



大豆 C₂H₂ 型锌指蛋白转录因子家族全基因组鉴定及表达分析

郭智滨¹, 张亚坤¹, 贺紫航¹, 周强华¹, 蔡占东^{1,2}, 马启彬^{1,2}, 年海^{1,2}

(1. 华南农业大学 农学院/国家大豆改良中心广东分中心, 广东 广州 510642; 2. 岭南现代农业科学与技术广东省实验室, 广东 广州 510642)

摘要:为进一步寻找与大豆产量相关的潜在候选基因,利用生物信息学与转录组学技术,通过拟南芥 (*Arabidopsis thaliana* L.) C₂H₂ 型锌指蛋白 (C₂H₂ zinc finger protein) 序列 BLAST 映射,将从大豆 (*Glycine max*) 的基因组数据库中筛选出的 44 个 C₂H₂ 锌指蛋白家族成员分成 11 个亚家族,并对其进行系统进化、保守结构域、顺式作用元件、测序数据和转录组数据等相关分析。结果表明:C₂H₂ 锌指蛋白家族成员分布于 17 条大豆染色体中,且每个亚家族的基因结构域较为相近,呈高度保守性。成员启动子部分含有抗逆境胁迫以及激素相关的调控元件。RNA-seq 结果显示,有 8 个基因在不同植物组织内表达差异显著。荧光定量 PCR 分析表明,3 个基因的分析结果为 SUPERMAN 的转录调节因子, *GmC1-IiZFP42* 为 Ln 位点的相应基因,参与大豆叶形和籽粒的调控, *GmC1-IiZFP12*、*GmC1-IiZFP30* 为 SUPERMAN 基因的转录调节因子, *GmC1-IiZFP21* 可能参与大豆籽粒形成和发育的过程,为调控相关过程的候选基因。本研究为进一步寻找和挖掘大豆高产基因提供方向,也为 C₂H₂ 型锌指蛋白和生物信息学在植物中的利用提供参考和启示。

关键词:大豆; C₂H₂; 基因家族; 表达分析

Genome-wide Identification and Expression Analysis of C₂H₂ Zinc Finger Protein Gene Family in Soybean

GUO Zhibin¹, ZHANG Yakun¹, HE Zihang¹, ZHOU Qianghua¹, CAI Zhandong^{1,2}, MA Qibin^{1,2}, NIAN Hai^{1,2}

(1. College of Agriculture, South China Agricultural University/Guangdong Subcenter of National Center for Soybean Improvement, Guangzhou 510642, China; 2. Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510642, China)

Abstract: In order to further search for potential candidate genes related to soybean yield, 44 C₂H₂ zinc finger protein family members from the genome database of soybean (*Glycine max*) were classified into 11 subfamilies by BLAST mapping of *Arabidopsis thaliana* L. C₂H₂ zinc finger protein sequences using bioinformatics and transcriptomics technologies. The 44 C₂H₂ zinc finger protein family members screened from the soybean genome database were divided into 11 subfamilies, and analysed for phylogeny, conserved structural domains, *cis*-acting elements, sequencing data and transcriptome data. The results showed that the C₂H₂ zinc finger protein family members were distributed in 17 soybean chromosomes, and the structural domains of the genes in each subfamily were similar and highly conserved. The promoters of the members contained regulatory elements for stress and hormone resistance, and the RNA-seq results showed that eight genes were differentially expressed in different plant tissues. Quantitative real-time PCR analysis showed that three genes analysed were transcriptional regulators of the SUPERMAN, *GmC1-IiZFP42* was the corresponding gene in the Ln locus, which is involved in the regulation of soybean leaf shape and seed, *GmC1-IiZFP12* and *GmC1-IiZFP30* were transcriptional regulators of the SUPERMAN gene, and *GmC1-IiZFP21* may be involved in the process of soybean grain formation and development, and is a candidate gene for regulating related processes. This study provides directions for further searching and mining of soybean high yield genes, as well as references and insights for the use of C₂H₂-type zinc finger proteins and bioinformatics in plants.

Keywords: soybean; C₂H₂; gene family; expression analysis

锌指蛋白 (Zinc Finger Proteins, ZFPs) 是一类通过结合 Zn²⁺ 自我折叠、形成短而稳定的手指状结构的转录因子 (Transcription Factor, TF)。在基因表达调控、细胞分化、胚胎发育和植物抗逆等方面都发挥着重要作用^[1]。Krishna 等^[2] 根据组氨酸 (Histidine, His) 和半胱氨酸 [L (+)-cysteine, Cys] 残基围绕 Zn²⁺ 形成的不同空间结构将锌指蛋白分为 C₂H₂、C₂HC、C₂HC₅、C₃HC₄、C₃H、C₄、C₄HC₃、C₆

和 C₈ 等多种类型^[2-3]。所有锌指蛋白类型中, C₂H₂ 型锌指蛋白是植物中最大的转录因子家族之一, 其在植物叶的形态发生、花的发育以及抗逆境胁迫等方面发挥着重要作用^[4]。

C₂H₂ 型锌指蛋白属于 TF III 型, 是目前在真核生物基因组中分布和研究最为广泛的一类锌指蛋白^[5]。C₂H₂ 型锌指蛋白由约 25 ~ 30 个氨基酸组成的锌指结构, 保守氨基酸序列为 X₂-Cys-X₍₂₄₎-Cys-

收稿日期: 2023-11-06

基金项目:国家重点研发计划“农业生物种质资源挖掘与创新利用”重点专项 (2021YFD1201602); 广东省乡村振兴战略专项省级种业振兴行动项目 (006939908/2022-00744); 国家大豆产业技术体系建设项目 (CARS-04-PS09); 2021 年大学生创新创业训练计划项目 (202110564047)。

第一作者:郭智滨 (1999—), 男, 硕士研究生, 主要从事大豆遗传育种研究。E-mail: 20222015014@stu.scau.edu.cn。

通讯作者:年海 (1962—), 男, 博士, 教授, 主要从事大豆遗传育种研究。E-mail: hnian@scau.edu.cn。

X_{12} -His- $X_{(3,5)}$ -His(其中X代表任意氨基酸),序列中两个Cys和两个His围绕 Zn^{2+} 形成配位键,由1个反向平行的 β -折叠和1个 α -螺旋形成紧密的指状结构^[6-7]。相较于动物的 C_2H_2 型锌指蛋白,植物 C_2H_2 型锌指蛋白有两个特征^[8]。第一个特征是相邻两个锌指间的氨基酸数目更多,间隔较长,且长度可变;另一方面,植物 C_2H_2 型锌指蛋白在与DNA结合处有一段植物锌指蛋白特有的高度保守序列QALGGH^[8]。QALGGH序列以及序列后两个锌指之间间隔物对蛋白质与DNA靶点的亲和力很关键。这些结构特征表明,植物和其他生物识别和调控靶基因的机制可能不同^[2]。此外,根据是否具有这段特有的保守序列,Agarwal等^[9]将植物 C_2H_2 型锌指蛋白分为Q型和C型两种,在Q型锌指蛋白中,根据QALGGH序列临近的氨基酸序列的保守性又将其分为A型和B型两种^[10]。 C_2H_2 锌指蛋白还可根据其拥有的其他结构域或基序组成进行分类,如DLN-box、R-box、PL-box、L-box等特征序列,含有不同结构的锌指蛋白在功能上也呈现出显著差异^[11]。

自1992年Takatsuji等^[12]在矮牵牛(*Petunia hybrida* Vilm.)中发现了第一个植物 C_2H_2 型锌指蛋白EPF1以来,一大批 C_2H_2 型锌指蛋白相继被克隆并鉴定。拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、番茄(*Solanum lycopersicum*)、玉米(*Zea mays*)等植物中均有发现。通过生物信息学的预测分析,在拟南芥中共发现了176个 C_2H_2 型转录因子^[13],水稻中已经鉴定出189个 C_2H_2 型转录因子^[9],在高粱中共鉴定出145个 C_2H_2 型转录因子^[14],在黄瓜^[15]和马铃薯^[16]中也有 C_2H_2 型转录因子被鉴定。其广泛参与生物生长发育、抗逆等事件,并在这些方面表现出了诸多生物学功能^[17]。

大豆(*Glycine max*)是一种重要的经济作物和粮食作物。伴随着分子生物学等学科的深入探索,对于大豆基因组的研究可以进一步丰富人们对于大豆的理解和认知,进而继续向育成高产优质稳产的大豆品种的目标前进。 C_2H_2 锌指蛋白作为植物抗逆和生长发育极其重要的一类蛋白,在大豆抗寒抗旱以及耐盐等方面表现出广泛的生物学功能,具有巨大的研究价值。Yuan等^[18]通过检索数据库,鉴定出345个候选 C_2H_2 -ZFPs。Sun等^[19]通过研究发现*GmZAT4*可以增强大豆对PEG以及干旱和盐胁迫的抵抗作用。Kim等^[20]验证了大豆 C_2H_2 型锌指蛋白SCOF-1可以增强转基因植物的耐寒性。Huang等^[21]研究发现*GmZFP1*可能会在大豆的花

瓣、雄蕊和发育后期发挥作用。Feng等^[22]发现*GmZFP7*可以调控苯丙烷途径并增强大豆异黄酮含量。越来越多的研究表明 C_2H_2 型转录因子对于大豆是一类重要的基因家族。

Englbrecht等^[13]使用BLAST参照拟南芥基因组,将拟南芥 C_2H_2 型锌指蛋白分为A、B、C亚家族,不同亚家族间以不同氨基酸个数为分类指标进行进一步分类。本研究参照其分类方法,对大豆 C_2H_2 型锌指蛋白基因家族进行全面筛选,旨在挖掘与大豆产量性状相关的基因,为提高大豆产量提供理论基础,并为生育期相关的基因研究提供参考和启示作用。

1 材料与方法

1.1 材料

供试材料为本实验室育成大豆品种华夏10号。

1.2 方法

1.2.1 大豆单 C_2H_2 锌指蛋白的鉴定 以拟南芥和大豆 C_2H_2 ZFP基因的保守氨基酸序列为查询对象,参照Englbrecht等^[13]方法,在Phytozome v13数据库中,利用BLASTP程序在大豆基因组数据库中进行检索。拟南芥中单 C_2H_2 锌指基因的氨基酸序列来自拟南芥数据库(<https://www.arabidopsis.org/>),大豆的基因组序列版本为*Glycine max* Wm82. a4. v1,来源于Phytozome数据库(<https://phytozome-next.jgi.doe.gov/>)。为鉴定大豆中的单 C_2H_2 锌指基因,使用BLASTB搜索使用保守的氨基酸序列作为查询(e 值 $=1 \times 10^{-5}$)。通过MEGA-X的序列比对挖掘保守的单一 C_2H_2 锌指结构域,所有假定的单一 C_2H_2 锌指基因都通过NCBI的保守结构域数据库CDD进行保守结构域验证,具有高置信度的基因被认为是大豆的单 C_2H_2 锌指基因,并命名为*GmCl-1iZFP1*~*GmCl-1iZFP44*。

1.2.2 系统进化树的构建和共线性分析 以拟南芥和大豆的Cl-1i型ZFP的全长氨基酸序列进行系统发育分析。使用MEGA-X软件对这些基因进行多序列比对。检索序列的比对用于通过相邻连接方法构建系统进化树,Bootstrap值设为1000次,而后利用iTOL在线软件(<https://itol.embl.de/>)对系统进化树进行美化。 C_2H_2 锌指蛋白家族成员的共线性分析和绘图使用TBtools进行^[23]。

1.2.3 保守基序分析和结构表征 使用ProtParam(<http://www.expasy.org>)分析*GmCl1iZFPs*的基本物理和化学特性,包括基因位置、相对分子质量、氨

氨基酸数、分子量和理论等电点。用 WoLFPSORT (<http://www.genscript.com>) 预测亚细胞定位。使用在线软件 GSDS(<http://gsds.cbi.pku.edu.cn>) 构建大豆 C₂H₂ 锌指蛋白基因结构,包括外显子、内含子、CDS 以及 UTR 区域^[24]。通过 MEME (<http://meme-suite.org>) 确定 GmC1-1iZFPs 序列中的保守基序,最大 Motif 数量设置为 10,使用 TBtools 对基因结构以及保守基序可视化^[23]。

1.2.4 GmC1-1iZFP 基因启动子中顺式元件的分析

利用 TBtools 截取 GmC1-1iZFP 基因上游 2 000 bp 核苷酸序列,将序列提交到在线数据库 PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 进行启动子顺式作用元件的分析,并使用 TBtools 将结果可视化^[23]。

1.2.5 RNA-seq 数据分析 利用中黄 13 基因组的 Illumina RNA-seq 数据,共包括 27 个大豆不同组织和同组织不同时期的数据^[25]。采用每千个碱基的转录每百万映射读取 reads (Fragments Per Kilobase of exon model per Million mapped fragments, FPKM) 的方法对表达 reads 进行归一化处理,并对表达量数据取 log₂ (FPKM + 1) 值,使用 TBtools 软件绘制大豆各组织不同时期的热图^[23],分析大豆中各 C₂H₂ ZFP 基因在不同植物组织和时期的表达模式。

1.2.6 荧光定量 PCR 分析 选择发育正常、籽粒

饱满且相对一致的华夏 10 号大豆种子。将种子均匀播种于华南农业大学农场,田间精细管理无杂草病害。待植株生长后参照 Shen 等^[25] 研究的取样时期和取样部位,分别取花芽分化期的老叶 (leaf-2)、处于枯萎期的花 (flo-5)、种子发育 21 d 的豆荚 (pod-1) 以及结荚后 21 d 的种子 (seed-1),液氮冷冻保存。使用 TRNzol Universal (DP424, TIANGEN, 北京) 从保存的材料中提取 RNA。使用 PrimeScript™ RT reagent Kit with gDNA Eraser 试剂盒 (RR047A, Takara Bio, 日本) 合成 cDNA。

使用 TB Green™ Premix Ex Taq™ II (RR820, Takara Bio, 日本) 搭配 CFX96 Real-Time System (BioRad, Hercules, CA, 美国) 进行荧光定量 PCR (quantitative real-time PCR, qPCR)。qPCR 反应体系为 25 μL,其中包含 12.5 μL SYBR 混合物,2 μL cDNA 模板,正向和反向引物各 1 μL 以及 8.5 μL 灭菌水。qPCR 反应程序设置为:95 ℃ 30 s,95 ℃ 5 s,60 ℃ 40 s,40 个循环;95 ℃ 15 s,60 ℃ 30 s,95 ℃ 15 s。引物序列详见表 1。所有材料分别进行 3 次独立的生物重复和 3 次技术重复的扩增反应。使用 2^{-ΔΔCt} 方法分析 GmC1-1iZFPs 的相对表达水平,内参基因为大豆 β-Actin 基因^[26]。在 IBM SPSS Statistics 26 中使用 Duncan 法 (*P* < 0.05) 评估数据的统计显著性。使用 GraphPad Prism 8 处理数据并制图。

表 1 qPCR 引物序列
Table 1 Primers sequence of PCR

基因名称 Gene name	基因 ID Gene ID	正向引物序列(5'→3') Forward primer sequence(5'→3')	反向引物序列(5'→3') Reverse primer sequence(5'→3')	产物长度 Product size/bp
GmC1-1iZFP12	Glyma.03G160700	TCAAGCTTTGGGTGTCACA	CCCATGATGACGGGGGTAAG	80
GmC1-1iZFP21	Glyma.10G034600	GGAGCACAAGGGTTCCAGA	CGAGTCGAAGCTCCAAATCCA	104
GmC1-1iZFP29	Glyma.12G230200	TGCTCAGGCACTTGCTGG	GGCCAGCAGCTTGACCTT	104
GmC1-1iZFP30	Glyma.13G120900	GGGAACTCCATGGCCTGC	GGCATTGCGAAACCCACG	149
GmC1-1iZFP32	Glyma.13G269400	TGCTCAGGCACTTGCTGG	GGGCCTTGAGCTTGACCT	111
GmC1-1iZFP35	Glyma.18G036000	GTGGGGACATCATCACCACAT	AACACTGCACCTTGTCCAC	138
GmC1-1iZFP37	Glyma.18G186000	GCAGCTTGCTCTGCTTGC	GCCACCTAAAGCCTGTGC	90
GmC1-1iZFP42	Glyma.20G116200	CACCGCAACGACCATACCTA	TCCTGTGCACCTGAATAGCC	277

2 结果与分析

2.1 大豆中 GmC1-1iZFP 基因的鉴定

基于多序列比对结果,将所有结果序列进行 BLAST 映射以确认保守域。44 个具有单 C₂H₂ 锌指结构域的序列被认为是单 C₂H₂ ZFP 基因,命名为 GmC1-1iZFP1 ~ 44。GmC1-1iZFP 蛋白的长度从 86 个氨基酸 (GmC1-1iZFP40) 到 380 个氨基酸 (GmC1-

1iZFP37) 不等,蛋白的相对分子质量为 9.73 ~ 43.28 kDa,等电点为 5.23 (GmC1-1iZFP13) ~ 10.19 (GmC1-1iZFP41)。大多数 GmC1-1iZFP 基因只存在于单个区域;其中 GmC1-1iZFP41 定位于细胞核、细胞质核和过氧化物酶体中。44 个 GmC1-1iZFP 基因的染色体位置、编码序列长度、氨基酸数、等电点、相对分子质量、亚细胞定位预测等信息如表 2 所示。

表 2 大豆 C₂H₂ 锌指蛋白转录因子 C1-li 亚家族成员基本信息

Table 2 Basic information about C1-li subfamily members of soybean C₂H₂ zinc finger protein transcription factors

基因名称	基因 ID	基因位置	染色体	氨基酸数	等电点	相对分子质量	亚细胞定位
Gene name	Gene ID	Genomic positon	Chr.	Amino acids	pI	Molecular weight/Da	Subcellular location
GmC1-1iZFP1	Glyma. 01G036600	3851393. . . 3852529	1	241	7. 11	27314. 93	细胞核
GmC1-1iZFP2	Glyma. 01G101800	34486594. . . 34488302	1	313	8. 24	34884. 89	细胞核
GmC1-1iZFP3	Glyma. 01G198700	54363589. . . 54364798	1	250	6. 14	27387. 33	细胞核
GmC1-1iZFP4	Glyma. 02G029000	2621862. . . 2622341	2	159	6. 41	17348. 59	细胞核、胞外
GmC1-1iZFP5	Glyma. 02G029100	2630709. . . 2631224	2	171	9. 06	18982. 46	细胞核、细胞间质_细胞核
GmC1-1iZFP6	Glyma. 02G029400	2680290. . . 2681618	2	246	5. 81	27920. 38	细胞核
GmC1-1iZFP7	Glyma. 02G162200	21871464. . . 21872198	2	244	9. 21	27064. 69	细胞核
GmC1-1iZFP8	Glyma. 02G170366	28298349. . . 28301667	2	338	6. 83	38061. 41	细胞核
GmC1-1iZFP9	Glyma. 02G242400	44921251. . . 44924972	2	257	5. 83	28033. 03	细胞核
GmC1-1iZFP10	Glyma. 03G067700	14804814. . . 14806495	3	318	7. 62	35676. 56	细胞核
GmC1-1iZFP11	Glyma. 03G147200	37472327. . . 37473106	3	259	8. 44	28920. 51	细胞核
GmC1-1iZFP12	Glyma. 03G160700	38760715. . . 38761864	3	205	9. 16	22739. 00	细胞核
GmC1-1iZFP13	Glyma. 03G160900	38778766. . . 38781785	3	178	5. 23	20223. 20	细胞核
GmC1-1iZFP14	Glyma. 03G217900	43334657. . . 43335291	3	166	10. 08	18989. 54	细胞核、过氧化物酶
GmC1-1iZFP15	Glyma. 04G127600	16530090. . . 16530998	4	302	6. 46	33135. 58	细胞核
GmC1-1iZFP16	Glyma. 05G111400	29673454. . . 29675517	5	224	6. 29	24801. 89	细胞核
GmC1-1iZFP17	Glyma. 05G139000	33186274. . . 33191642	5	249	6. 79	26769. 78	细胞核
GmC1-1iZFP18	Glyma. 06G311500	49537870. . . 49538781	6	303	6. 12	33341. 46	细胞核
GmC1-1iZFP19	Glyma. 07G135900	15940297. . . 15941628	7	305	8. 33	34606. 76	细胞核
GmC1-1iZFP20	Glyma. 09G107400	19442809. . . 19444468	9	307	6. 76	34219. 19	细胞核
GmC1-1iZFP21	Glyma. 10G034600	2997826. . . 2998992	10	219	6. 55	24704. 47	细胞核
GmC1-1iZFP22	Glyma. 10G034700	3007118. . . 3008529	10	196	5. 73	22407. 25	细胞核
GmC1-1iZFP23	Glyma. 10G101900	18402081. . . 18402773	10	230	9. 24	25435. 67	细胞核、过氧化物酶
GmC1-1iZFP24	Glyma. 10G184500	41861988. . . 41863119	10	175	5. 57	19956. 79	细胞核
GmC1-1iZFP25	Glyma. 10G272300	49523123. . . 49525895	10	283	6. 86	29659. 74	细胞核
GmC1-1iZFP26	Glyma. 11G043200	3170373. . . 3172092	11	260	7. 12	28642. 80	细胞核、叶绿体
GmC1-1iZFP27	Glyma. 11G214300	35635042. . . 35638612	11	249	5. 93	27671. 68	细胞核
GmC1-1iZFP28	Glyma. 11G221500	36534021. . . 36535352	11	264	8. 60	29768. 41	细胞核
GmC1-1iZFP29	Glyma. 12G230200	40463407. . . 40464310	12	272	8. 80	30272. 11	细胞核
GmC1-1iZFP30	Glyma. 13G120900	22413230. . . 22414489	13	211	6. 25	23665. 52	细胞核
GmC1-1iZFP31	Glyma. 13G121000	22422000. . . 22423238	13	193	5. 43	22098. 99	细胞核
GmC1-1iZFP32	Glyma. 13G269400	36560110. . . 36560877	13	255	9. 12	28386. 91	细胞核
GmC1-1iZFP33	Glyma. 14G212000	48561071. . . 48563936	14	257	5. 44	28151. 17	细胞核
GmC1-1iZFP34	Glyma. 17G155600	13140973. . . 13142764	17	231	6. 24	25699. 91	细胞核
GmC1-1iZFP35	Glyma. 18G036000	2797251. . . 2798598	18	259	8. 59	29073. 72	细胞核
GmC1-1iZFP36	Glyma. 18G042300	3564868. . . 3568227	18	251	6. 17	27631. 69	细胞核
GmC1-1iZFP37	Glyma. 18G186000	44983687. . . 44985014	18	380	8. 79	43279. 89	细胞核
GmC1-1iZFP38	Glyma. 19G150800	41553195. . . 41554501	19	255	8. 37	28702. 44	细胞核
GmC1-1iZFP39	Glyma. 19G162500	42771039. . . 42771674	19	211	9. 19	23472. 70	细胞核
GmC1-1iZFP40	Glyma. 19G174100	43897621. . . 43898345	19	86	9. 97	9732. 20	细胞核
GmC1-1iZFP41	Glyma. 19G214600	47306542. . . 47307683	19	194	10. 19	21923. 76	细胞核、细胞间质_ 细胞核、过氧化物酶
GmC1-1iZFP42	Glyma. 20G116200	35791114. . . 35793636	20	258	8. 55	27964. 99	细胞核
GmC1-1iZFP43	Glyma. 20G117800	36015798. . . 36017141	20	281	6. 86	29383. 62	细胞核
GmC1-1iZFP44	Glyma. 20G206000	44210590. . . 44211681	20	174	5. 80	19722. 60	细胞核

注:基因 ID 来源于 Phytozome V13 数据库。
Note: The gene ID derived from Phytozome V13 database.

2.2 大豆 GmC1-liZFP 的系统进化及共线性分析

前人在拟南芥中共发现 32 个单链 C₂H₂ ZFP 基因^[13]。基于氨基酸序列的保守性,构建这 32 个拟南芥成员与新发现的 44 个大豆 C₂H₂ ZFP 的系统进化树结果如图 1A 所示。这些基因分为 11 个亚家族,分别为 C1-liAa、C1-liAb、C1-liAc、C1-liAd、C1-liAe、C1-liAf、C1-liAg、C1-liAh、C1-liBx、C1-liCa 和 C1-liCb。其中,属于 C1-liAg 亚家族的成员最多,有 21 个,包括 16 个 GmC1-liZFP 家族成员以及 5 个拟南芥 C₂H₂ 锌指蛋白家族成员;其次是 C1-liAa 亚家族,有 20 个成员,包括 12 个 GmC1-liZFP 家族成员以及 8 个拟南芥 C₂H₂ 锌指蛋白家族成员;C1-liAc、C1-liAe 以及 C1-liCb 亚家族都只有 1 个成员。不同物种中同一亚家族 GmC1-liZFP 基因的亲缘关系比其与同一物种中的其他基因的亲缘关系更近一些,证明在同一亚家族中,不同物种的 GmC1-

liZFP 具有较高的共通性。

C₂H₂ 锌指蛋白家族基因成员间的共线性关系分析结果如图 1B 所示,共发现 66 对 C₂H₂ 锌指蛋白基因存在共线性关系。其中,3 号染色体与 19 号染色体间的共线性基因最多,有 4 对,分别是 GmC1-liZFP11 和 GmC1-liZFP38、GmC1-liZFP12 和 GmC1-liZFP39、GmC1-liZFP13 和 GmC1-liZFP39 以及 GmC1-liZFP14 和 GmC1-liZFP41;10 号和 20 号染色体间有 3 对共线性基因、11 号和 18 号染色体之间有 2 对共线性基因,10 号染色体和 20 号染色体之间存在共线性关系的基因为 GmC1-liZFP24 和 GmC1-liZFP44、GmC1-liZFP25 和 GmC1-liZFP43 以及 GmC1-liZFP25 和 GmC1-liZFP42,11 号和 18 号染色体之间的为 GmC1-liZFP28 和 GmC1-liZFP35、GmC1-liZFP27 和 GmC1-liZFP36、GmC1-liZFP28 和 GmC1-liZFP35。

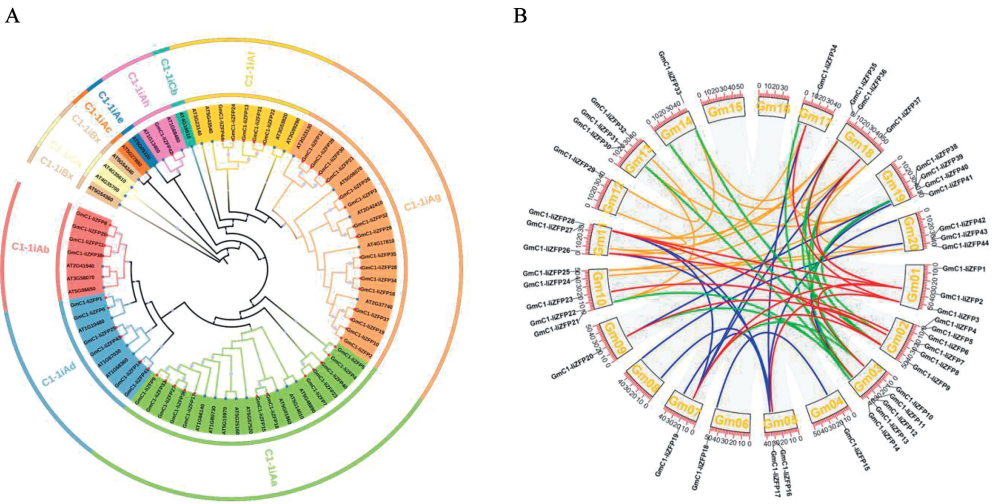


图 1 大豆 C₂H₂ 转录因子系统进化关系 (A) 及共线性分析 (B)

Fig. 1 Phylogenetic relationship (A) and collinearity analysis (B) of C₂H₂ transcription factor in soybean

2.3 GmC1-liZFP 的染色体定位和保守基序分析

利用 GmC1-liZFP 基因的核苷酸序列,分析该家族基因在染色体上的分布的结果如图 2 所示,44 个 GmC1-liZFP 较为随机地分布在大豆的 20 条染色体上。除第 8、15、16 条染色体外,每条染色体均有 GmC1-liZFP 基因。2 号染色体上 GmC1-liZFP 基因最多,有 6 个,3 号染色体和 5 号染色体上各有 5 个 GmC1-liZFP 基因。第 4、6、7、9、12、14、17 条染色体上各有 1 个,是所有家族成员中最少的。且在大豆部分染色体上存在基因簇,分别为 2 号染色体上 GmC1-liZFP4、GmC1-liZFP5、GmC1-liZFP6,3 号染色体上的 GmC1-liZFP12、GmC1-liZFP13,10 号

染色体上的 GmC1-liZFP21、GmC1-liZFP22,13 号染色体上的 GmC1-liZFP30、GmC1-liZFP31,18 号染色体上的 GmC1-liZFP35、GmC1-liZFP36 以及 20 号染色体上的 GmC1-liZFP42 和 GmC1-liZFP43。上述形成基因簇的基因可能存在相同或相似的调控作用,也有可能参与相关的生理生化过程,值得进一步挖掘和研究。

各基因的结构和基序如图 3 所示,大部分的 GmC1-liZFP 基因中没有内含子,GmC1-liZFP8、GmC1-liZFP28、GmC1-liZFP35 中各含有 1 个内含子,GmC1-liZFP31 中含有两个内含子,但在其中值得注意的是,在 GmC1-liZFP42 中共有 4 个内含子

和4个外显子。其中,所有 *GmC1-1iZFP* 都具有基序1(Moitif 1),且每个 *GmC1-1iZFP* 基因都至少含有两个保守基序,表明此个基因家族的所有成员都

具有高度保守的结构域。所有 *GmC1-1iZFP* 基因都包含基序1(Moitif 1)和基序2(Moitif 2), *GmC1-1iZFP* 家族成员的基序组合都有相近之处。

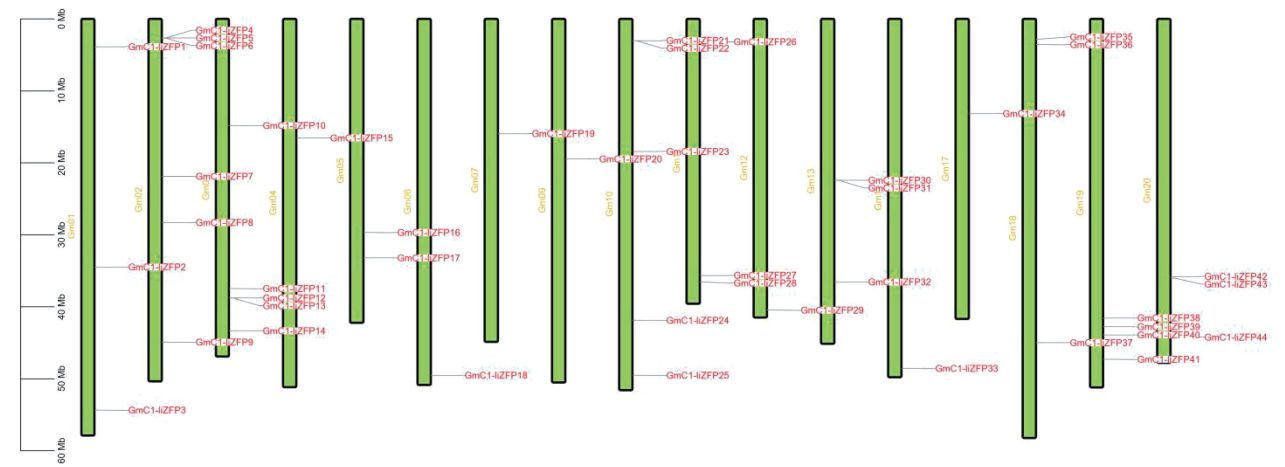
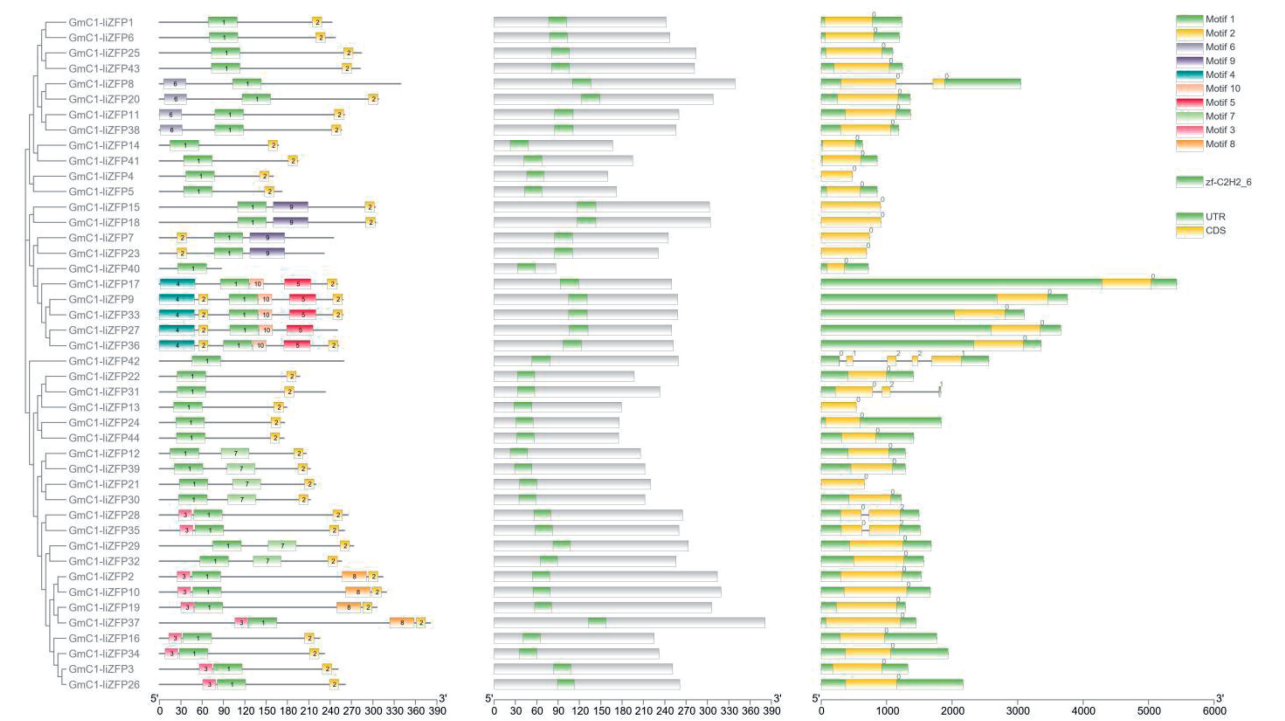


图2 大豆 C₂H₂ 锌指蛋白染色体定位

Fig. 2 Chromosome location of C₂H₂ zinc finger protein in soybean



注:最左一列展示了 *GmC1-ZFPs* 的系统进化树;第二列展示了 *GmC1-ZFPs* 结构域的保守基序组成;第三列为 *GmC1-ZFPs* 的结构域;最右侧的图片为 *GmC1-ZFPs* 的基因结构,图中灰色直线为内含子。

Note: The leftmost column shows the phylogenetic tree of *GmC1-ZFPs*; The second column shows the conserved motif composition of the *GmC1-ZFPs* domain; The third is the domains of *GmC1-ZFPs*; The image on the far right shows the gene structure of *GmC1-ZFPs*, with introns represented by gray lines.

图3 *GmC1-1iZFPs* 的结构和基序组成

Fig. 3 Gene structures and motif compositions of *GmC1-1iZFPs*

2.4 大豆 C₂H₂ 锌指蛋白家族基因启动子元件

对大豆 C₂H₂ 锌指蛋白家族的 44 个成员的 2 000 bp 启动子区域进行元件分析结果如图 4 所示,其中光响应相关元件在各基因启动子中含量最为丰富,植物体自身防御和应激反应模块在大部分启动子中表达,水杨酸(Salicylic Acid, SA)、茉莉酸(Jasmonic acid, MeJA)、赤霉素(Gibberellin Acid, GA)、生长素(Auxin, IAA)等植物激素响应模块以

及参与低温响应的反应模块在部分启动子中也有分布。结果说明该家族基因在植物体非生物胁迫以及激素响应相关的功能表达较多,在植物的光合作用、呼吸作用等生理生化途径中发挥重要作用。*GmC1-liZFP16* 和 *GmC1-liZFP23* 中含有参与玉米醇溶蛋白(Zein)代谢调节的顺式作用元件,玉米醇溶蛋白自身及水解产物具有抗氧化性,可能具有相关调控功能。

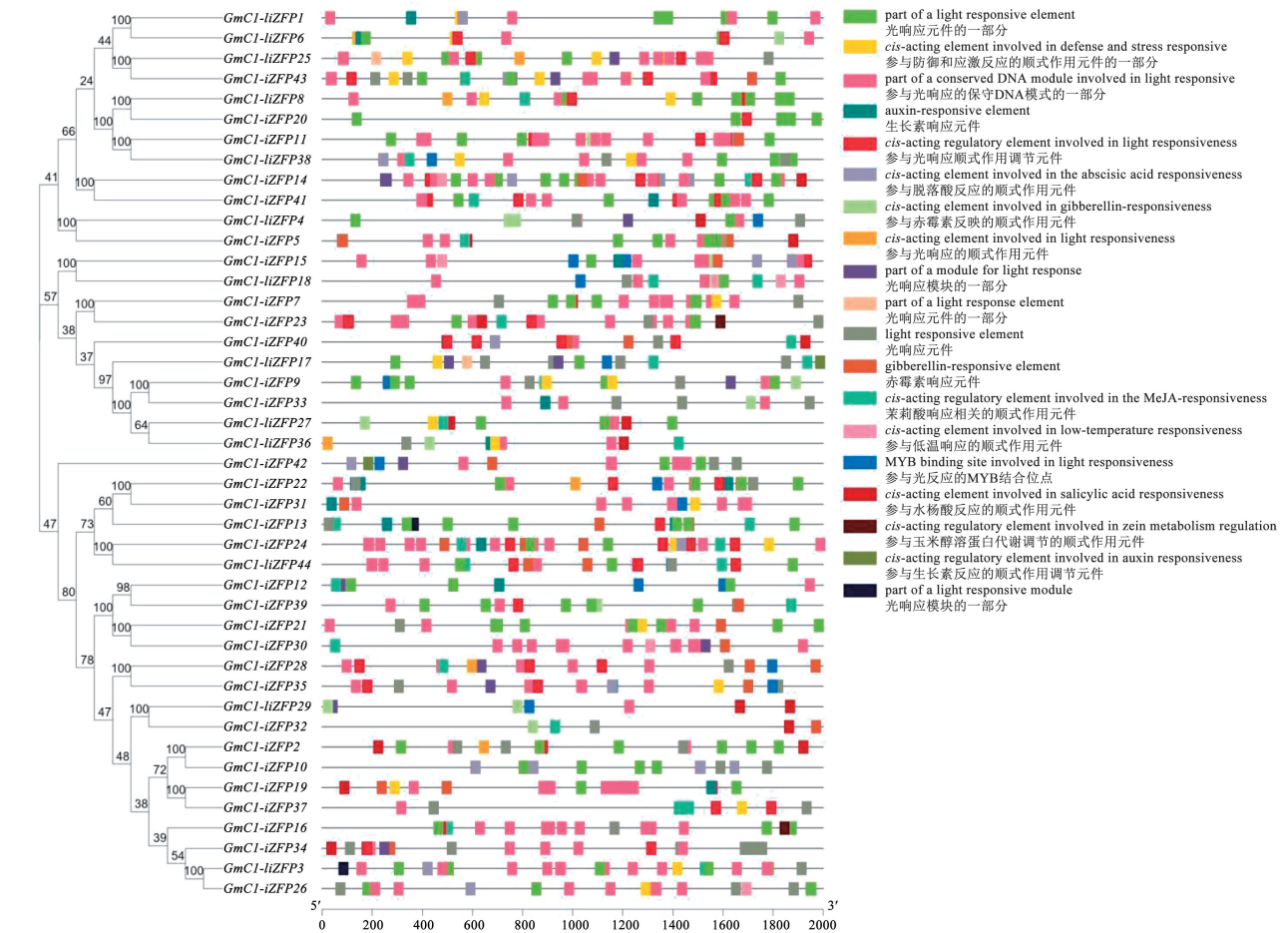


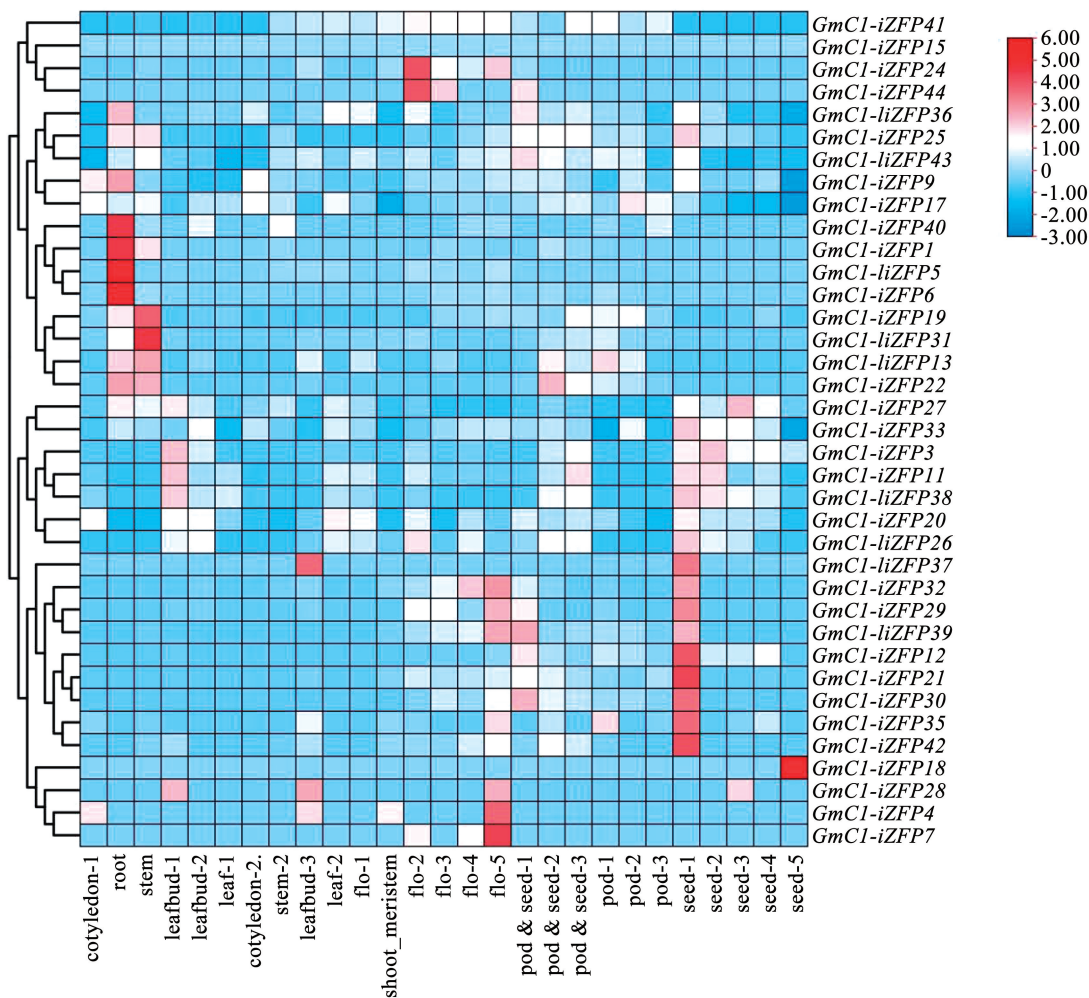
图 4 GmC1-liZFP 家族成员的启动子元件组成

Fig. 4 Composition of promoter elements of GmC1-liZFP family members

2.5 大豆 GmC1-liZFP 组织特异性表达分析

使用中黄 13 的 Illumina RNA-seq 数据^[25]对 *GmC1-liZFP* 基因在不同组织器官及不同生长发育时期 *GmC1-liZFP* 的表达量构建的热图如图 5 所示,参考 Shen 等^[25]研究获得对应不同大豆植物组织及不同发育时期的详细描述。热图中有部分基因在某些组织或时期中没有表达,推测可能是受到特定环境影响才会进行相应的表达过程。部分基因在发芽期的根、茎、枯萎期的花、发育 21 d 和发育 70 d 的种子等组织中表达显著,如 *GmC1-liZFP1*、*GmC1-liZFP5*、*GmC1-liZFP6*、*GmC1-liZFP40* 在大

豆植株根部表达量极高,说明这 4 个基因可能与大豆根的生长发育有关;*GmC1-liZFP24*、*GmC1-liZFP44* 在花蕾时期表达显著,说明可能参与了花蕾时期的调控;*GmC1-liZFP18* 在籽粒发育 70 d 时期的表达量较高;*GmC1-liZFP12*、*GmC1-liZFP21*、*GmC1-liZFP29*、*GmC1-liZFP30*、*GmC1-liZFP32*、*GmC1-liZFP35*、*GmC1-liZFP37* 和 *GmC1-liZFP42* 在籽粒发育 21 d 时期的表达显著,表明这些基因可能与大豆籽粒 21 d 时期的发育有关。这 8 个基因的描述信息如表 3 所示。



注:cotyledon-1 和 cotyledon-2 分别代表萌芽期的子叶和长出三出复叶后的衰老子叶;root 代表萌芽期的根;stem-1 和 stem-2 分别代表萌芽期的茎和三叶期的茎;leafbud-1、leafbud-2 和 leafbud-3 分别表示发芽期的新叶、三叶期的新叶以及花芽分化期的新叶;shoot_meristem 表示花芽分化期的叶芽 flo-1、flo-2、flo-3、flo-4 和 flo-5 分别表示花芽分化期的花蕾、开花前的花蕾、处于花期的花、开花期第 5 天的花以及枯萎期的花;pod & seed-1、pod & seed-2、pod & seed-3 分别表示种子发育初期 14 d 的豆荚、种子发育初期 27 d 的豆荚以及种子发育初期 28 d 的豆荚;pod-1、pod-2 和 pod-3 分别表示种子开始发育 21 d 的豆荚、发育 28 d 的豆荚以及发育 35 d 的豆荚;seed-1、seed-2、seed-3、seed-4 和 seed-5 分别表示发育 21 d 的种子、发育 35 d 的种子、发育 42 d 的种子、发育 56 d 的种子以及发育 70 d 的种子。

Note:cotyledon-1 and cotyledon-2 represent cotyledons at the budding stage and senescent cotyledons after growing ternate compound leaves, respectively; root represents roots at the budding stage; stem-1 and stem-2 represent stems at the budding stage and stems at the ternate stage, respectively; leafbud-1, leafbud-2, and leafbud-3 represent new leaves at the germination stage, new leaves at the three-leaf stage, and new leaves at the stage of flower bud differentiation, respectively; shoot_meristem denotes leaf buds at the stage of flower bud differentiation flo-1, flo-2, flo-3, flo-4, and flo-5 denote flower buds at the stage of flower bud differentiation, flower buds prior to blooming, flowers at the stage of blooming, flowers on the 5th day of blooming, and flowers at the stage of wilting, respectively; pod & seed-1, pod & seed-2, pod & seed-3, and pod & seed-4 denote pods at two weeks of initial seed development, pods at three weeks of initial seed development, and pods at four weeks of initial seed development, respectively; pod-1, pod-2 and pod-3 denote pods at 21 days of seed initiation, pods at 28 days of development, and pods at 35 days of development, respectively; Seed-1, seed-2, seed-3, seed-4 and seed-5 denote seeds 21 days into development, seeds 35 days into development, seeds 42 days into development, seeds 56 days into development, and seeds 70 days into development, respectively.

图 5 *GmC1-liZFPs* 在大豆不同组织中的表达热图

Fig. 5 Heat map of *GmC1-liZFPs* expression in different tissues of soybean

在 NCBI (National Center for Biotechnology Information)数据库中查询热图中表达显著的基因,其中 *GmC1-liZFP12*、*GmC1-liZFP19*、*GmC1-liZFP21*、*GmC1-liZFP29*、*GmC1-liZFP30*、*GmC1-liZFP32* 和 *GmC1-liZFP39* 的功能显示为 SUPERMAN 基因的转录调节因子。SUPERMAN 基因在植物的胚、花和种子发育等方面起到重要作用^[27]。推测上述相关基因可能也与大豆在花、胚和种子发育等方面相关。

GmCl-1iZFP4 属于 JAGGED 基因,参与大豆叶型和荚粒数的调控。Yuan 等^[18] 在 Phytozome 数据库中检索大豆共生关系,推断 *GmCl-1iZFP37* 可能参与结核发育形成的相关过程,本研究鉴定出的大部分

GmCl-liZFPs 与其存在共同鉴定到的基因。*GmCl-1iZFP24* 和 *GmCl-1iZFP44* 为端粒酶激活剂(TAC10)的相关基因^[28],而 *GmCl-liZFP42* 则在前人中研究中被确定为调控大豆叶形和荚粒数的 Ln 位点^[29]。

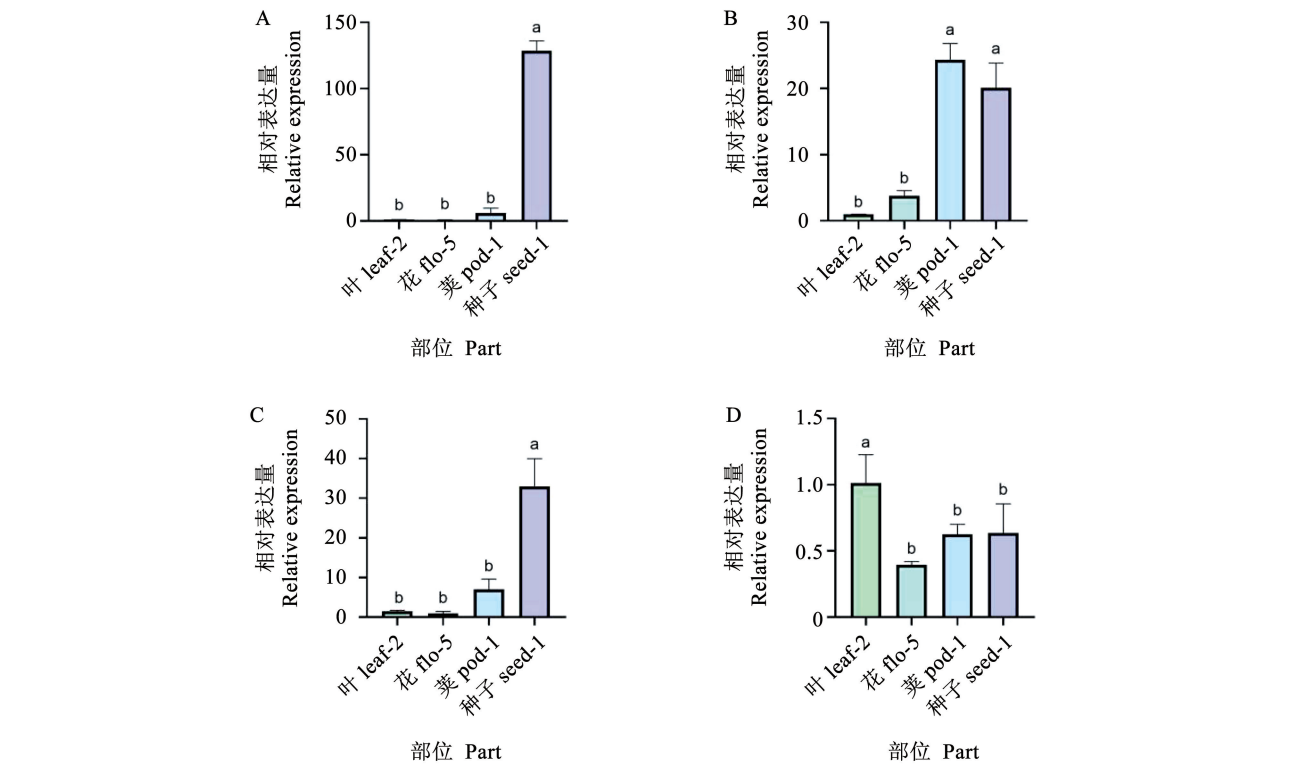
表 3 8 个高表达量基因的描述
Table 3 Description of eight expressing salient genes

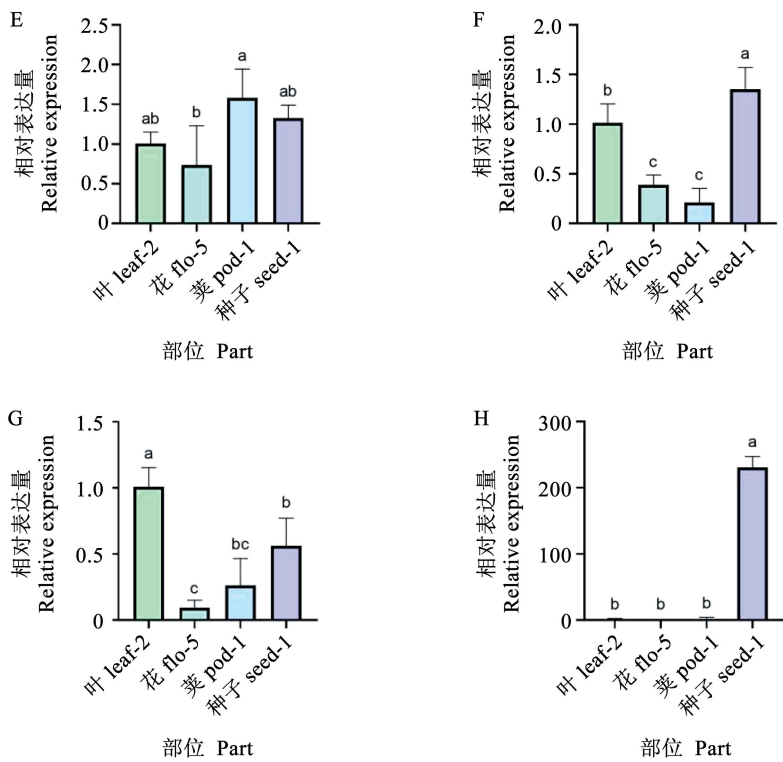
基因名称 Gene name	基因 ID Gene ID	描述 Description
<i>GmCl-1iZFP12</i>	<i>Glyma. 03G160700</i>	SUPERMAN 的转录调节因子
<i>GmCl-1iZFP21</i>	<i>Glyma. 10G034600</i>	N/A
<i>GmCl-1iZFP29</i>	<i>Glyma. 12G230200</i>	N/A
<i>GmCl-1iZFP30</i>	<i>Glyma. 13G120900</i>	SUPERMAN 的转录调节因子
<i>GmCl-1iZFP32</i>	<i>Glyma. 13G269400</i>	N/A
<i>GmCl-1iZFP35</i>	<i>Glyma. 18G036000</i>	N/A
<i>GmCl-1iZFP37</i>	<i>Glyma. 18G186000</i>	N/A
<i>GmCl-1iZFP42</i>	<i>Glyma. 20G116200</i>	Ln; 调控大豆叶形和荚粒数

2.6 荧光定量 PCR 分析

进一步对不同部位中 *GmCl-1iZFP12*、*GmCl-1iZFP21*、*GmCl-1iZFP29*、*GmCl-1iZFP30*、*GmCl-1iZFP32*、*GmCl-1iZFP35*、*GmCl-1iZFP37* 和 *GmCl-1iZFP42* 等 8 个基因的相对表达量进行分析,以花芽分化期的老叶(图中简述为叶)表达量为参照。结果如图 6 所示,*GmCl-1iZFP42* 在种子发育 21 d

时的相对表达是花芽分化期老叶的 165.4 倍,是荚发育 21 d 时表达量的 107.1 倍,是花枯萎期表达量的 635.9 倍,对籽粒相关的基因的调控作用显著($P < 0.05$)。*GmCl-1iZFP30* 在种子发育形成 21 d 时相对表达量是花芽分化期老叶中表达量的 22.45 倍,呈显著($P < 0.05$),表明该基因可能参与大豆籽粒发育和形成阶段。





注:A ~ H 分别为 *GmC1-liZFP12*、*GmC1-liZFP21*、*GmC1-liZFP30*、*GmC1-liZFP29*、*GmC1-liZFP32*、*GmC1-liZFP35*、*GmC1-liZFP37* 和 *GmC1-liZFP42*;不同字母表示不同组织材料之间的差异显著性($P < 0.05$),数据以 3 次生物重复的平均值 \pm 标准差表示。

Note: A-H indicated *GmC1-liZFP12*, *GmC1-liZFP21*, *GmC1-liZFP30*, *GmC1-liZFP29*, *GmC1-liZFP32*, *GmC1-liZFP35*, *GmC1-liZFP37* and *GmC1-liZFP42*; Different letters indicate the significance of differences between different tissue materials($P < 0.05$), data were expressed as mean \pm standard deviation of three biological replicates.

图 6 部分大豆组织中 *GmC1-liZFP* 基因的相对表达量

Fig. 6 Relative expression of *GmC1-liZFP* gene in some soybean tissues

GmC1-liZFP12 在籽粒发育 21 d 时期的表达量是花芽分化时期老叶表达量的 128.59 倍,相关性显著,在荚发育 21 d 时期的表达量是花芽分化期的老叶表达量的 5.98 倍,证明该基因极有可能与大豆籽粒形成的调控过程相关,可能参与了大豆荚形成的生理过程。*GmC1-liZFP21* 在荚发育 21 d 时期的表达量是花芽分化时期老叶时期的 24.3 倍,籽粒发育 21 d 时期表达量是花芽分化期的老叶时期的 20.13 倍,相关性均显著,花枯萎时期的表达量是花芽分化时期老叶的 3.75 倍。结果证明该基因可能同时参与了大豆荚形成和籽粒发育形成两个过程,并参与花期调控,可以继续通过研究来进行验证。

3 讨论

大豆是我国重要的粮食作物和经济作物,是关系到我国饲料供给以及食品粮食安全的一大重要作物^[30]。从各个方面和角度来提升大豆的产量、品质,保证供给需求的稳定并减少对进口大豆的依赖,是我们目前亟需攻克的重要课题。锌指蛋白是

植物中最大的转录因子家族之一,而 C_2H_2 型锌指蛋白作为目前研究较为广泛的锌指蛋白,在植物的生长发育阶段以及抵抗非生物胁迫所带来的影响方面表现出优秀的研究和应用前景^[31-33]。植物 C_2H_2 型锌指蛋白具有一段特殊的保守氨基酸结构域 QALGGH,以及比动物 C_2H_2 型锌指蛋白更长的间隔,表明植物和动物锌指蛋白间调控机制存在不同^[13]。

本研究分析了大豆 *GmC1-liZFP* 亚家族的系统进化和共线性关系,构建了系统进化树,并将筛选出的 C_2H_2 型锌指蛋白基因分为 11 个亚家族,每个亚家族中的转录因子在序列和结构域类型方面都比较一致,说明 C_2H_2 型锌指蛋白在进化中高度保守。共线性分析结果表明这些转录因子之间一共存在 67 对共线性关系,表明该基因家族的成员关系较为紧密。染色体定位和保守基序分析,除第 8、15、16 条染色体上没有 *GmC1-liZFP* 基因分布外,其他的染色体上均有基因分布,大部分基因在染色体上的位置相对聚集,少量基因位置分散,说明该

基因家族成员具有一定的保守性。通过保守基序分析发现,大部分的基因中不存在内含子,比较特殊的是 *GmC1-liZFP42*,一共有 4 个内含子和 4 个外显子,各个亚家族所含有的 Motif 略有不同,说明这 44 个基因中的每个亚家族的保守性都很高。

同时分析了其启动子元件,以寻求其顺式作用元件功能的异同,所有转录因子的启动子元件中都含有与光响应相关的元件,植物应激反应和自身防御模块在大部分启动子序列中也存在,部分转录因子中含有一些植物激素的响应作用元件,表明 C₂H₂ 型锌指蛋白在植物抵抗逆境胁迫和激素调控方面发挥重要作用。

使用 RNA-seq 数据绘制该亚家族基因的特异性表达热图发现, *GmC1-liZFP12*、*GmC1-liZFP19*、*GmC1-liZFP21* 等 8 个基因在大豆籽粒形成 21 d 时期的材料表达显著。其中, *GmC1-liZFP42* (Glyma. 20G116200) 已被前人研究发是调控大豆叶型和荚粒数的 Ln 位点^[29]。种植材料并选取大豆 4 个植物组织进行 qPCR 分析, qPCR 结果显示 *GmC1-liZFP42* 在籽粒形成 21 d 时期的相对表达量极高,表明该基因可能参与了大豆荚粒数相关的调控,这与 Cai 等^[34] 的研究结果相一致, *GmC1-liZFP12*、*GmC1-liZFP30* 属于 SUPERMAN 基因的转录调节因子,可能参与了大豆胚、花和籽粒形成等的相关过程。 *GmC1-liZFP12* 在籽粒形成 21 d 时期的相对表达量相较于其他基因表现出了较高的水平,很有可能参与大豆花和籽粒发育形成阶段, *GmC1-liZFP30* 在此时期也有较为显著的表达量,可以通过实验进一步确定其功能; *GmC1-liZFP21* 在荚形成 21 d 和籽粒形成 21 d 两个时期的相对表达都较为显著,在大豆花的枯萎期也有一定的表达量,表明此基因有可能同时参与调控大豆荚和籽粒的发育,对于大豆花的发育可能也具有一定调控作用。以上两个基因在大豆花、荚、籽粒的表达均暗示其功能可能与大豆产量相关,值得深入研究。

4 结论

本研究以生物信息学方法为手段,利用前人的相关研究和数据库的已知数据,对大豆中还没有被研究但可能存在功能的基因家族进行了较为系统的分析。结合已知的拟南芥 C₂H₂ 型锌指蛋白分类信息对大豆 C₂H₂ 型锌指蛋白进行分类并构建系统进化树,命名为 *GmC1-liZFPs*。结合 TBtools 等工具分析了该家族成员的启动子结构域以及保守性。通过转录组测序数据对大豆 *GmC1-liZFP* 基因家族

在不同组织和时期的表达进行了分析,并通过 qPCR 研究发现 *GmC1-liZFP12* 和 *GmC1-liZFP30* 在大豆籽粒形成 21 d 时期的表达量较高,将以上两个基因列为与大豆籽粒形成和发育调控相关的候选基因。 *GmC1-liZFP21* 在荚形成 21 d 时期的表达同样较高,可能参与荚形成或发育相关的调控,将其列为与大豆荚发育相关的候选基因。研究结果对大豆产量相关基因的研究奠定基础,也为大豆 C₂H₂ 锌指蛋白相关基因的研究提供参考信息。

参考文献

[1] 赵楠, 赵飞, 李玉花. 锌指蛋白结构及功能研究进展[J]. 生物技术通讯, 2009, 20(1): 131-134. (ZHAO N, ZHAO F, LI Y H. Advances in research on zinc finger protein[J]. Letters in Biotechnology, 2009, 20(1): 131-134.)

[2] KIELBOWICZ-MATUK A. Involvement of plant C(2)H(2)-type zinc finger transcription factors in stress responses [J]. Plant Science: an International Journal of Experimental Plant Biology, 2012, 185/186: 78-85.

[3] MOORE M, ULLMAN C. Recent developments in the engineering of zinc finger proteins[J]. Briefings in Functional Genomics & Proteomics, 2003, 1(4): 342-355.

[4] LIAO X, WANG L, ZHU S, et al. Identification, genomic organization, and expression profiles of single C₂H₂ zinc finger transcription factors in tomato (*Solanum lycopersicum*) [J]. Journal of Applied Genetics, 2021, 62(1): 1-15.

[5] 宋冰, 洪洋, 王武, 等. 植物 C₂H₂ 型锌指蛋白的研究进展 [J]. 基因组学与应用生物学, 2010, 29(6): 1133-1141. (SONG B, HONG Y, WANG W, et al. Advances on plant C₂H₂-type zinc finger protein [J]. Genomics and Applied Biology, 2010, 29(6): 1133-1141.)

[6] 黄骥, 王建飞, 张红生. 植物 C₂H₂ 型锌指蛋白的结构与功能 [J]. 遗传, 2004, 26(3): 414-418. (HUANG J, WANG J F, ZHANG H S. Structure and function of plant C₂H₂ zinc finger protein[J]. Hereditas(Beijing), 2004, 26(3): 414-418.)

[7] PABO C O, PEISACH E, GRANT R A. Design and selection of novel Cys2His2 zinc finger proteins [J]. Annual Review of Biochemistry, 2001, 70: 313-340.

[8] TAKATSUJI H. Zinc-finger proteins: The classical zinc finger emerges in contemporary plant science[J]. Plant Molecular Biology, 1999, 39(6): 1073-1078.

[9] AGARWAL P, ARORA R, RAY S, et al. Genome-wide identification of C₂H₂ zinc-finger gene family in rice and their phylogeny and expression analysis[J]. Plant Molecular Biology, 2007, 65(4): 467-485.

[10] 张佳, 刘俊芳, 赵婷婷, 等. 植物 C₂H₂ 型锌指蛋白研究进展 [J]. 分子植物育种, 2018, 16(2): 427-433. (ZHANG J, LIU J F, ZHAO T T, et al. Research progress of C₂H₂ zinc finger protein in plant [J]. Molecular Plant Breeding, 2018, 16(2): 427-433.)

[11] LI Y, SUN A, WU Q, et al. Comprehensive genomic survey,

structural classification and expression analysis of C₂H₂-type zinc finger factor in wheat (*Triticum aestivum* L.) [J]. BMC Plant Biology, 2021, 21(1): 380.

[12] TAKATSUJI H, MORI M, BENFEY P N, et al. Characterization of a zinc finger DNA-binding protein expressed specifically in petunia petals and seedlings[J]. The EMBO Journal, 1992, 11(1): 241-249.

[13] ENGLBRECHT C C, SCHOOFF H, BÖHM S. Conservation, diversification and expansion of C₂H₂ zinc finger proteins in the *Arabidopsis thaliana* genome[J]. BMC Genomics, 2004, 5(1): 39.

[14] CUI H, CHEN J, LIU M, et al. Genome-wide analysis of C₂H₂ zinc finger gene family and its response to cold and drought stress in *Sorghum* [*Sorghum bicolor* (L.) moench] [J]. International Journal of Molecular Sciences, 2022, 23(10): 5571.

[15] CHEN Y, WANG G, PAN J, et al. Comprehensive genomic analysis and expression profiling of the C₂H₂ zinc finger protein family under abiotic stresses in cucumber (*Cucumis sativus* L.) [J]. Genes, 2020, 11(2): 171.

[16] LIU Z, COULTER J A, LI Y, et al. Genome-wide identification and analysis of the Q-type C₂H₂ gene family in potato (*Solanum tuberosum* L.) [J]. International Journal of Biological Macromolecules, 2020, 153: 327-340.

[17] 唐玉萍, 彭凡嘉, 郭莉莉, 等. 陆地棉 C2H2 型锌指蛋白全基因组鉴定及表达分析[J/OL]. 分子植物育种, 2022: 1-23 [2023-07-06]. (TANG Y P, PEGN F J, GUO L L, et al. Genome-wide identification and stress expression analysis of C₂H₂ type zinc finger protein family in *Gossypium hirsutum* [J/OL]. Molecular Plant Breeding, 2022: 1-23 [2023-07-06].)

[18] YUAN S, LI X, LI R, et al. Genome-wide identification and classification of soybean C₂H₂ zinc finger proteins and their expression analysis in legume-Rhizobium symbiosis[J]. Frontiers in Microbiology, 2018, 9: 126.

[19] SUN Z, LIU R, GUO B, et al. Ectopic expression of GmZAT4, a putative C₂H₂-type zinc finger protein, enhances PEG and NaCl stress tolerances in *Arabidopsis thaliana* [J]. 3 Biotech, 2019, 9(5): 166.

[20] KIM J C, LEE S H, CHEONG Y H, et al. A novel cold-inducible zinc finger protein from soybean, SCOF-1, enhances cold tolerance in transgenic plants[J]. The Plant Journal; for Cell and Molecular Biology, 2001, 25(3): 247-259.

[21] HUANG F, CHI Y, MENG Q, et al. GmZFP1 encoding a single zinc finger protein is expressed with enhancement in reproductive organs and late seed development in soybean (*Glycine max*) [J]. Molecular Biology Reports, 2006, 33(4): 279-285.

[22] FENG Y, ZHANG S, LI J, et al. Dual-function C₂H₂-type zinc-finger transcription factor GmZFP7 contributes to isoflavone accumulation in soybean [J]. The New Phytologist, 2023, 237(5): 1794-1809.

[23] CHEN C, 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.

[24] 郭安源, 朱其慧, 陈新, 等. GSDS: 基因结构显示系统[J]. 遗传, 2007, 29(8): 1023-1026. (GUO A Y, ZHU Q H, CHEN X, et al. GSDS: A gene structure display server[J]. Hereditas, 2007, 29(8): 1023-1026.)

[25] SHEN Y, DU H, LIU Y, et al. Update soybean Zhonghuang 13 genome to a golden reference[J]. Science China Life Sciences, 2019, 62(9): 1257-1260.

[26] LIVAK, K J, T D SCHMITTGEN. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} method[J]. Methods (San Diego, Calif.), 2001, 25(4): 402-408.

[27] BOWMAN J L, SAKAI H, JACK T, et al. SUPERMAN, a regulator of floral homeotic genes in *Arabidopsis* [J]. Development, 1992, 114(3): 599-615.

[28] SUN S, YI C, MA J, et al. Analysis of spatio-temporal transcriptome profiles of soybean (*Glycine max*) tissues during early seed development [J]. International Journal of Molecular Sciences, 2020, 21(20): 7603.

[29] JEONG N, SUH S J, KIM M H, et al. Ln is a key regulator of leaflet shape and number of seeds per pod in soybean [J]. The Plant Cell, 2013, 24(12): 4807-4818.

[30] 李若云. 新时期国家粮食安全战略的思考 [J]. 农业经济, 2014(9): 3-6. (LI R Y. Thoughts on national food security strategy in the new period [J]. Agricultural Economy, 2014(9): 3-6.)

[31] SUGANO S, KAMINAKA H, RYBKA Z, et al. Stress-responsive zinc finger gene ZPT2-3 plays a role in drought tolerance in petunia [J]. The Plant Journal; for Cell and Molecular Biology, 2003, 36(6): 830-841.

[32] VAN DER KROL A R, VAN POECKE R M P, VORST O F J, et al. Developmental and wound-, cold-, desiccation-, ultraviolet-B-stress-induced modulations in the expression of the Petunia zinc finger transcription factor GeneZPT2-2 [J]. Plant Physiology, 1999, 121(4): 1153-1162.

[33] ZHANG S, LIU J, ZHONG G, et al. Genome-wide identification and expression patterns of the C₂H₂-zinc finger gene family related to stress responses and catechins accumulation in *Camellia sinensis* [L.] O. kuntze [J]. International Journal of Molecular Sciences, 2021, 22(8): 4197.

[34] CAI Z, XIAN P, CHENG Y, et al. CRISPR/Cas9-mediated gene editing of GmJAGGED1 increased yield in the low-latitude soybean variety Huachun 6 [J]. Plant Biotechnology Journal, 2021, 19(10): 1898-1900.