

基于转录组信息的甘松 MADS-box 转录因子家族分析

崔 琪^{1,3,4}, 俸明康^{1,3,4}, 丰日落^{1,3,4}, 王 涛^{1,3,4}, 张绍山^{2,3,4}, 李 莹^{1,3,4}, 陈 晨^{2,3,4*}, 刘 圆^{3,4*}

1. 西南民族大学药学院, 四川 成都 610041

2. 西南民族大学青藏高原研究院, 四川 成都 610225

3. 四川省羌彝药用资源保护与利用技术工程实验室, 四川 成都 610225

4. 青藏高原民族药用资源保护与利用国家民委重点实验室, 四川 成都 610225

摘要: 目的 基于转录组数据筛选鉴定分析甘松 *Nardostachys jatamansi* MADS-box 转录因子家族。方法 利用生物信息学的方法全面分析甘松 MADS-box 基因家族成员的蛋白理化性质、亚细胞定位、基因结构、导肽、信号肽及跨膜结构域、系统进化树。结果 在甘松转录组数据中共鉴定出 20 个 MADS-box 转录因子, 含有 90~365 个氨基酸, 相对分子质量在 10 179.96~41 707.68, 理论等电点范围为 4.87~10.43, 主要在细胞核表达, 均含有 MADS 保守结构域, 均不含信号肽、导肽以及跨膜结构域。系统发育分析表明甘松 MADS 蛋白主要属于 TypeII 型。结论 利用生物信息学分析手段, 鉴定了甘松 MADS-box 家族转录因子, 为甘松深入研究甘松 MADS-box 蛋白的生物学功能提供参考, 为甘松的药用成分合成调控机制提供研究基础, 为培育优质甘松新品种提供依据。

关键词: 甘松; MADS-box; 转录因子分析; 花器官; 亚细胞定位

中图分类号: R286 文献标志码: A 文章编号: 0253-2670(2023)03-0898-09

DOI: 10.7501/j.issn.0253-2670.2023.03.024

Analysis of MADS-box transcription factor family based on transcriptome sequencing in *Nardostachys jatamansi*

CUI Qi^{1,3,4}, FENG Ming-kang^{1,3,4}, FENG Ri-luo^{1,3,4}, WANG Tao^{1,3,4}, ZHANG Shao-shan^{2,3,4}, LI Ying^{1,3,4}, CHEN Chen^{2,3,4}, LIU Yuan^{3,4}

1. College of Pharmacy, Southwest Minzu University, Chengdu 610041, China

2. Institute of Qinghai-Tibetan Plateau Research, Southwest Minzu University, Chengdu 610225, China

3. Sichuan Provincial Qiang-Yi Medicinal Resources Protection and Utilization Technology Engineering Laboratory, Chengdu 610225, China

4. Tibetan Plateau Ethnic Medicinal Resources Protection and Utilization Key Laboratory of National Ethnic Affairs Commission of China, Chengdu 610225, China

Abstract: Objective To identify and analyze MADS-box transcription factor family in Gansong (*Nardostachys jatamansi* DC.) based on transcriptome data. **Methods** The protein physicochemical properties, subcellular localization, gene structure, conducting peptide, signal peptide, transmembrane domain and phylogenetic tree of MADS-box gene family members in *N. jatamansi* were comprehensively analyzed by bioinformatics methods. **Results** A total of 20 MADS-box transcription factors were identified in the transcriptome data of *N. jatamansi*. And they have 90—365 amino acids, and the molecular weight ranged from 10 179.96 to 41 707.68. The theoretical isoelectric point (pI) ranged from 4.87 to 10.43. And most of MADS-box protein located in nucleus. All of them contained MADS conserved domains, none of which contains a signal peptide, a guide

收稿日期: 2022-10-06

基金项目: 国家重点研发计划 (2018YFC1706101); 四川省自然科学基金项目 (2022NSFSC1588); 2021 年国家民委领军人才支持计划, 西南民族大学研究生创新型科研项目 (CX2021SZ94)

作者简介: 崔 琪 (1997—), 硕士研究生, 研究方向为中药鉴定学。Tel: 18280563878 E-mail: 1747396783@qq.com

*通信作者: 陈 晨 (1987—), 助理研究员, 博士, 主要从事药用植物发育分子机制及民族药资源研究。

Tel: (028)83372832 E-mail: lilychenqhu@163.com

刘 圆 (1968—), 教授, 从事民族药品种、品质评价及新药资源保护与利用研究。Tel: (028)85528812 E-mail: 499769896@qq.com

peptide or a transmembrane domain. The results of phylogenetic analysis showed that the MADS protein mainly in *N. jatamansi* belonged to Type II. **Conclusion** In this study, the MADS-box gene family members were identified by bioinformatics analysis, which provides a reliable reference for further study on the biological function of MADS-box protein in *N. jatamansi*, offers a research basis for the regulation mechanism of medicinal composition synthesis of *N. jatamansi*, which further provides a basis for cultivating new varieties of *N. jatamansi* with high quality.

Key words: *Nardostachys jatamansi* DC.; MADS-box; transcription factor analysis; floral organ; subcellular localization

败酱科甘松属植物甘松 *Nardostachys jatamansi* DC. 主要分布于四川省、甘肃省、青海省、云南省、西藏等地海拔 2800~4500 m 的高山草甸区, 历版《中国药典》均有收载, 其药用部位为干燥根及根茎, 具有理气止痛、开郁醒脾, 外用祛湿消肿的功效; 用于脘腹胀满、食欲不振、呕吐; 外用治牙痛、脚气肿毒^[1], 是古印度阿育吠陀 (Ayurveda) 和尤纳尼医 (Unani) 学体系的常用药材, 在我国为藏、蒙、维、傈僳、纳西等传统民族医学临床常用品种^[2], 也是藏香的主要原料之一^[3]。另外, 甘松在精油、日化品等领域也有较大的经济价值。

MADS-box 蛋白家族是一类 N 端存在 1 个氨基酸数目为 58~60 的保守结构域-MADS-box 结构域^[4]的重要的转录因子, 广泛存在于真核生物中, 其名称源自酿酒酵母的 MCMI、拟南芥的 AGAMOUS、金鱼草的 DEFICIENS 和人类的 SRF4 4 种类^[5]。常被分为 Type I 型和 Type II 型, Type I 型包括 M α 、M β 、M γ 、M δ 亚族, 目前仅少数 Type I 型蛋白的生物学功能已被探明, 且与植物胚和胚乳的发育有关^[6], Type II 型包括 MIKC^e 和 MIKC^{*}亚族^[7], MIKC^{*}通常参与调控雄配子的生长发育, MIKC^e 对花器官生长发育有重要作用^[8], 在花育的 ABCDE 模型中, Type II 型蛋白中所涉及的花四聚体能调节不同花序中的特定表达过程^[9]。MADS-box 家族转录因子大多参与花器官的发育、开花时间和果实成熟等过程的调控^[7], 少数参与调控植物侧根的形成^[10]。

甘松的药用部位为营养器官-根茎, 开花时间及侧根的发育影响其产量及品质。因此, 甘松花及侧根发育分子调控机制的研究, 可为其优质高产及优良种质资源的培育提供理论基础和技术手段。MADS-box 家族基因对开花时间调控、花器官发育及侧根发育至关重要, 但甘松 MADS-box 家族转录因子未见报道, 因此, 基于课题组前期的转录组数据, 鉴定和分析甘松 MADS-box 转录因子家族的

物信息学功能, 为进一步探究甘松 MADS-box 家族转录因子在花器官形成机制及侧根发育的分子调控机制提供重要的理论基础。

1 材料与方法

1.1 样品与处理

甘松植株经西南民族大学青藏高原研究院刘圆教授鉴定为甘松 *Nardostachys jatamansi* DC., 现蕾后套袋收集同一株的种子。

将采集的甘松种子播种于花盆中, 共 6 盆, 在培养箱中 (温度 20 °C; 相对湿度 50%; 8 h/16 h 光照/黑暗) 培育至长出两片真叶。3 盆喷洒 50 mg/mL 吲哚乙酸溶液, 另外 3 盆喷洒同样数量的清水作空白对照, 24 h 后取处理后的甘松叶片, 每份样品采集 3 个生物学重复, 液氮速冻, -80 °C 保存, 送至上海欧易生物医学科技有限公司进行转录组测序。

1.2 甘松 MADS-box 家族的来源、筛选及序列分析

甘松 MADS-box 转录因子家族序列来源于课题组上传于 NCBI 的 SRA 数据库中的转录组数据, 检索号分别为 SRR21656214、SRR21656215、SRR21656216、SRR21656217、SRR21656218、SRR21656219、SRR21656220、SRR21656221、SRR21656222、SRR21656223、SRR21656224、SRR21656225、SRR21656226、SRR21656227、SRR21656228、SRR21656229、SRR21656230、SRR21656231。利用数据中 NR 和 Swiss 注释搜索甘松中潜在的 MADS-box 转录因子, 筛选的 *E* 值为 1×10^{-5} 。利用数据在 NCBI 查找其开放阅读框 (open reading frame, ORF), 再利用 CDD、pfam 对保守结构域进行鉴定, 筛选并剔除不完整的甘松 MADS-box 转录因子, 最终获得 20 条 MADS-box 转录因子序列, 将其命名为 NcMADS01~NcMADS20。

1.3 甘松 MADS-box 家族蛋白理化性质、二级结构、三级结构和保守基序 (motif) 分析

对筛选出的 MADS-box 家族蛋白序列, 利用

ExPasy (<https://web.expasy.org/protparam/>) 分析甘松 MADS-box 转录因子家族成员的蛋白相对分子量、理论等电点等基本理化性质, 利用 SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) 预测甘松 MADS-box 蛋白二级结构, 利用 MEME (<https://memesuite.org/meme/tools/meme>) 分析甘松 MADS-box 蛋白保守基序及其分布, 利用 SWISS MODEL (<https://swissmodel.expasy.org/>) 预测甘松 MADS-box 蛋白的三维结构。

1.4 甘松 MADS-box 家族蛋白亚细胞定位、导肽、信号肽、跨膜结构预测分析

利用 Cello (<http://cello.life.nctu.edu.tw/>) 进行亚细胞定位预测分析。利用 TargetP-2.0 (<https://services.healthtech.dtu.dk/service.php?TargetP-2.0>) 预测导肽, 利用 SignalP-4.1 (<http://www.cbs.dtu.dk/services/SignalP-4.1/>) 预测信号肽, 跨膜结构预测通过 TM-HMM Server.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) 完成。

1.5 甘松 MADS-box 家族蛋白系统进化树构建

使用 MEGA6 软件以邻位连接法 (neighbor-joining, NJ) 构建甘松 MADS-box 蛋白序列与 TAIR 拟南芥官网 (https://www.arabidopsis.org/browse/genefamily/mads_tffamily.jsp) 中下载的拟南芥 MADS-box 蛋白序列的系统发育树, 并用 iTOL 软件进行美化。

2 结果与分析

2.1 甘松 MADS-box 家族蛋白的筛选

甘松相关数据来源于课题组前期的转录组测序数据。本研究在 Swiss-Prot、NR 数据库中筛选得到了 39 个 MADS-box 家族转录因子, 再通过 ORFfinder 对其开放阅读框进行分析, 进一步通过 CDD 和 Pfam 进行结构域预测分析, 删除没有完整保守结构域以及重复的序列, 最终得到 20 个 MADS-box 蛋白, 按照转录组数据编号从小到大的顺序进行编号, 分别编号为 NcMADS01 ~ NcMADS20。

2.2 甘松 MADS-box 蛋白序列理化性质分析

利用 Protparam 在线工具, 对 20 条甘松 MADS-box 蛋白序列进行理化性质分析, 结果显示, 20 条 NcMADS 蛋白序列长度在 90~365 aa, 相对分子量在 10 179.96~41 707.68。理论等电点 (PI) 在 4.87~10.43, 其中 9 个蛋白等电点 < 7 显酸性, 11 个蛋白等电点 > 7 显碱性。20 个 NcMADS 蛋白中, 3 个蛋白序列的不稳定指数 < 40, 为稳定蛋白, 17 个蛋白序列的不稳定指数 > 40, 为不稳定蛋白。甘松 MADS-box 蛋白脂肪系数在 73.12~108.44, 其中有 2 个脂肪系数大于 100。20 个甘松 MADS-box 蛋白的平均疏水指数 -0.84~-0.15, 均为负值, 其中 16 个小于 -0.5, 占比 80%, 可推出大部分甘松 MADS-box 蛋白质是亲水蛋白质, 两性蛋白质为 4 个, 占比 20% (表 1)。

表 1 甘松 MADS-box 转录因子家族信息

Table 1 Information of the MADS-box transcription factor family in *N. jatamansi*

编号	基因 ID	氨基酸长度/aa	相对分子量	PI	不稳定指数	脂肪系数	平均疏水指数
NcMADS01	TRINITY_DN11162_c0_g1_i1_6	262.00	29 910.02	6.20	57.35	83.36	-0.67
NcMADS02	TRINITY_DN12589_c0_g1_i4_6	93.00	10 774.83	10.43	50.41	103.76	-0.41
NcMADS03	TRINITY_DN12773_c0_g1_i1_8	360.00	41 707.68	6.49	60.08	76.08	-0.84
NcMADS04	TRINITY_DN13263_c0_g1_i2_4	211.00	24 489.03	8.86	56.29	86.45	-0.74
NcMADS05	TRINITY_DN18675_c0_g1_i4_5	144.00	16 712.41	9.86	37.73	90.07	-0.72
NcMADS06	TRINITY_DN18884_c0_g1_i3_2	90.00	10 179.96	9.52	44.05	108.44	-0.15
NcMADS07	TRINITY_DN19268_c0_g1_i6_4	227.00	25 795.03	5.34	59.43	85.51	-0.74
NcMADS08	TRINITY_DN19624_c0_g1_i1_5	365.00	41 644.76	5.63	54.79	76.96	-0.68
NcMADS09	TRINITY_DN20209_c0_g1_i1_4	244.00	28 008.67	7.63	46.04	75.16	-0.78
NcMADS10	TRINITY_DN20649_c0_g1_i1_8	176.00	20 392.56	9.21	49.29	95.80	-0.47
NcMADS11	TRINITY_DN21982_c0_g1_i1_1	259.00	29 405.05	4.87	53.64	78.73	-0.63
NcMADS12	TRINITY_DN22730_c0_g1_i2_8	205.00	23 549.80	9.02	41.39	88.88	-0.68
NcMADS13	TRINITY_DN22731_c0_g1_i2_9	238.00	26 651.46	6.03	47.71	84.87	-0.57
NcMADS14	TRINITY_DN23076_c0_g1_i2_5	231.00	26 768.43	6.63	69.54	83.16	-0.79
NcMADS15	TRINITY_DN24808_c0_g1_i3_5	92.00	10 399.94	9.79	41.67	83.70	-0.57
NcMADS16	TRINITY_DN26219_c0_g1_i13_9	92.00	10 732.39	9.48	33.23	78.37	-0.54
NcMADS17	TRINITY_DN27867_c1_g3_i1_5	262.00	29 687.69	6.67	60.26	78.51	-0.67
NcMADS18	TRINITY_DN32808_c0_g1_i15_5	157.00	18 753.22	9.80	48.12	97.52	-0.24
NcMADS19	TRINITY_DN3508_c0_g1_i1_8	204.00	23 338.50	5.08	32.84	80.29	-0.54
NcMADS20	TRINITY_DN9885_c0_g1_i1_2	256.00	29 338.30	8.53	49.38	73.12	-0.74

2.3 甘松 MADS-box 蛋白二级结构与三级结构分析

蛋白质的二级结构是蛋白质中有规则重复的构象，研究蛋白质二级结构是研究蛋白质结构和功能的基础。甘松 MADS-box 蛋白序列的二级结构均含有 α-螺旋、β-转角、无规则卷曲和延伸链，α-螺旋和无规则卷曲作为二级结构的主要构成元件，占二级结构的 70%左右，这对蛋白质的特殊结构构象具有一定的作用，无规则卷曲和 β-转角占比相对较小（表 2）。

利用在线软件 SWISS-MODEL 对甘松的 20 个 MADS-box 家族基因进行结构预测（图 1），可从构建的模型中清楚地观察到各家族蛋白的空间结构（α-螺旋、β-转角、无规则卷曲、延伸链）。结果显示甘松 9 个亚族之间以及同一亚族不同蛋白之间的 α-螺旋、β-转角、无规则卷曲、延伸链的比例均不相同，从而导致其空间结构存也存在一定差异，结构的差异使得各自呈现不同的功能。

表 2 甘松 MADS-box 转录因子家族二级结构
Table 2 Secondary structure of the MADS-box transcription factor family in *N. jatamansi*

编号	α-螺旋/%	β-转角/%	延伸链/%	无规则卷曲/%
NcMADS01	50.38	11.83	1.91	35.88
NcMADS02	40.86	22.58	9.68	26.88
NcMADS03	34.72	12.78	4.72	47.78
NcMADS04	65.88	12.32	3.32	18.48
NcMADS05	61.81	15.28	6.25	16.67
NcMADS06	43.33	21.11	6.67	28.89
NcMADS07	61.23	10.13	3.52	25.11
NcMADS08	30.14	13.15	3.29	53.42
NcMADS09	52.05	7.79	4.51	35.66
NcMADS10	64.20	14.77	3.41	17.61
NcMADS11	51.35	13.90	1.16	33.59
NcMADS12	49.81	15.44	3.86	30.89
NcMADS13	54.62	13.03	4.20	28.15
NcMADS14	61.04	8.23	1.73	29.00
NcMADS15	50.00	22.83	6.52	20.65
NcMADS16	25.00	32.61	17.39	25.00
NcMADS17	47.71	14.12	3.82	34.35
NcMADS18	42.04	24.84	11.46	21.66
NcMADS19	43.14	15.69	6.37	34.80
NcMADS20	52.34	9.77	4.30	33.59

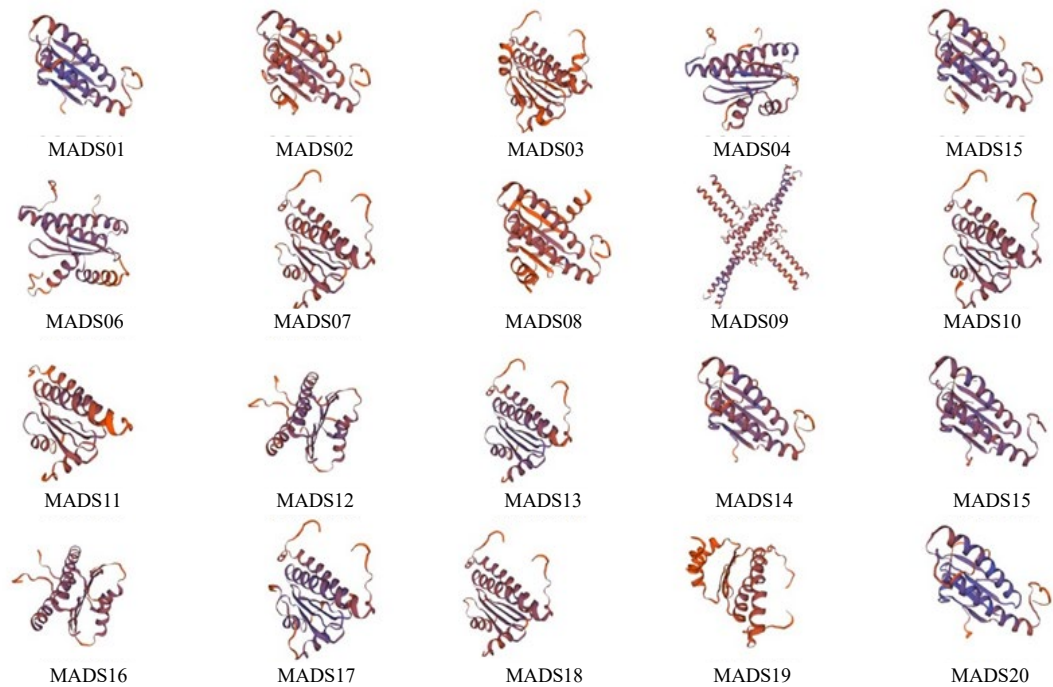


图 1 甘松 MADS-box 转录因子家族三维结构预测

Fig. 1 Tertiary structures prediction of MADS-box transcription factor family in *N. jatamansi*

2.4 甘松 MADS-box 蛋白基序分析

利用 MEME 在线软件分析了甘松 MADS-box 家族蛋白的保守基序，得到 MADS-box 蛋白的 10 个保守基序，命名为 motif1~motif10，长度为 12~

50 aa。
其中 motif1 为 MADS 结构域，motif2 和 motif5 为 K-box 结构域。通过分析发现，20 个 NcMADS-box 蛋白序列含有 2~7 个保守基序（图 2）。

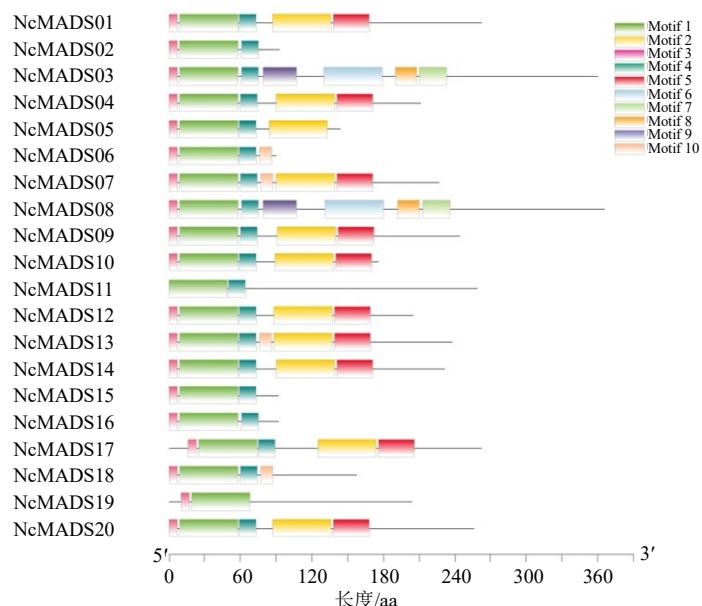


图 2 甘松 MADS-box 转录因子家族 motif 分布

Fig. 2 Motif analysis of the MADS-box transcription factor family in *N. jatamansi*

2.5 甘松 MADS-box 蛋白亚细胞定位

利用 Cello v.2.5 在线数据库对 20 个甘松 MADS-box 蛋白进行亚细胞定位预测分析, 结果表明甘松大部分 MADS-box 蛋白定位在细胞核, 说明它们作为转录因子大部分能调控下游基因的表达^[1], 而少数 MADS-box 蛋白在细胞质、线粒体中也有定位, 其中 2 个定位到细胞质, 2 个定位到线粒体。NcMADS19 定位到细胞质中, NcMADS16 定位到线粒体中; 其余转录因子均定位到细胞核内, 说明大部分 MADS-box 蛋白作为转录因子, 不进行蛋白转运, 直接在细胞核中发挥转录调控作用 (表 3)。

2.6 导肽、信号肽和跨膜结构域预测分析

导肽是将新合成的肽链引入不同细胞器的一段识别序列^[12]。信号肽是导肽的一种, 其位于蛋白质的 N 端, 具有指导分泌型蛋白到内质网上合成的作用。导肽和信号肽在蛋白定位过程中起重要作用。预测分析导肽和信号肽, 利于进一步研究蛋白的功能和作用途径。利用 TargetP2.0 分析 22 个甘松 MADS-box 蛋白是否具有信号肽和导肽结果表明: 20 个 NcMADS 蛋白均不含信号肽和导肽, 为非分泌蛋白。同时利用 TM-HMM Server.2.0 在线软件对其跨膜结构进行分析, 发现 20 个 NcMADS 蛋白均无跨膜结构特征, 这与亲水性的分析结果是一致的。同时表明甘松 22 个 MADS-box 蛋白无法跨膜运输或者作为转录因子通过核孔运输 (表 4)。

2.7 甘松 MADS-box 家族成员系统进化树构建

利用 MEGA6 软件, 对筛选获取的 20 个甘松 MADS-box 的蛋白序列和 TAIR 数据库查询得到的 97 个拟南芥 MADS-box 蛋白序列进行同源对比, 采用邻接法, 构建系统进化树, 在通过 iTOL 进行美化处理。结果显示 (图 3), 20 个甘松 MADS 蛋白聚类可分为 9 个亚族。其中 4 个 Type I 型 NcMADS 蛋白聚类到拟南芥 Type I 型 (Mα 与 Mδ) 中 Mα 中仅包含 NcMADS19, Mδ 中包含 NcMADS02、NcMADS03、NcMADS08; 其余 16 个 NcMADS 蛋白分别聚到拟南芥 MICK 型蛋白的不同亚族。其中 AP3、STK/SHP、FUL、SOC1 亚族各包含 1 个成员, 分别为 NcMADS04、NcMADS05、NcMADS14、NcMADS16; SEP 亚族包含 2 个成员, 分别是 NcMADS09、NcMADS20; SVP 亚族中包含 4 个成员, 分别是 NcMADS06、NcMADS07、NcMADS13、NcMADS18; 还有 6 个成员归属到 ANR1 亚族, 分别是 NcMADS01、NcMADS10、NcMADS11、NcMADS12、NcMADS15 和 NcMADS17。

3 讨论

3.1 甘松 MADS-box 家族转录因子是参与多项生理活动的重要调控因子

MADS-box 家族转录因子参与调控植物生长发育和生殖发育过程中的多种生物学功能, 尤其在花序、花、果实的生长发育过程中起重要作用^[13]。中

表 3 甘松 MADS-box 转录因子家族亚细胞定位

Table 3 Subcellular location prediction of the MADS-box transcription factor family in *N. jatamansi*

编号	预测评分											
	细胞核	细胞质	胞外	线粒体	细胞骨架	高尔基体	叶绿体	内质网	液泡	质膜	过氧化物酶体	溶酶体
NcMADS01	4.344*	0.302	0.134	0.088	0.037	0.029	0.019	0.014	0.011	0.011	0.007	0.005
NcMADS02	2.230*	0.350	0.131	1.530	0.037	0.018	0.589	0.013	0.009	0.063	0.023	0.006
NcMADS03	4.721*	0.096	0.049	0.048	0.040	0.003	0.007	0.006	0.008	0.009	0.008	0.006
NcMADS04	3.459*	0.665	0.094	0.373	0.212	0.073	0.076	0.007	0.005	0.014	0.018	0.003
NcMADS05	2.601*	0.300	0.068	1.737	0.035	0.008	0.194	0.008	0.007	0.012	0.026	0.004
NcMADS06	1.648*	1.047*	0.084	1.003*	0.021	0.035	0.896	0.039	0.014	0.152	0.052	0.009
NcMADS07	3.537*	1.123	0.067	0.062	0.080	0.016	0.051	0.009	0.009	0.029	0.011	0.006
NcMADS08	4.699*	0.094	0.041	0.023	0.022	0.053	0.010	0.010	0.024	0.010	0.005	0.009
NcMADS09	4.277*	0.315	0.056	0.268	0.013	0.006	0.009	0.022	0.020	0.002	0.006	0.007
NcMADS10	3.667*	0.440	0.022	0.641	0.035	0.019	0.136	0.003	0.004	0.023	0.007	0.003
NcMADS11	4.353*	0.334	0.044	0.039	0.071	0.052	0.033	0.008	0.029	0.026	0.007	0.003
NcMADS12	4.346*	0.303	0.018	0.213	0.013	0.018	0.047	0.007	0.007	0.017	0.003	0.009
NcMADS13	4.281*	0.503	0.052	0.051	0.030	0.014	0.020	0.007	0.027	0.008	0.005	0.003
NcMADS14	3.412*	0.623	0.076	0.289	0.347	0.036	0.123	0.026	0.013	0.029	0.020	0.007
NcMADS15	2.268*	0.401	0.086	1.317	0.025	0.020	0.684	0.022	0.010	0.133	0.021	0.013
NcMADS16	1.278	0.688	0.403	2.046*	0.018	0.025	0.202	0.032	0.023	0.221	0.049	0.014
NcMADS17	4.087*	0.259	0.149	0.310	0.047	0.029	0.071	0.009	0.011	0.013	0.011	0.004
NcMADS18	2.908*	0.351	0.172	0.751	0.028	0.018	0.165	0.013	0.012	0.520	0.053	0.009
NcMADS19	0.963	2.573*	0.035	0.321	0.022	0.092	0.554	0.206	0.063	0.039	0.121	0.011
NcMADS20	4.387*	0.331	0.059	0.121	0.013	0.007	0.019	0.008	0.016	0.005	0.026	0.008

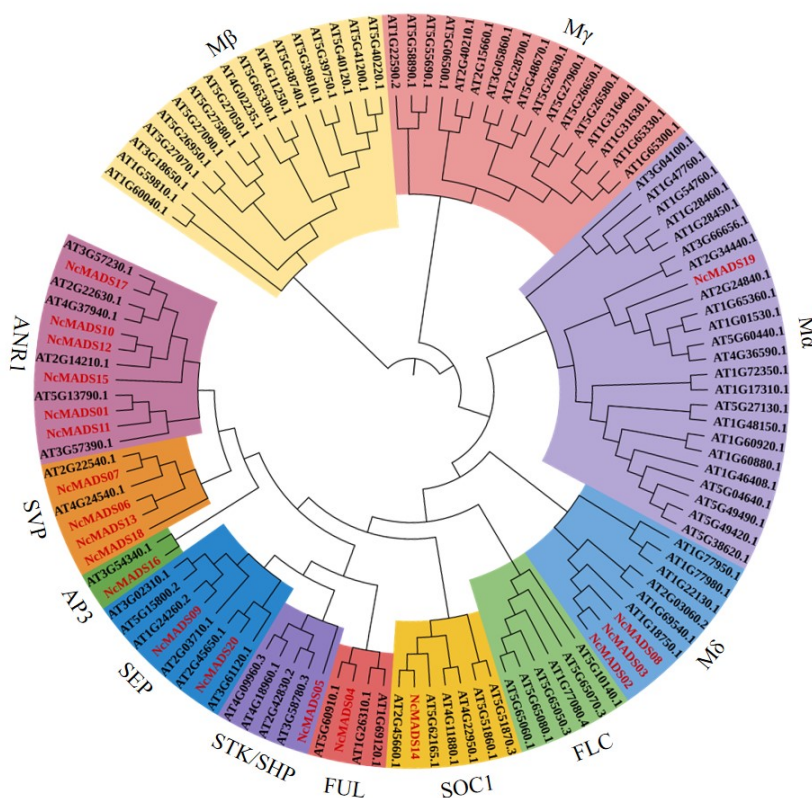
*表示该蛋白存在于该亚细胞的概率较大

*indicates that the protein is more likely to exist in the subcellular

表 4 甘松 MADS-box 转录因子导肽预测

Table 4 Leader peptide prediction of MADS-box transcription factors in *N. jatamansi*

编号	预测评分				
	信号肽	线粒体转运肽	叶绿体转运肽	类囊体腔转运肽	跨膜结构
NcMADS01	0.000 1	0.000 0	0.000 0	0.000 0	无
NcMADS02	0.000 3	0.007 8	0.000 0	0.000 0	无
NcMADS03	0.000 1	0.000 5	0.000 0	0.000 0	无
NcMADS04	0.000 0	0.000 2	0.000 0	0.000 0	无
NcMADS05	0.000 2	0.000 5	0.000 0	0.000 0	无
NcMADS06	0.000 2	0.000 4	0.000 0	0.000 0	无
NcMADS07	0.000 0	0.000 5	0.000 0	0.000 0	无
NcMADS08	0.000 1	0.000 5	0.000 0	0.000 0	无
NcMADS09	0.000 0	0.000 0	0.000 0	0.000 0	无
NcMADS10	0.000 2	0.001 2	0.000 0	0.000 0	无
NcMADS11	0.000 1	0.000 0	0.000 0	0.000 0	无
NcMADS12	0.000 1	0.000 8	0.000 0	0.000 0	无
NcMADS13	0.000 0	0.000 0	0.000 0	0.000 0	无
NcMADS14	0.000 1	0.001 3	0.000 0	0.000 0	无
NcMADS15	0.000 3	0.001 7	0.000 0	0.000 0	无
NcMADS16	0.001 7	0.002 9	0.000 0	0.000 0	无
NcMADS17	0.005 6	0.052 5	0.016 9	0.005 6	无
NcMADS18	0.000 1	0.001 4	0.000 0	0.000 0	无
NcMADS19	0.000 0	0.000 0	0.000 0	0.000 0	无
NcMADS20	0.000 0	0.000 0	0.000 0	0.000 0	无



图中不同物种 MADS-box 蛋白用不同颜色标记，红色代表甘松，黑色代表拟南芥

MADS-box proteins of different species were marked with different colors, red represents *N. jatamansi*, and black represents *Arabidopsis thaliana*

图 3 甘松与拟南芥 MADS-box 家族转录因子系统进化树

Fig. 3 Phylogenetic tree of MADS-box family transcription factors in *N. jatamansi* and *Arabidopsis thaliana*

药火麻仁中共鉴定出 39 个 MADS-box 家族转录因子，其参与调控火麻仁的营养和生殖结构、开花时间、花器官形成^[14]，茶树中 MADS-box 蛋白 CsGLO1 和 CsGLO2 可形成二聚体调控第 2 轮花瓣和第 3 轮雄蕊的发育^[15]。马铃薯中 XAL1 蛋白参与调控其块茎的发育^[16]。甘松 MADS-box 家族转录因子尚未见报，本研究从甘松中鉴别出 20 个功能未知的 MADS-box 家族转录因子，首次全面分析了甘松的 MADS-box 家族转录因子，有利于进一步对甘松 MADS-box 家族转录因子进行功能分析及后续验证。

与拟南芥相比，甘松 MADS-box 蛋白相对较少。一方面与所测转录组序列完整度有关，另一方面可能与甘松 MADS-box 家族转录因子的重复率较高或重复后拼接丢失率较高有关。这一结果表明，不同物种间 MADS-box 家族转录因子的差异较大，导致不同物种间 MADS-box 家族转录因子不同，各自的功能也差别较大，具有不同的进化限制^[17]。

3.2 甘松 MADS-box 家族转录因子与其药用价值的息息相关

抽薹开花是植物从营养生长到生殖生长的关键环节，通常称为成花转变^[18]。多数以根茎入药的药材抽薹开花后药效显著降低。如当归抽薹过程中，多个生理指标发生变化，开花后，营养物质大量消耗，根部木质化，丧失药用价值^[19]。防风以干燥根入药，抽薹前后，其主要有效成分人参炔醇含量逐渐自根部向地上部分转移，随后根部开始木质化并中空腐烂，药用价值显著降低^[20]。研究表明，MADS-box 家族中转录因子较多与植物抽薹开花相关。模式植物拟南芥中 MADS-box 家族中 AP1 蛋白能调控与抽薹相关的结构基因，与拟南芥 AP1 蛋白同源的春甘蓝 AP1 蛋白也对植株抽薹有调控作用^[21]，十字花科植物芥菜中 SOC1 蛋白可直接作用于抽薹的决定基因 *LFY*，此外 SVP 蛋白结合 SOC1 能调控 AP1 蛋白从而会影响其抽薹。MADS-box 家族转录因子多参与调控植物抽薹过程以及开花时

间^[22]。莴苣 LsRGL1 蛋白可通过调控赤霉素途径控制抽薹开花时间^[18], LSMADS16 和 LSMADS37 蛋白对莴苣高温抽薹具有负调控作用。甘松以干燥根及根茎入药, 也面临抽薹开花后有效成分含量降低的问题, 本研究鉴定出 20 个甘松 MADS-box 家族转录因子, 其中 NcMADS14 与拟南芥 SOC1 同源, NcMADS06、NcMADS07、NcMADS13、NcMADS18 与拟南芥 SVP 蛋白同源, 推测这 5 个蛋白可能参与调控甘松抽薹开花过程, 为明晰甘松 MADS-box 家族转录因子如何调控其抽薹开花时间, 减慢甘松根茎木质化过程, 保证药材药效提供理论基础。

3.3 甘松 MADS-box 家族转录因子参与调控花器官发育与药用部位的优质生产

MADS-box 家族转录因子在控制植物开花时间及花器官发育中起重要作用^[23], 部分参与调控植物根的发育。2001 年, Theissen^[24]提出了四聚体模型即 ABCDE 模型: A 类和 E 类基因控制花萼的形成, A 类、B 类和 E 类基因共同控制花瓣的形成, B 类、C 类和 E 类基因共同控制雄蕊的形成, 而心皮由 C 类和 E 类基因共同控制, 胚珠由 C 类、D 类和 E 类基因共同控制。系统发育分析发现, NcMADS04 与拟南芥 A 类基因 *FUL* 同源性高, 其可能与甘松花萼形成相关, NcMADS16 与拟南芥 B 类基因 *AP3* 同源性高, 推测 NcMADS16 可能参与调控甘松花瓣与雄蕊的形成过程。NcMADS05 与拟南芥 D 类基因 *SHP* 同源性高, 可能参与甘松心皮的形成过程。NcMADS09、NcMADS20 与拟南芥 E 类基因 *SEP* 聚为一类, 则二者可能为甘松的 E 类基因, 与 ABCD 类基因共同参与调控花器官的形成过程。本研究通过同源比对发现, 甘松 20 个 MADS-box 家族中未有与拟南芥同源的 C 类基因, 可能是转录组拼接错误导致甘松 C 类基因缺失, 需继续试验验证。

SOC1 为开花整合因子, 可通过整合光周期途径、春化途径、赤霉素途径、自主途径、温度途径和年龄途径等相关开花信号来控制开花时间^[25], 已有研究表明脐橙中 *CsSOC1* 基因能响应低温信号, 表达量上调, 促进脐橙成花^[26], 胡萝卜 *DcSOC1-1* 是光周期成花途径的整合因子, 长日照使胡萝卜抽薹延迟^[27]。本研究发现 NcMADS14 与拟南芥 SOC1 同源, 推测其具有与 SOC1 类似的功能。

NcMADS06、NcMADS07、NcMADS13、NcMADS18 与拟南芥 SVP 蛋白同源, SVP 为开花抑制因子^[28], 可直接结合在开花途径整合因子

SOC1 的启动子上, 从而调节 *SOC1* 的转录, 抑制其表达, 延迟抽薹开花^[29]。若将甘松中与拟南芥 *SOC1*、*SVP* 同源的基因合理地通过基因工程手段改造植株, 则可获得晚开花的植株。以延长甘松采收时期, 加大生产。

甘松以根及根茎入药, 本课题组前期研究表明, 仿野生栽培甘松须根根尖数明显多于野生品, 且主根变细, 严重影响仿野生栽培甘松的品质^[30], *ANR1* 是在根中特异性表达的基因, 也是 MADS-box 家族基因中第一个被鉴定的唯一通过 NO_3^- 对信号调控侧根形成^[31]。研究表明, *ANR1* 在拟南芥根中特异性表达, 菊花中 *CmANR1* 通过同源/异源二聚体调节侧根发育^[32]。*ANR1* 在水稻中的同源基因 *OsMADS25* 在 NO_3^- 存在时过表达, 能显著促进水稻主根和侧根生长^[10]。系统发育分析结果显示, 甘松 MADS-box 家族基因转录因子中 NcMADS01、NcMADS10、NcMADS11、NcMADS12、NcMADS15、NcMADS17 与拟南芥 *ANR1* 同源性高, 可能参与调控甘松根的生长发育过程。后期可通过探明其如何参与根发育调控, 从而减少须根数, 保持主根粗壮, 从而获得根及根茎性状符合药用规定且统一的优良甘松品种。

4 结论

甘松 MADS-box 家族转录因子的鉴定和挖掘有待进一步完善。本研究通过了解甘松 MADS-box 家族转录因子的生物信息学, 预测其生物学功能及参与的生理过程, 为后续试验奠定基础, 也为甘松的分子生物学研究提供一定的支持。

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]. 四部. 2020: 123.
- [2] 贾敏如, 张艺. 中国民族药辞典 [Z]. 北京: 中国医药科技出版社, 2016: 552.
- [3] 宇妥·元丹贡布等著. 李永年译. 四部医典 [M]. 北京: 人民卫生出版社, 1983: 93-96.
- [4] Schwarz-Sommer Z, Huijser P, Nacken W, et al. Genetic control of flower development by homeotic genes in *Antirrhinum majus* [J]. *Science*, 1990, 250(4983): 931-936.
- [5] de Bodt S, Raes J, van de Peer Y, et al. And then there were many: MADS Goes genomic [J]. *Trends Plant Sci*, 2003, 8(10): 475-483.
- [6] Tiwari S, Spielman M, Schulz R, et al. Transcriptional profiles underlying parent-of-origin effects in seeds of

- Arabidopsis thaliana* [J]. *BMC Plant Biol*, 2010, 10: 72.
- [7] Masiero S, Colombo L, Grini P E, *et al.* The emerging importance of type I MADS box transcription factors for plant reproduction [J]. *Plant Cell*, 2011, 23(3): 865-872.
- [8] Kwantes M, Liebsch D, Verelst W. How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes [J]. *Mol Biol Evol*, 2012, 29(1): 293-302.
- [9] 李成儒, 董钠, 李笑平, 等. 兰科植物花发育调控 MADS-box 基因家族研究进展 [J]. 园艺学报, 2020, 47(10): 2047-2062.
- [10] 于春燕. ANR1 在水稻中的同源基因 OsMADS25、OsMADS27 和 OsMADS57 的克隆和功能分析 [D]. 杭州: 浙江大学, 2015.
- [11] 黄文娟, 许鑫, 陈靳松, 等. 川白芷 NAC 家族的生物信息学及表达模式分析 [J]. 中国中药杂志, 2021, 46(7): 1769-1782.
- [12] 董娇, 周军, 辛培尧, 等. 不同植物 LDOX/ANS 基因的生物信息学分析 [J]. 基因组学与应用生物学, 2010, 29(5): 815-822.
- [13] Becker A, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants [J]. *Mol Phylogenet Evol*, 2003, 29(3): 464-489.
- [14] 万志庭, 鲁梦, 吴沙沙, 等. 中药火麻仁基原植物大麻 MIKC 型 MADS-box 基因家族鉴定与表达分析 [J]. 药学学报, 2021, 56(11): 3173-3183.
- [15] 靳春梅, 周坤, 张今今. 茶树花发育 MADS-box 转录因子 CsGLO1、CsGLO2 与 CsAG 之间的互作关系研究 [J]. 植物科学学报, 2017, 35(1): 79-86.
- [16] Gao H H, Wang Z M, Li S L, *et al.* Genome-wide survey of potato MADS-box genes reveals that StMADS1 and StMADS13 are putative downstream targets of tuberigen StSP6A [J]. *BMC Genomics*, 2018, 19(1): 726.
- [17] Wang Y S, Zhang J L, Hu Z L, *et al.* Genome-wide analysis of the MADS-box transcription factor family in *Solanum lycopersicum* [J]. *Int J Mol Sci*, 2019, 20(12): 2961.
- [18] 郝新迪, 许梦男, 董波波, 等. 基于转录组信息的叶用莴苣 MADS-box 转录因子家族分析 [J]. 北京农学院学报, 2021, 36(4): 14-19.
- [19] 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.
- [20] 滕虹伯, 赵岩, 孙夏霖, 等. 防风抽薹期人参炔醇含量的变化分析 [J]. 时珍国医国药, 2022, 33(2): 450-453.
- [21] 王神云, 刘振宁, 李建斌, 等. 春甘蓝抽薹的生理与相关基因转录分析 [J]. 分子植物育种, 2011, 9(5): 605-610.
- [22] Ning K, Han Y Y, Chen Z J, *et al.* Genome-wide analysis of MADS-box family genes during flower development in lettuce [J]. *Plant Cell Environ*, 2019, 42(6): 1868-1881.
- [23] Ma H. The unfolding drama of flower development: Recent results from genetic and molecular analyses [J]. *Genes Dev*, 1994, 8(7): 745-756.
- [24] Theissen G. Development of floral organ identity: Stories from the MADS house [J]. *Curr Opin Plant Biol*, 2001, 4(1): 75-85.
- [25] Lee J, Lee I. Regulation and function of SOC1, a flowering pathway integrator [J]. *J Exp Bot*, 2010, 61(9): 2247-2254.
- [26] 贺新兴, 杨杰, 高钱辉, 等. 脐橙 SOC1 基因的克隆及其在脐橙花发育过程中的表达分析 [J]. 分子植物育种, 2022, 20(15): 4948-4957.
- [27] 刘莉洁, 欧承刚, 赵志伟, 等. 胡萝卜 SOC1 同源基因对光周期的响应及其功能 [J]. 园艺学报, 2016, 43(6): 1099-1106.
- [28] Ahmar S. Development of mutants with varying flowering times by targeted editing of multiple SVP gene copies in *Brassica napus* L [J]. *Crop J*, 2022, 10(1): 67-74.
- [29] Helliwell C A, Wood C C, Robertson M, *et al.* The *Arabidopsis* FLC protein interacts directly *in vivo* with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex [J]. *Plant J*, 2006, 46(2): 183-192.
- [30] 李莹, 钟海蓉, 李奕松, 等. 基于高通量测序研究甘松根际细菌及药用部位内生细菌群落结构及多样性 [J]. 中草药, 2022, 53(3): 835-847.
- [31] Zhang H, Forde B G. An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture [J]. *Science*, 1998, 279(5349): 407-409.
- [32] Sun C H, Yu J Q, Wen L Z, *et al.* *Chrysanthemum* MADS-box transcription factor CmANR1 modulates lateral root development via homo-/heterodimerization to influence auxin accumulation in *Arabidopsis* [J]. *Plant Sci*, 2018, 266: 27-36.

[责任编辑 时圣明]