3Ld*; d. *GmELF3Le*。

图 10 *GmELF3* 基因组织特异性表达分析

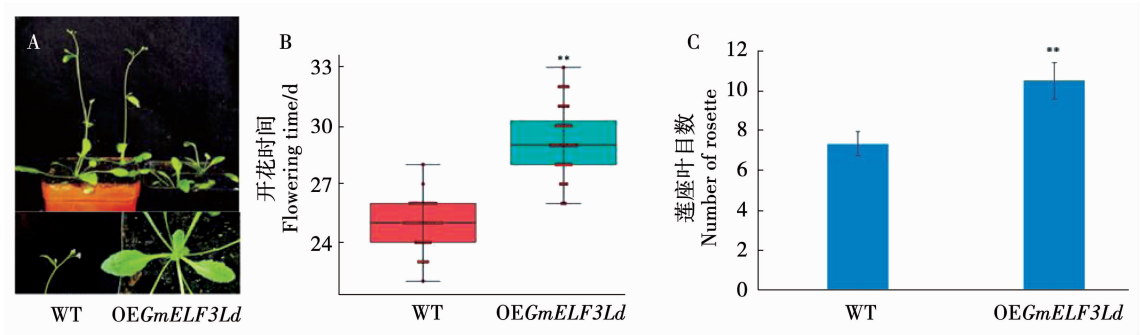
Fig. 10 The tissue expression pattern of *GmELF3* genes



## 2.5 转基因拟南芥的表型和表达分析

分析结果表明, *GmELF3Ld* 基因过表达比野生型拟南芥开花时间明显延迟(图 11A 和 B), 莲座叶数目显著增多(图 11A 和 C)。同时在转基因拟南芥中, *GmELF3Ld* 基因显著高表达, 而与拟南芥开花

相关的 *CO* 和 *FT* 基因的表达量却明显下调(图 12)。该结果与前人报道的 *GmELF3a/J* 促进拟南芥开花<sup>[15]</sup> 的功能完全相反, 说明 *GmELF3s* 在进化中发生了功能分化。



注: A. 表型观察; B. 开花时间; C. 花期莲座叶片数目。  
Note: A. Phenotypes observation; B. Flowering time; C. Rosette number.

图 11 过表达 *GmELF3Ld* 拟南芥表型分析

Fig. 11 The phenotypic analysis of *GmELF3Ld* transgenic *Arabidopsis*

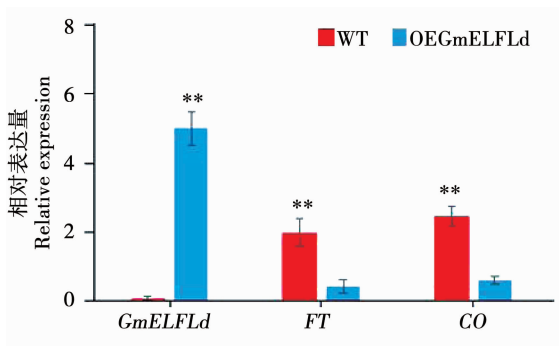


图 12 过表达 *GmELF3Ld* 拟南芥花期调控基因表达分析  
Fig. 12 The analysis of flowering time genes expression pattern in *GmELF3Ld* overexpression *Arabidopsis*

## 3 讨论

大豆对光温反应极为敏感, 短日高温下往往植株矮小、产量低下, 突变体 *Gmelf3a/j* 在短日高温条件下表现出晚花高产的表型<sup>[15]</sup>, 通过对拟南芥的研究表明 *ELF3* 是光温反应的重要调节因子, 同时也是生物节律钟的核心成员, *ELF3* 的缺失会造成拟南芥生物节律的失调<sup>[8]</sup>, 但大豆突变体 *Gmelf3a/j* 的生物节律钟依旧保持完整的节律性<sup>[15]</sup>, 这可能与同源基因 *ELF3b-Le* 的功能相关, 但这些基因的具体功能、在光温反应中的角色及在生物钟系统中的作用机制还需进一步研究。全球气候变暖对大豆种植的影响极其显著, 近年来黄淮海地区的大豆症青现象几乎造成一些产区大豆绝产, 有研究表明大豆症青的发生与花期高温有关<sup>[24]</sup>, *ELF3* 作为重要的植物温度响应基因, 其同源基因 *GmELF3s* 对高温胁迫有可能具有响应, 对 *GmELF3s* 改造和利用有可能

对大豆高温胁迫育种具有重要贡献。光照长度是大豆品种推广的重要限制性因素, *ELF3* 作为光信号的整合因子, 具有根据光质量调控生物节律的功能, 大豆 *GmELF3a/J* 基因的利用在这方面就表现出了极大的优势, *GmELF3s* 基因的进一步研究和利用将对广适性大豆品种的选育具有重要意义。

本研究发现大豆的 5 个 *ELF3* 同源基因在结构上表现出一定的分化, 根据序列分化程度将其分为 3 个聚类, 其中 B 类蛋白中 *GmELF3Ld* 与 *GmELF3Le* 的遗传距离为 0.1, A 类蛋白中 *GmELF3b* 与 *GmELF3c* 的遗传距离为 0.2, 而 *GmELF3a/J* 与 A2 和 B 类的遗传距离分别为 0.5 和 1.2, 说明这 5 个基因在进化上朝着 3 个不同的方向进行, 并分化出不同的功能。本研究中 *GmELF3Ld* 的功能与 *GmELF3a/J* 有明显的差异, 也同时印证了 *GmELF3s* 在大豆中的功能发生了分化。同样的结果出现在基因组、蛋白质二级和三级结构分析中。有趣的是, 作为古四倍体大豆, 同一家族的同源基因多为 2 的整数倍<sup>[16, 25]</sup>, 而 *ELF3* 的则为 5 个, 是由于其在进化中 *GmELF3a/J* 缺失了 1 个拷贝, 或是 *GmELF3a/J* 由其他两个 *GmELF3s* 融合而来, 可能需要进一步在野生豆中进行考证。

本研究发现 *GmELF3s* 蛋白为亲水性蛋白, 没有跨膜结构域, 且被预测定位于细胞核, 同时 *GmELF3s* 蛋白存在多个丝氨酸和酪氨酸磷酸化位点, 其中酪氨酸磷酸化位点是蛋白结合形成多蛋白复合体的前提条件。 *GmELF3a/J* 在夜晚出现表达高峰, 而拟南芥 *EC* 夜晚复合物具有调控作用<sup>[2-3, 11]</sup>, 因此推测 *GmELF3s* 的核定位信号可能在

生物节律钟系统调控中起到重要作用。本研究利用生物信息学手段详尽对比分析了各 *GmELF3s* 基因的结构,并借助转基因手段进一步分析说明 B 类蛋白 *GmELF3Ld* 的抑制开花功能与 A 类蛋白 *GmELF3a* 的促进开花功能作用相反,而引起 *GmELF3s* 功能分化及调控下游开花基因的分子机制还有待分子生物学研究进一步揭示。

4 结论

*GmELF3s* 在进化中分化成 3 类基因: A1 (*GmELF3a/J*)、A2 (*GmELF3b* 和 *GmELF3c*) 和 B (*GmELF3b* 和 *GmELF3c*)。3 类蛋白在进化过程中发生了不同程度的分化,分子量和 pI 值最大且最不稳定的是 *GmELF3Le*, 脂肪指数最大的是 *GmELF3b*。B 类基因在营养器官中高表达;而 A2 类基因在生殖器官中高表达。*GmELF3Ld* 抑制拟南芥开花,与 *GmELF3a* 促进拟南芥开花的功能完全不同。本研究说明 *GmELF3s* 在进化中发生了功能分化,研究结果为深入研究 *GmELF3s* 生物学功能及利用 *GmELF3s* 改良大豆品种提供了新思路。

参考文献

[1] HICKS K A, MILLAR A J, CARRE I A, et al. Conditional circadian dysfunction of the *Arabidopsis* early-flowering 3 mutant [J]. *Science*, 1996, 274(5288): 790-792.

[2] COVINGTON M F, PANDA S, LIU X L, et al. ELF3 modulates resetting of the circadian clock in *Arabidopsis* [J]. *Plant Cell*, 2001, 13(6): 1305-1316.

[3] THINES B, HARMON F G. Ambient temperature response establishes ELF3 as a required component of the core *Arabidopsis* circadian clock [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107 (7): 3257-3262.

[4] NUSINOW D A, HELFER A, HAMILTON E E, et al. The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth[J]. *Nature*, 2012, 475(7356): 398-402.

[5] DIXON L E, KNOX K, KOZMA-BOGNAR L, et al, Temporal repression of core circadian genes is mediated through EARLY FLOWERING 3 in *Arabidopsis* [J]. *Current Biology*, 2011, 21 (2): 120-125.

[6] CHOW B Y, HELFER A, NUSINOW D A, et al. ELF3 recruitment to the PRR9 promoter requires other Evening Complex members in the *Arabidopsis* circadian clock[J]. *Plant Signaling & Behavior*, 2012, 7 (2): 170-173.

[7] LU S X, WEBB C J, KNOWLES S M, et al. CCA1 and ELF3 interact in the control of hypocotyl length and flowering time in *Arabidopsis*[J]. *Plant Physiology*, 2012, 158(2): 1079-1088.

[8] ZAGOTTA M T, HICKS K A, JACOBS C I, et al. The *Arabidopsis* *ELF3* gene regulates vegetative photomorphogenesis and the photo-periodic induction of flowering[J]. *Plant Journal*, 2010, 10(4): 691-702.

[9] TOMOMI T, LINSEY N, ALYSSA B, et al. Light and the circadian

clock mediate time-specific changes in sensitivity to UV-B stress under light/dark cycles[J]. *Journal of Experimental Botany*, 2014, 65(20): 6003-6012.

[10] HE H, ALVAREZ S, BINDBEUTEL R K, et al. Identification of Evening Complex associated proteins in *Arabidopsis* by affinity purification and mass spectrometry [J]. *Molecular & Cellular Proteomics*, 2016, 15(1): 201-217.

[11] NIETO C, LÓPEZ-SALMERÓN V, DAVIÈRE J, et al. ELF3-PIF4 interaction regulates plant growth independently of the Evening Complex[J]. *Current Biology*, 2015, 25(2): 187-193.

[12] BOX M S, HUANG B E, DOMIJAN M, et al. ELF3 controls thermoresponsive growth in *Arabidopsis* [J]. *Current Biology*, 2015, 25(2): 194-199.

[13] TAKESHI M, YUJI N, HARUKA O, et al. Ambient temperature signal feeds into the circadian clock transcriptional circuitry through the EC night-time repressor in *Arabidopsis thaliana*[J]. *Plant and Cell Physiology*, 2014, 55(5): 958-976.

[14] ZHAO J, HUANG X, OUYANG X, et al. OsELF3-1, an ortholog of *Arabidopsis* EARLY FLOWERING 3, regulates rice circadian rhythm and photoperiodic flowering[J]. *PLoS One*, 2012, 7(8): e43705.

[15] YUE Y, LIU N, JIANG B, et al. A single nucleotide deletion in J encoding *GmELF3* confers long juvenility and is associated with adaption of tropic soybean[J]. *Molecular Plant*, 2017, 10(4): 656-658.

[16] LIU W, JIANG B, MA L, et al. Functional diversification of Flowering Locus T homologs in soybean: *GmFT1a* and *GmFT2a/5a* have opposite roles in controlling flowering and maturation [J]. *New Phytologist*, 2018, 217(3): 1335-1345.

[17] TAMURA K, PETERSON D, PETERSON N, et al. Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods [J]. *Molecular Biology and Evolution*, 2011, 28(10): 2731-2739.

[18] FLORENCE C. Multiple sequence alignment with hierarchical clustering[J]. *Nucleic Acids Research*, 1988, 16(22): 10881-10890.

[19] LUPAS A, VAN DYKE M, STOCK J. Predicting coiled coils from protein sequences[J]. *Science*, 1991, 252(5009): 1162-1164.

[20] ROY A, KUCUKURAL A, ZHANG Y. I-TASSER: A unified platform for automated protein structure and function prediction [J]. *Nature Protocols*, 2010, 5(4): 725-738.

[21] WILKINS M R, GASTEIGER E, BAIROCH A, et al. Protein identification and analysis tools in the ExPASy Server [J]. *Methods in Molecular Biology*, 1999, 112: 531-552.

[22] BLOM N, GAMMELTOFT S, BRUNAK S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites [J]. *Journal of Molecular Biology*, 1999, 294(5): 1351-1362.

[23] BRINKMAN F S L. PSORTb 3. 0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes [J]. *Bioinformatics*, 2010, 26(13): 1608-1615.

[24] 郭建秋, 马雯, 雷全奎, 等. 黄淮海夏大豆“症青”现象发生原因初步探讨[J]. *河南农业科学*, 2012, 41(4): 45-48, 53. (GUO Z, MA W, LEI Q, et al. Tentative analysis of ‘Zheng Qing’ phenomena of soybean in the Huanghuai Valleys [J]. *Journal of Henan Agricultural Science*, 2012, 41(4): 45-48, 53.)

[25] JIA Z, JIANG B, GAO X, et al. *GmFULa*, a *FRUITFULL* homolog, functions in the flowering and maturation of soybean [J]. *Plant Cell Reports*, 2015, 34(1): 121-132.