

当归 MADS-box 生物信息学及 SOC1 克隆与表达分析

崔秀文¹, 刘 迪¹, 黄天苗¹, 李美玲¹, 栗孟飞^{1*}, 魏建和^{2*}

1. 甘肃农业大学生命科学技术学院 干旱生境作物学国家重点实验室, 甘肃 兰州 730070

2. 中国医学科学院北京协和医学院 药用植物研究所, 北京 100193

摘要:目的 对当归 *Angelica sinensis* MADS-box 基因家族进行筛选与生物信息学分析, 并对 *SOC1-4* 基因进行克隆与表达验证。方法 基于当归全长转录组筛选 MADS-box 基因家族, 利用在线工具进行生物信息学分析, 并利用 qRT-PCR 检测 *SOC1-4* 基因在不同材料中的表达水平。结果 当归全长转录组中含有 29 个 MADS-box 基因家族成员, 可分为 10 个亚家族, 共含有 6 个保守基序, 蛋白质序列长度为 49~422 aa, 相对分子质量为 5 697.56~49 624.90, 等电点为 5.06~11.00, 亚细胞结果显示当归 MADS-box 基因家族成员分别定位于叶绿体、细胞核、细胞质和线粒体, 二级结构预测及 3D 建模显示同一亚家族结构相似; *SOC1-4* 基因克隆片段与全长转录组测序结果一致; *SOC1-4* 基因随种苗春化作用和植株发育时期延长呈现高表达, 在抽薹植株、茎和叶中高表达, 而在冷冻规避春化作用种苗中呈现低表达。结论 当归含有 29 个 MADS-box 成员, 各成员理化性质和结构存在一定的差异, *SOC1-4* 基因表达水平与当归抽薹开花生理调控一致, 为当归 MADS-box 基因家族调控抽薹开花相关研究提供借鉴。

关键词: 当归; MADS-box; 生物信息学; *SOC1-4*; 基因克隆; 抽薹开花

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)05-1551-10

DOI: 10.7501/j.issn.0253-2670.2023.05.022

Analysis on MADS-box bioinformatics, *SOC1* clone and expression pattern in *Angelica sinensis*

CUI Xiu-wen¹, LIU Di¹, HUANG Tian-miao¹, LI Mei-ling¹, LI Meng-fei¹, WEI Jian-he²

1. State Key Laboratory of Aridland Crop Science, College of Life Science and Technology, Gansu Agricultural University, Lanzhou 730070, China

2. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China

Abstract: Objective To screen and analyze MADS-box gene family bioinformatics of *Angelica sinensis*, clone the *SOC1-4* gene and perform bioinformatic analysis. **Methods** The MADS-box gene family was screened based on the full-length transcriptomes of *A. sinensis*, the bioinformatics was analyzed using the online tools, and the expression levels of *SOC1-4* gene in different materials were validated by qRT-PCR. **Results** A total of 29 MADS-box gene family members were identified from *A. sinensis*, with 10 subfamily classified and six motifs included; the sequence length of MADS-box proteins ranged from 49 to 422 aa, the relative molecular mass ranges from 5 697.56 to 49 624.90, and the isoelectric point ranged from 5.06 to 11.00; the sub-cellular results showed that the members of the *A. sinensis* MADS-box gene family was located in the chloroplast, nucleus, cytoplasm and mitochondria; and secondary structure prediction and 3D modeling indicated that there was a similar structure for the same subfamily. The cloned fragment of the *SOC1-4* gene was consistent with the full-length transcript. The *SOC1-4* gene were over-expressed with the extension of seedlings vernalization and plants development, as well as in bolted plants, stems and leaves, respectively; while down-regulated in the seedlings with frozen-avoided vernalization. **Conclusion** There are 29 MADS-box gene family members in *A. sinensis*, and certain differences exist in physical and chemical properties and structure among the 29 members; the expression levels of *SOC1-4* gene are in accordance with the physiological regulation of bolting and flowering of *A. sinensis*, The

收稿日期: 2022-08-09

基金项目: 国家自然科学基金项目 (32160083); 干旱生境作物学国家重点实验室 (甘肃农业大学) 基金项目 (GSCS-2021-Z03); 道地药材生态种植与质量保障项目 (202103003); 财政部和农业农村部: 国家现代农业产业技术体系 (CARS-21)

作者简介: 崔秀文 (1996—), 女, 甘肃靖远人, 在读硕士研究生, 主要从事当归抽薹开花调控方面研究。E-mail: xiuwen2021@163.com

***通信作者:** 栗孟飞 (1980—), 男, 河南驻马店人, 博士, 教授, 主要从事药用植物学研究。E-mail: lmf@gsau.edu.cn

魏建和 (1970—), 男, 福建建阳人, 博士, 教授, 主要从事药用植物栽培、分子育种和次生代谢产物调控研究。

E-mail: jhwei@implad.ac.cn

results will provide theoretical basis of MADS-box gene family for regulating bolting and flowering of *A. sinensis*.

Key words: *Angelica sinensis* (Oliv.) Diels; MADS-box; bioinformatics; *SOC1-4*; gene clone; bolting and flowering

当归 *Angelica sinensis* (Oliv.) Diels 为伞形科多年生草本植物,干燥根是我国传统中药材,素有“十方九归”之称,具有补血活血、调经止痛、润肠通便等功效^[1]。现代药理学研究表明,当归根中的有机酸类、挥发油类和多糖类等化学成分在消炎、抗癌和治疗心脑血管疾病等方面具有显著效果^[2]。目前,当归年需求量超 3 万 t,年种植面积达 4.35 万 hm²^[3-4]。然而,当归在第 2 年肉质根成药过程中,出现高达 50% 植株早薹开花的现象,导致肉质根木质化,药用有效成分含量降低且大量减产,药农经济收入减少,影响当归种质资源及产业发展^[5-8]。

MADS-box 基因家族是一类广泛存在于真核生物中的转录因子,在植物花器官分化、开花时间调节、以及果实发育等方面起到重要的调控作用^[9];其通常分为 I 型和 II 型,其中,II 型有 4 个保守结构域: M-domain (MADS-domain)、I-domain (intervening domain)、K-domain (keratin-like domain) 和 C-domain (C-terminal domain)^[10]。目前,在模式植物拟南芥中发现了 *deficiens* (DEF)、*SQUAMOSA* (SQUA) 和 *TOMATO gene 3* (TM3-like) 等 12 个 MADS-box 亚家族^[11]。前期在当归转录组学的研究中发现, MADS-box 亚家族基因,如 *SOC1* (suppressor of overexpression of *constans1*)、*FLC* (flowering locus C) 和 *AGL26* (agamouslike 26) 等,在抽薹开花过程中差异表达^[12-16]。

研究证实, *SOC1* 基因是花器官分生组织形成过程中的关键基因,其通过整合外界环境和内在因素等各种成花途径信号,激活下游花器官发育所需的基因,如 *LEAFY* (*LFY*)、*APETALA 1* (*AP1*) 和 *AGAMOUS* (*AG*),进而促进植株开花^[17]。前期当归研究中发现, *SOC1* 基因在完成春化作用的种苗根茎顶端分生组织以及早薹植株中高表达^[12-13],其上游基因 [如 *COL16* (*constans-like*)] 和下游基因 (*AGL62*) 在抽薹开花过程中也呈现高表达^[14]。综合以上研究表明,尽管前人及本课题组已通过转录组测序获得了当归抽薹开花相关的 MADS-box 基因家族序列,并对关键基因 (如 *SOC1*、*AG* 和 *VRN1*) 的表达水平进行了 qRT-PCR 检测与分析^[12-16, 18],然而,针对当归 MADS-box 基因家族以及关键基因

SOC1 的生物信息学分析等方面的系统研究尚未有报道。因此,本研究基于前期当归全长转录组测序结果,开展了 MADS-box 基因家族生物信息学系统分析、*SOC1-4* 基因克隆及表达验证,旨在深入揭示 MADS-box 基因家族的生物学功能,为有效抑制抽薹开花提供理论基础。

1 材料与仪器

1.1 材料

当归 MADS-box 编码序列来源于: (1) 不同品种 (岷归 1 号和岷归 2 号) 两年生大田植株叶片和叶柄的全长转录组,结果已提交 NCBI (Access: PRJNA782300),样品采集等详细信息见 Zhu 等^[18]; (2) 岷归 1 号种苗春化作用过程中 [T1 (0 °C、14 d; 未通过春化)、T2 (0 °C、60 d; 通过春化) 和 T3 (-3 °C、125 d; 低温规避春化)] 根茎顶端分生组织的全长转录组,结果已提交 NCBI (Access: PRJNA789039),样品采集等详细信息见 Luo 等^[12],样品原植物均由甘肃农业大学栗孟飞教授鉴定为当归 *A. sinensis* (Oliv.) Diels。

1.2 仪器

台式高速离心机 (德国 SORVAL 公司); ABI QuantStudio 5 实时荧光定量 PCR 仪 (美国 ABI 公司); 超微量分光光度计 (上海宝予德科学仪器有限公司)。

2 方法

2.1 MADS-box 家族鉴定、理化性质及结构分析

蛋白质序列 (包括氨基酸长度、相对分子质量和理论等电点) 分析利用 ExPASy (<https://web.expasy.org/protparam/>); 亚细胞定位利用 Cell-PLoc (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>); 蛋白质二级结构预测利用 PRABI-Gerland (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=NPSA/npsa_sopma.html); 蛋白质 3D 建模利用 Phyre 2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi>)。

2.2 MADS-box 家族系统发育树构建

在 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov>) 选择模式植物拟南芥 *Arabidopsis thaliana* L. 和伞形科植物黄胡萝卜 *Daucus carota* subsp. *sativus* (Hoffm.) 中与当归 29 个 MADS-box 基因家

族置信度较高的 116 个蛋白质, 利用 MEGA7.0 软件最大似然法 (maximum likelihood estimate, MLE) 构建系统进化树 (重复次数 1000 次, 其他参数为默认值)。蛋白质保守基序分析利用 MEME (<https://memesuite.org/meme/tools/meme>), 并利用 TBtools 进行可视化。利用 DNAMAN 软件进行蛋白质多序列比对 (深蓝色表示相似度 100%、粉色 > 75%、浅蓝色 > 50%)。

2.3 当归 *SOC1-4* 基因克隆

以岷归 1 号温室栽培植株功能叶片为材料 (“1.1” 项中 T2: 0 °C、60 d 通过春化, 种苗移栽生长 40 d), 种苗种植及生长环境等详细信息见 Liu 等^[19]。总 RNA 提取使用 Plant RNA Kit R6827 (Omega Bio-Tek, Norcross, GA, United States), 其纯度和浓度检测使用微量分光光度计 (Micro Drop, 上海宝予德科学仪器有限公司); 反转录使用 First-Strand cDNA Synthesis SuperMix (北京百泰克生物技术有限公司) 得到 cDNA; 利用 NCBI Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) 设计 *SOC1-4* 基因引物序列, forward: 5'-TGAGGGGAAAGACTCAGA-3' 和 reverse: 5'-CTGTTTCGACATCGGAAT-3'; 扩增产物利用 1%TAE 琼脂糖凝胶电泳进行检测; 胶回收使用琼脂糖凝胶纯化试剂盒 TIANGel Midi Purification Kit (天根生化科技有限公司), 具体反应体系及条件见说明书。基因克隆使用平末端克隆试剂盒 pHANDY®-Blunt Cloning Kit (哈尔滨博健生物科技有限公司); 引物合成及阳性克隆测序由兰州天启基因生物科技有限公司完成。

2.4 当归 *SOC1-4* 基因表达分析

基因表达材料: (1) 岷归 1 号两年生大田种植早薹 (early bolting, EB) 和非早薹 (un-early bolting, Un-EB) 植株叶和侧根 (混合 1:1), 植株生长环境和样品采集等详细信息见 Li 等^[13]; (2) 岷归 1 号三年生不同生长发育期 (S1: 营养生长期、S2: 营养生长到生殖生长过渡期、S3: 抽薹初期、S4: 抽薹伸长期) 植株叶和侧根 (混合 1:1), 植株生长环境和样品采集等详细信息见 Li 等^[14]的实验; (3) 岷归 1 号不同春化期 (T1、T2 和 T3) 种苗根茎顶端分生组织; (4) 岷归 1 号温室栽培植株 (“1.1” 项中 T2: 0 °C, 60 d 通过春化, 种苗移栽生长 60 d) 不同器官 (根、茎、叶)。

利用 NCBI Primer-BLAST 设计 *SOC1-4* 基因

qRT-PCR 表达引物, forward: 5'-CGAAACGGCGA-AATGGACTG-3' 和 reverse: 5'-CTGAATGTCTTG-CCCAGCAG-3', 引物合成由上海生工生物工程技术服务有限公司完成。以 *Actin* 作为内参基因^[20], 设计引物序列, forward: 5'-TGGTATTGTGCTG-GATTCTGGT-3' 和 reverse: 5'-TGAGATCACCA-CCAGCAAGG-3', 利用 FastKing cDNA Kit (KR116, 天根生化科技有限公司) 合成 cDNA; 利用 SuperReal PreMix Plus (SYBR Green) 进行 qRT-PCR 检测, 具体反应体系及条件见说明书。利用 $2^{-\Delta\Delta C_t}$ 法计算 *SOC1-4* 基因的相对表达水平^[21]。

2.5 统计分析

在 qRT-PCR 实验中, 所采集材料进行了 3 个生物学重复和 3 个技术重复。运用 Excel 2019 进行数据的计算及制图。

3 结果与分析

3.1 当归 *MADS-box* 基因家族鉴定及理化性质分析

通过对当归不同品种叶片和叶柄、种苗春化作用过程中根茎顶端分生组织的全长转录组进行筛选, 发现有 29 个 *MADS-box* 基因, 基于所编码蛋白质序列和亚细胞定位分析, 结果显示, 蛋白质序列长度为 49 (*AP1-1*) ~ 422 (*At5g65490-1*)、相对分子质量为 5 697.56 ~ 49 624.90、等电点为 5.06 (*J*) ~ 11.00 (*AGL14-1*), 亚细胞定位分析预测 29 个 *MADS-box* 蛋白分别位于叶绿体 (*AGL14-1*、*At3g28050-2* 和 *At5g65490-1*)、细胞核 (*AGL14-2*、*AGL16* 和 *AGL19-1* 等 22 个)、细胞质 (*AP1-1*、*AP1-2* 和 *At3g28050-1* 等 7 个) 和线粒体 (*SVP*) (表 1)。

3.2 当归 *MADS-box* 基因家族蛋白质结构分析

通过对 29 个 *MADS-box* 基因所编码的蛋白质进行二级结构预测以及 3D 建模, 结果显示, 蛋白质二级结构由 α 螺旋、延伸链、 β 转角和无规则卷曲组成 (表 2); 三级结构预测结果 (图 1) 与二级结构 (表 2) 相符合, 主要由 α 螺旋、无规则卷曲和延伸链进一步折叠组装为三级结构, 各亚家族内蛋白质三级结构较为相似, 如 *AGL14-2*、*SOC1-1*、*SOC1-2*、*SOC1-3* 和 *SOC1-4* 均以 α 螺旋为主, 无规则卷曲次之, β 转角最少。但各亚家族之间空间结构差异较大, 如 *AGL65*、*SOK2* 及 *AP1-2* 等以无规则卷曲为主、而 *AP1-1* 无延伸链 (图 1)。

表1 当归 *MADS-box* 基因家族蛋白质特征及亚细胞定位Table 1 Protein characteristic and subcellular location of *MADS-box* gene family in *A. sinensis*

序号	基因	同源基因	CDS 长度/bp	蛋白质大小/aa	相对分子量	等电点	亚细胞定位
1	<i>AGL14-1</i>	AT5G59890.3	189	63	7 160.35	11.00	叶绿体
2	<i>AGL14-2</i>	AT5G14740.9	651	217	24 768.39	9.16	细胞核
3	<i>AGL16</i>	AT1G34460.4	663	221	24 926.41	6.87	细胞核
4	<i>AGL19-1</i>	AT5G53280.1	639	213	24 661.17	6.88	细胞核
5	<i>AGL19-2</i>	AT1G68010.1	639	213	24 721.31	9.24	细胞核
6	<i>AGL27</i>	AT5G55990.2	649	216	24 570.04	5.58	细胞核
7	<i>AGL31-1</i>	AT5G55990.2	432	144	16 133.41	8.66	细胞核
8	<i>AGL31-2</i>	AT5G55990.2	648	216	24 695.12	6.46	细胞核
9	<i>AGL65</i>	AT5G64630.1	954	318	36 406.77	5.28	细胞核
10	<i>AGL8</i>	AT1G20450.2	723	241	27 846.49	6.22	细胞核
11	<i>AP1-1</i>	AT5G05090.1	147	49	5 697.56	8.35	细胞核、细胞质
12	<i>AP1-2</i>	AT1G77230.1	1215	405	44 790.06	8.40	细胞质
13	<i>AP1-3</i>	AT1G20450.2	723	241	27 846.49	6.22	细胞核
14	<i>At3g28050-1</i>	AT1G80060.4	768	256	28 381.08	6.37	细胞质
15	<i>At3g28050-2</i>	AT3G28130.6	834	278	29 881.93	8.82	叶绿体
16	<i>At5g65490-1</i>	AT4G31073.1	1266	422	49 624.90	9.55	叶绿体、细胞质
17	<i>At5g65490-2</i>	AT4G11780.1	1101	367	12 146.37	9.93	细胞质
18	<i>FLC</i>	AT5G55990.2	651	217	24 839.14	6.25	细胞核
19	<i>J</i>	AT1G34460.4	477	159	17 539.96	5.06	细胞核、细胞质
20	<i>MADS8</i>	AT3G54480.2	534	178	20 651.51	8.97	细胞核
21	<i>SCM1</i>	AT2G01540.1	210	70	8 095.52	9.86	细胞核
22	<i>SOC1-1</i>	AT5G14740.9	651	217	24 768.39	9.16	细胞核
23	<i>SOC1-2</i>	AT3G02470.4	558	186	21 447.56	9.13	细胞核
24	<i>SOC1-3</i>	AT4G22950.2	651	217	24 939.61	9.28	细胞核
25	<i>SOC1-4</i>	AT1G72860.1	627	209	23 861.12	5.91	细胞核
26	<i>SOK2</i>	AT1G80790.3	1089	363	40 264.47	7.78	细胞核、细胞质
27	<i>SVP</i>	AT1G07705.1	690	230	24 845.24	9.77	线粒体
28	<i>TM6-1</i>	AT3G29770.2	537	179	20 962.83	9.05	细胞核
29	<i>TM6-2</i>	AT3G29770.2	675	225	26 200.86	9.31	细胞核

表2 当归 *MADS-box* 基因家族蛋白质二级结构Table 2 Secondary structure of *MADS-box* gene family in *A. sinensis*

序号	基因	α 螺旋	占比/%	延伸链	占比/%	β 转角	占比/%	无规则卷曲	占比/%
1	<i>AGL14-1</i>	131	76.16	18	10.47	9	5.23	14	8.14
2	<i>AGL14-2</i>	139	60.70	24	10.48	12	5.24	54	23.58
3	<i>AGL16</i>	130	59.09	27	12.27	8	3.64	55	25.00
4	<i>AGL19-1</i>	132	64.08	27	13.11	8	3.88	39	18.93
5	<i>AGL19-2</i>	83	50.00	19	11.45	10	6.02	54	32.53
6	<i>AGL27</i>	140	62.50	30	13.39	8	3.57	46	20.54
7	<i>AGL31-1</i>	107	63.69	23	13.69	10	5.95	28	16.67
8	<i>AGL31-2</i>	128	62.75	24	11.76	13	6.37	39	19.12
9	<i>AGL65</i>	43	16.93	35	13.78	13	5.12	163	64.17
10	<i>AGL8</i>	113	50.00	24	10.62	10	4.42	79	34.96
11	<i>AP1-1</i>	128	96.24	0	0.00	1	0.75	4	3.01
12	<i>AP1-2</i>	123	29.64	95	22.89	34	8.19	163	39.28
13	<i>AP1-3</i>	118	51.08	24	10.39	10	4.33	79	34.20
14	<i>At3g28050-1</i>	144	51.99	46	16.61	20	7.22	67	24.19
15	<i>At3g28050-2</i>	53	24.88	62	29.11	9	4.23	89	41.78
16	<i>At5g65490-1</i>	134	32.52	74	17.96	34	8.25	170	41.26
17	<i>At5g65490-2</i>	100	67.11	16	10.74	7	4.70	26	17.45
18	<i>FLC</i>	109	56.77	26	13.54	10	5.21	47	24.48
19	<i>J</i>	22	28.95	15	19.74	7	9.21	32	42.11
20	<i>MADS8</i>	127	64.80	21	10.71	10	5.10	38	19.39
21	<i>SCM1</i>	120	71.43	27	16.07	7	4.17	14	8.33
22	<i>SOC1-1</i>	141	61.04	24	10.39	12	5.19	54	23.38
23	<i>SOC1-2</i>	128	65.98	18	9.28	8	4.12	40	20.62
24	<i>SOC1-3</i>	70	47.95	20	13.70	9	6.16	47	32.19
25	<i>SOC1-4</i>	138	63.01	18	8.22	12	5.48	51	23.29
26	<i>SOK2</i>	108	26.93	53	13.22	28	6.98	212	52.87
27	<i>SVP</i>	131	61.50	19	8.92	13	6.10	50	23.47
28	<i>TM6-1</i>	108	60.34	19	10.61	11	6.15	41	22.91
29	<i>TM6-2</i>	131	58.22	32	14.22	20	8.89	42	18.67

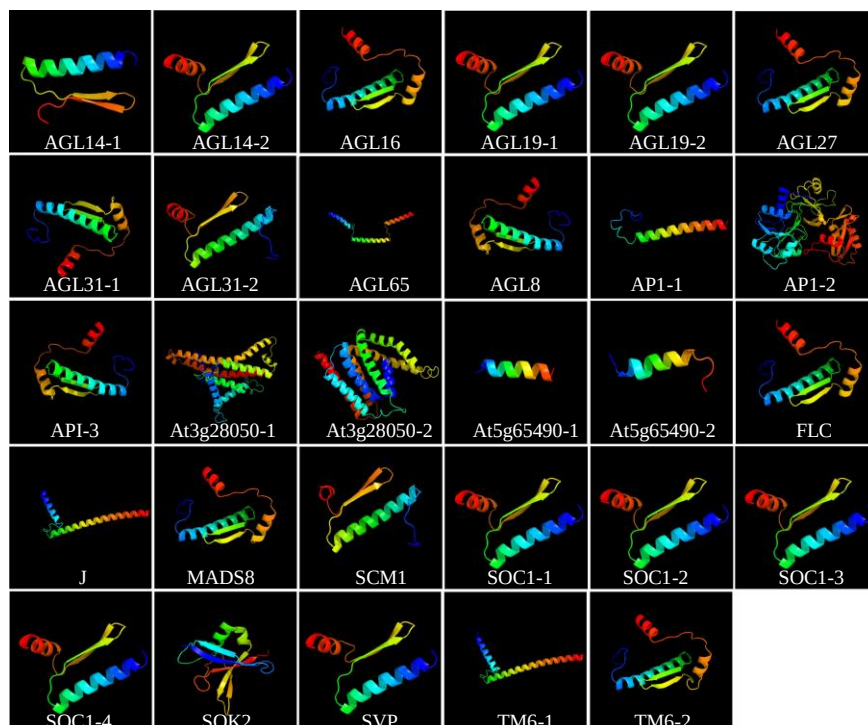


图 1 当归 MADS-box 基因家族蛋白质三级结构

Fig. 1 Tertiary structure of MADS-box gene family in *A. sinensis*

3.3 当归 MADS-box 基因家族发育树构建

通过对 29 个当归 MADS-box 基因家族蛋白质、61 个拟南芥和 55 个黄胡萝卜 MADS-box 蛋白质进行系统进化树构建，基于蛋白质序列相似性，145 个 MADS-box 蛋白质分为 10 个亚家族，可分为 I

型和 II 型，其中，I 型包括 suppressor-like，II 型包括 SOC1、TM3、DEF、MADS8、SQUA、FLC、STMADS11、MIKC 和 SOK；另外，SOC1 亚家族包括 5 个当归 MADS-box 成员（SOC1-1、SOC1-2、SOC1-3、SOC1-4 和 AGL14-2）（图 2）。

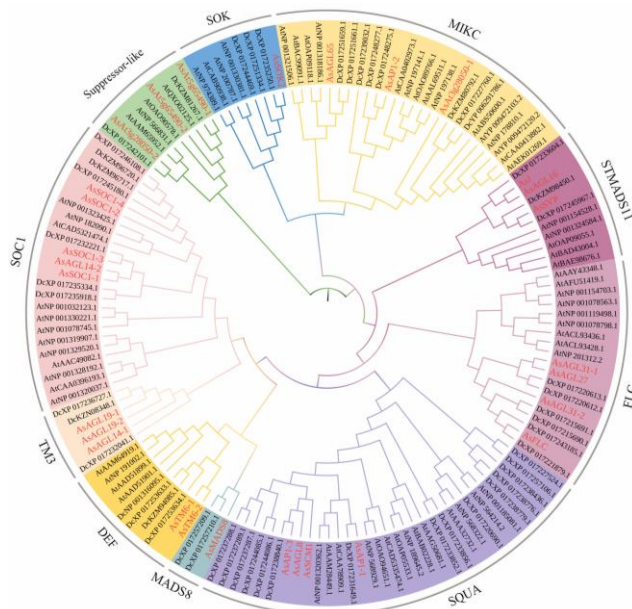


图 2 当归、拟南芥和黄胡萝卜 MADS-box 蛋白质系统进化树

Fig. 2 Phylogenetic tree of MADS-box family proteins in *A. sinensis*, *A. thaliana* and *D. carota* subsp. *sativus*

3.4 当归 MADS-box 基因家族基序分析

通过对以上 29 个当归 MADS-box 基因家族蛋白质、61 个拟南芥和 55 个黄萝卜 MADS-box 蛋白质进行结构域和基序分析,结果显示,MADS-box 蛋白质分为 10 个亚家族,与图 2 系统进化树关系一致,序列含有 6 个保守基序(表 3);当归 MADS-box 同一亚家族蛋白质基序具有高度相似性,其中,SOC1 亚家族均含有 Motif 1~Motif 5;不同亚家族蛋白质基序存在较大差异,比如,Motif 6 只存在于 suppressor-like 亚家族(图 3)。

3.5 当归 SOC1 亚家族蛋白质多序列比对

通过对当归 SOC1 亚家族 5 个成员(SOC1-1、SOC1-2、SOC1-3、SOC1-4 和 AGL14-2)及来自拟南芥(AtNP_182090.1)和黄萝卜(DcXP_017232221.1)2 个物种的 SOC1 同源蛋白质等 7 个蛋白质进行多序列比对,结果显示,SOC1 亚家族 5 个蛋白质均含 MADS-domain、I-domain、K-domain 和 C-domain 结构域,其中,MADS-domain 结构域高度保守,C-domain 结构域保守性较低(图 4);通过对 SOC1-1、SOC1-2、SOC1-3、SOC1-4 进行进一步比对,发现 SOC1-1、SOC1-3 和 SOC1-4 蛋白质的 4 个结构域较为完整、且保守性较高,而 SOC1-2 在 C-domain 不完整;其中,SOC1-1、SOC1-3 和 SOC1-4 的相似性为 73.27%。

3.6 当归 SOC1-4 基因克隆

为了保证全长转录组中当归 SOC1-4 基因碱基序列的准确性,基于全长转录组中 SOC1-4 基因碱基序列设计扩增引物,以岷归 1 号功能叶片中 RNA 反转录所得的 cDNA 为模板,进行 SOC1-4 基因克隆。琼脂糖凝胶电泳显示,SOC1-4 基因扩增片段大小在 500~750 bp(图 5);产物回收与碱基测序显示,SOC1-4 基因克隆长度为 585 bp;通过与全长转录组测序获得的 SOC1-4 进行序列比对,根据引物设计所得克隆结果与测序序列 5-590 bp 相似度为 100%(图 6)。

3.7 当归 SOC1-4 基因表达分析

为了进一步验证 SOC1-4 基因的生物学功能,本研究对当归不同材料中 SOC1-4 基因的表达水平进行了 qRT-PCR 检测与分析。结果显示,早薹植株相对非早薹植株,SOC1-4 基因表达量上调 6.42 倍;不同生长期植株中,SOC1-4 基因表达水平随着时间延长呈现逐渐增加,S2、S3 和 S4 时期相对于 S1 时期分别增加 2.16、2.31 和 5.79 倍;不同春化期种苗根茎顶端分生组织中,通过春化作用(T2)处理相对于未春化作用(T1)SOC1-4 基因表达水平呈现 3.39 倍高表达,而规避春化作用(T3)处理相对于 T1 处理 SOC1-4 基因表达水平呈现 0.48 倍降低;不同器官中,茎和叶相对于根 SOC1-4 基因表达水平分别上调 9.23 和 4.90 倍(图 7)。

4 讨论

目前,早薹开花导致根木质化不能入药仍是困扰当归生产、质量和效益提升的重大难题^[4]。研究发现,当归抽薹开花受到内在(如种质、苗龄和种苗大小等)和外在(如温度、光照和干旱等)多种因素的影响^[4, 22]。此外,当归为“低温长日照型”植物,即植株由营养生长转入生殖生长必须同时满足低温春化作用和长日照^[22-24]。大量研究表明,MADS-box 基因家族,尤其是开花抑制因子 *FLC* 和整合因子 *SOC1*,在植物花器官分化和开花时间调节等方面起到核心调节作用^[17]。尽管 MADS-box 基因家族在模式植物拟南芥及其他植物中调节开花的作用已进行了深入研究,然而在药用植物当归中的生物学功能还鲜见报道。本研究基于前期当归全长转录组测序,发现有 29 个 MADS-box 基因分布于 10 个亚家族,所编码蛋白质序列含有 6 个保守基序,其中,亚家族有 5 个成员;另外,还对 SOC1-4 基因进行了克隆和表达分析,为后续探究 SOC1 在当归抽薹开花的分子调控研究奠定了基础。

模式植物拟南芥中有 DEF、STMADS11 和 TM3-like 等 12 个 MADS-box 亚家族^[11];山茶中有

表 3 MADS-box 蛋白质 6 个保守基序及其序列

Table 3 Six motifs and their conserved sequences of MADS-box proteins

基序	长度/aa	保守序列
Motif 1	50	IKRIENKTNRQVTFKRRNGLLKKAFELSVLCDAEVALIIFSSRGKLYEF
Motif 2	29	SLSIEELQQLEQQLETSLKQIRARKTQLM
Motif 3	29	QQHLKLETAKLKKKIELLZRSKRKLLGEG
Motif 4	29	MEQIEZLQEKEKLLQEENKRLRKKIKERE
Motif 5	21	SSSSMEKIJERYQKYTKDDRS
Motif 6	50	WTLPIFHANGWSYPWGIAAVGGTNVCLRKFDAPLIYRLIRDHGVTHMCGA



图3 当归、拟南芥和黄萝卜 MADS-box 蛋白质保守基序

Fig. 3 Conserved sequences of MADS-box proteins in *A. sinensis*, *A. thaliana* and *D. carota* subsp. *sativus*

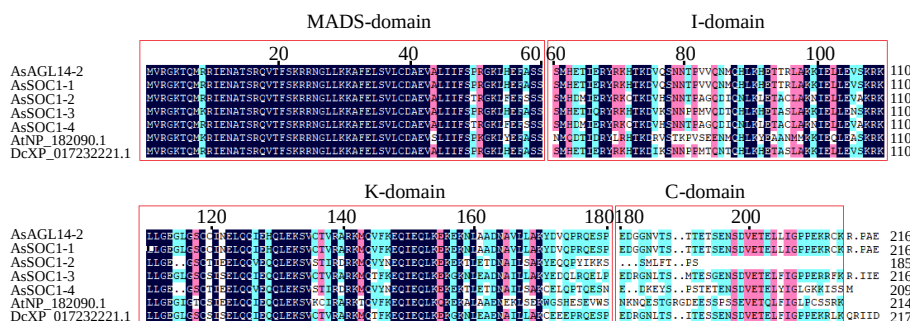


图4 当归 SOC1 亚家族蛋白质多序列比对

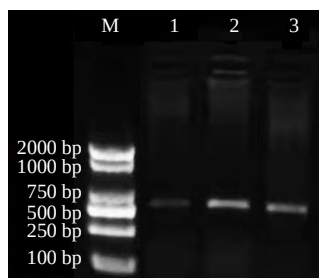
Fig. 4 Multiple sequence alignment of SOC1 subfamily proteins in *A. sinensis*M-Marker 1~3-SOC1-4 克隆片段
M-Marker 1—3-Cloned SOC1-4 sequence

图5 当归 SOC1-4 基因扩增产物琼脂糖凝胶电泳

Fig. 5 Agarose gel electrophoresis of amplification products of SOC1-4 gene in *Angelica sinensis*

83 个 MADS-box 基因, 可分为 I 型 (Mα、Mβ 和 Mγ 亚家族) 和 II 型 (MIKC^C 和 MIKC^{*}亚家族) [25]; 荔枝中有 101 个 MADS-box 基因, 可分为 I 型和 II 型, 其中, 50 个 I 型分为 Mα、Mβ 和 Mγ 3 个亚家族, 51 个 II 型进一步分为 MIKC^{*}、SOC1 和 FLC 等 13 个亚家族 [26]。本研究基于全长转录组鉴定出 29 个 MADS-box 基因家族成员, 基于蛋白质序列分为 10 个亚家族, 可分为 I 型 (Suppressor-like) 和 II 型 (SOC1、TM3、DEF、MADS8、SQUA、FLC、STMADS11、MIKC 和 SOK); 其中, SOC1 亚家族包括 SOC1-1、SOC1-2、SOC1-3、SOC1-4 和 AGL14-2。另外, 前人预测 22 种植物 (拟南芥、水



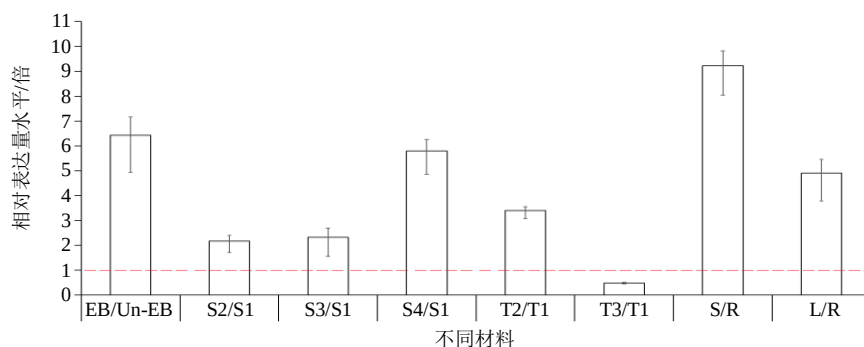
图6 当归 SOC1-4 基因克隆与全长转录组测序的序列比对

Fig. 6 Sequence alignment of SOC1-4 gene obtained from gene clone and full-length sequencing in *A. sinensis*

稻、烟草等) 的 SOC1 二级结构与亚细胞定位 [27], 与本研究 MADS-box 蛋白质理化性质和结构分析, 尤其是 SOC1 亚家族等的结果基本一致, 表明表明 SOC1 基因在进化上较为保守, 在植物的生殖发育过程起着重要作用。

植物中 MADS-box 的 I 型和 II 型蛋白质结构域各具有高度相似性, 其中 I 型只含有一个保守结构

域, 而 II 型均具有 M-domain、I-domain、K-domain 和 C-domain 结构域, 推测当归 MADS-box 基因功能存在多样性 [7, 11, 28-30]。蛋白质保守基序比对发现不同亚家族保守基序存在数目及位置差异, 预测不同亚家族功能有所差异。本研究中, SOC1 亚家族均含有 II 型的 4 个结构域, 通过对 5 个成员 SOC1-1、SOC1-2、SOC1-3、SOC1-4 和 AGL14-2 进行多序列



EB-抽薹 Un-EB-非早薹 S1-营养生长期 S2-营养生长到生殖生长过渡期 S3-抽薹初期 S4-抽薹伸长期 T1-未通过春化 T2-通过春化 T3-低温规避春化 S-茎 L-叶 R-根
EB-early bolting Un-EB-un-early bolting S1-stem-node pre-differentiation S2-differentiation stages S3-forming stages S4-elongating stages T1-un-completed vernalization T2-completed vernalization T3-frozen-avoided vernalization S-stem L-leaf R-root

图 7 当归 SOC1-4 基因在不同材料中的相对表达水平

Fig. 7 Relative expression level of SOC1-4 gene in different materials of *A. sinensis*

比对,发现 MADS-domain 结构域高度保守,而 C-domain 结构域保守性较低。这表明,5 个 SOC1 亚家族成员在当归抽薹开花过程中可能发挥相似或协同的功能。另外, SOC1-1、SOC1-3 和 SOC1-4 蛋白质的 4 个结构域较为完整,而 SOC1-2 在 C-domain 不完整,而樊世婷等^[31]克隆的当归 SOC1 基因(XM_017379845.1)与本研究 4 个 SOC1 基因比对存在差异,可能由于当归无参考基因组序列,转录组序列拼接参考不同物种而产生的差异;或全长转录组样本来自不同品种及材料所致。

尽管前期研究已获得了当归全长转录组,为了保证 SOC1 基因碱基序列的准确性、以及更深入探究 SOC1 基因的生物学特性,本研究通过克隆当归 SOC1-4 基因,获得了 585 bp 片段,与全长转录组测序中 SOC1-4 基因序列 5-590 bp 结果完全一致,大量研究证实,作为开花整合因子 SOC1,在植物花器官分化和开花时间调控过程中高表达^[17, 32]。本研究发现, SOC1 基因表达量随着种苗春化作用和植株生长发育时期延长随之增加、在抽薹植株中高于非抽薹植株、而在冷冻规避春化作用种苗中显著降低。

综合以上研究表明,本课题组首次对当归 MADS-box 基因家族进行了挖掘和生物信息学分析,并对开花整合因子 SOC1 基因进行了基因克隆与表达验证。但对于 SOC1 基因片段的完整性、以及其它调控当归抽薹开花的关键基因(如 FLC、AG 和 AP1 等)还需要进一步研究和验证。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国药典 [S]. 一部. 2020: 139.
- [2] Wei W L, Zeng R, Gu C M, et al. *Angelica sinensis* in China-a review of botanical profile, ethnopharmacology, phytochemistry and chemical analysis [J]. *J Ethnopharmacol*, 2016, 190: 116-141.
- [3] 晋玲. 当归生产加工适宜技术 [M]. 北京: 中国医药科技出版社, 2018: 2.
- [4] 栗孟飞, 康天兰, 晋玲, 等. 当归抽薹开花及其调控途径研究进展 [J]. *中草药*, 2020, 51(22): 5894-5899.
- [5] Zhang H Y, Bi W G, Yu Y, et al. *Angelica sinensis* (Oliv.) Diels in China: Distribution, cultivation, utilization and variation [J]. *Genet Resour Crop Evol*, 2012, 59(4): 607-613.
- [6] Li M L, Cui X W, Jin L, et al. Bolting reduces ferulic acid and flavonoid biosynthesis and induces root lignification in *Angelica sinensis* [J]. *Plant Physiol Biochem*, 2022, 170: 171-179.
- [7] 郭斌, 郎建军, 郭增祥, 等. 关于当归早期抽薹的危害及几点建议 [J]. *农民致富之友*, 2017, 565(20): 75.
- [8] 梁婕, 王辉, 睢宁. 伞形科根茎类药用植物早期抽薹的研究进展 [J]. *中国农学通报*, 2022, 38(13): 90-95.
- [9] 王莹, 穆艳霞, 王锦. MADS-box 基因家族调控植物花器官发育研究进展 [J]. *浙江农业学报*, 2021, 33(6): 1149-1158.
- [10] 李元元, 王鲁, 苏振刚, 等. MADS-box 基因控制植物开花的分子机理 [J]. *基因组学与应用生物学*, 2010, 29(6): 1122-1132.
- [11] Becker A, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants [J]. *Mol Phylogenetics Evol*, 2003, 29(3): 1122-1132.

- 464-489.
- [12] Luo M M, Liu X X, Su H Y, *et al.* Regulatory networks of flowering genes in *Angelica sinensis* during vernalization [J]. *Plants (Basel)*, 2022, 11(10): 1355.
- [13] Li M F, Li J, Wei J H, *et al.* Transcriptional controls for early bolting and flowering in *Angelica sinensis* [J]. *Plants (Basel)*, 2021, 10(9): 1931.
- [14] Li J, Li M L, Zhu T T, *et al.* Integrated transcriptomics and metabolites at different growth stages reveals the regulation mechanism of bolting and flowering of *Angelica sinensis* [J]. *Plant Biol (Stuttg)*, 2021, 23(4): 574-582.
- [15] Gao X, Guo F X, Chen Y, *et al.* Full-length transcriptome analysis provides new insights into the early bolting occurrence in medicinal *Angelica sinensis* [J]. *Sci Rep*, 2021, 11(1): 13000.
- [16] Yu G, Zhou Y, Yu J J, *et al.* Transcriptome and digital gene expression analysis unravels the novel mechanism of early flowering in *Angelica sinensis* [J]. *Sci Rep*, 2019, 9(1): 10035.
- [17] Lincoln Taiz, Eduardo Zeiger 著. 宋纯鹏, 王学路, 周云等译. 植物生理学 [M]. 北京: 科学出版社, 2015: 569-596.
- [18] Zhu T T, Zhang M H, Su H Y, *et al.* Integrated metabolomic and transcriptomic analysis reveals differential mechanism of flavonoid biosynthesis in two cultivars of *Angelica sinensis* [J]. *Molecules*, 2022, 27(1): 306.
- [19] Liu X X, Luo M M, Li M F, *et al.* Depicting precise temperature and duration of vernalization and inhibiting early bolting and flowering of *Angelica sinensis* by freezing storage [J]. *Front Plant Sci*, 2022, 13: 853444.
- [20] Xu R, Xu J, Li Y C, *et al.* Integrated chemical and transcriptomic analyses unveils synthetic characteristics of different medicinal root parts of *Angelica sinensis* [J]. *Chin Herb Med*, 2020, 12(1): 19-28.
- [21] Willems E, Leyns L, Vandesompele J. Standardization of real-time PCR gene expression data from independent biological replicates [J]. *Anal Biochem*, 2008, 379(1): 127-129.
- [22] 李明世, 于兆英. 当归及其防止早期抽苔的研究 [J]. 中草药, 1977, 8(12): 34-38.
- [23] 王文杰, 张正民. 当归的抽苔特性和控制途径 [J]. 西北植物研究, 1982, 2(2): 95-104.
- [24] 王文杰. 对当归早期抽苔特性的分析和控制 [J]. 西北大学学报: 自然科学版, 1977, 7(2): 29-36.
- [25] Zhang Z B, Jin Y J, Wan H H, *et al.* Genome-wide identification and expression analysis of the MADS-box transcription factor family in *Camellia sinensis* [J]. *J Appl Genetics*, 2021, 62(2): 249-264.
- [26] Guan H L, Wang H, Huang J J, *et al.* Genome-wide identification and expression analysis of MADS-box family genes in *Litchi (Litchi chinensis sonn.)* and their involvement in floral sex determination [J]. *Plants (Basel)*, 2021, 10(10): 2142.
- [27] 齐联联, 宿强, 张珂. *SOC1* 调控植物开花时间的分子机制 [J]. 草业科学, 2022, 39(1): 149-160.
- [28] 张加强, 朱开元, 史小华. 芍药 *MADS-box* 基因家族的鉴定及适应性进化分析 [J]. 分子植物育种, 2019, 17(21): 6959-6966.
- [29] 郑玲, 谢爱玲, 韩建明. 高粱 *MADS-box* 家族基因的鉴定与分析 [J]. 东北农业科学, 2019, 44(5): 26-29.
- [30] 钱金森, 田雅, 王飞鉴, 等. 基于转录组测序信息的德国鸢尾 *MADS-box* 基因家族分析 [J]. 分子植物育种, 2022, 20(1): 49-63.
- [31] 樊世婷, 李云, 罗香怡, 等. 当归抽苔关键基因 *FT* 和 *SOC1* 的生物信息学分析 [J]. 分子植物育种, 2022, 20(11): 3562-3569.
- [32] Lee J, Lee I. Regulation and function of *SOC1*, a flowering pathway integrator [J]. *J Exp Bot*, 2010, 61(9): 2247-2254.

[责任编辑 时圣明]