

• 药材与资源 •

大麻 GRAS 转录因子全基因组研究

张苗苗¹, 于江珊¹, 施江¹, 杨雨¹, 孟雪¹, 孙伟², 万会花^{2*}, 薛建平^{1*}

1. 淮北师范大学安徽省特色资源植物利用工程实验室, 安徽 淮北 235000

2. 中国中医科学院中药研究所, 北京 100700

摘要: 目的 GRAS 转录因子在植物响应逆境胁迫和调控次生代谢方面起重要作用。为了研究 GRAS 转录因子在大麻素生物合成途径中的潜在功能, 对大麻 *Cannabis sativa* 基因组中的 GRAS 基因进行全基因组鉴定、表达模式和功能分析。方法 使用 NCBI、PlantTFDB、ExPASy 等在线网站以及 TBtools、MEGA、DNAMAN 等软件对 CsGRAS 基因进行鉴定、可视化分析和功能预测。结果 在大麻基因组中共鉴定出 44 个 GRAS 基因, 在 10 条染色体上不均匀分布, 基因之间存在串联重复现象; CsGRAS 基因编码氨基酸数目为 436~757, 等电点介于 4.77~7.24, 相对分子质量为 49 164.54~85 748.52; 亚细胞定位分析显示 CsGRAS 主要定位在细胞核中; 系统发育分析将 CsGRAS 分为 10 个亚家族; CsGRAS 基因启动子区含有光响应元件、GA 响应元件和 MeJA 响应元件等; CsGRAS 在不同品种不同组织中表达量差异显著, 其中 CsGRAS4、CsGRAS13、CsGRAS15、CsGRAS17 和 CsGRAS44 在花中特异性高表达, 推测其可能调控大麻素生物合成。结论 全面分析了大麻基因组中的 GRAS 家族, 有助于更好地挖掘 GRAS 基因在大麻的大麻素生物合成调控和响应逆境胁迫中的作用。

关键词: 大麻; GRAS 基因家族; 基因表达; 系统进化; 功能预测

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2021)05-1423-11

DOI: 10.7501/j.issn.0253-2670.2021.05.024

Study on GRAS transcription factors of cannabis genome

ZHANG Miao-miao¹, YU Jiang-shan¹, SHI Jiang¹, YANG Yu¹, MENG Xue¹, SUN Wei², WAN Hui-hua², XUE Jian-ping¹

1. Plant Utilization Engineering Laboratory of Anhui Characteristic Resources, Huaibei Normal University, Huaibei 235000, China

2. Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

Abstract: Objective In plants, GRAS transcription factors play important roles in stress response and regulation of secondary metabolism. In this study, to investigate the potential roles of GRAS transcription factors in cannabinoid biosynthesis, genome-wide identification and expression and functional analysis of GRAS genes family in hemp (*Cannabis sativa*) were performed. **Methods** Identification, visualized analysis and function prediction of CsGRAS genes were carried out by online websites such as NCBI, PlantTFDB, ExPASy and several softwares including TBtools, MEGA and DNAMAN. **Results** A total of 44 CsGRAS genes were identified in hemp genome, which were unevenly distributed on 10 chromosomes and displayed tandem duplication between genes. The number of amino acids encoded by CsGRAS genes was between 436 and 757, the isoelectric point of which was between 4.77 and 7.24, and the molecular weight of which was between 49 164.54 and 85 748.52. Subcellular localization analysis showed that most of the CsGRAS were located in the nucleus. CsGRAS genes were divided into 10 subfamilies by phylogenetic tree analysis. Light response elements, GA response elements and MeJA response elements were found in the promoter region of CsGRAS genes. The expression profiles of CsGRAS were significantly different in varied cultivars and tissues. CsGRAS4, CsGRAS13, CsGRAS15, CsGRAS17, and CsGRAS44 specifically and highly expressed in flowers, suggesting that these CsGRAS may play key roles in regulating cannabinoid biosynthesis. **Conclusion** This study comprehensively analyzed the GRAS family in hemp genome, which provides significant insights into the potential function of GRAS gene in the regulation of cannabinoid biosynthesis and stress response in hemp.

Key words: *Cannabis sativa* L.; GRAS gene family; gene expression; phylogenetic evolution; functional prediction

收稿日期: 2020-08-06

基金项目: 中央级公益性科研院所基本科研业务费专项 (ZXKT19019); 中央级公益性科研院所基本科研业务费专项 (ZZ11-096)

作者简介: 张苗苗, 硕士研究生, 从事药用植物生物工程研究。E-mail: 1747244078@qq.com

*通信作者: 薛建平, 男, 硕士生导师, 教授, 研究方向为中药资源学。E-mail: xuejp@163.com

万会花, 女, 助理研究员, 博士, 研究方向为药用植物资源创新与应用。E-mail: hhwan@icmm.ac.cn

转录因子的研究是功能基因组学的重要组成部分。自从第一个转录因子在玉米中被发现^[1],至今大量的转录因子已经被证明参与高等植物的各种生理过程和调控网络。GRAS 蛋白是一个重要的植物特异性蛋白家族,是以该家族中最早鉴定的 3 个成员 GAI (gibberellic-acid insensitive)、RGA (repressor of GAI) 和 SCR (scarecrow)^[2-3]命名。GRAS 转录因子作为植物所特有的因子,参与植物的信号转导、生长发育、逆境胁迫等过程。该家族蛋白具有保守的羧基末端,其中包括 LRI、VHIID、LRII、PFYRE 和 SAW 5 个基序,在 N 端各不相同的序列是基因功能特异性的主要原因。到目前为止,GRAS 基因家族已经在多种植物中进行了全基因组研究,包括杨树、拟南芥、水稻、大白菜、梅花、松树等^[4-7]。在模式植物拟南芥和水稻基因组中分别有 34、60 个 GRAS 基因家族成员。以拟南芥 GRAS 家族成员的分类为依据,将大麻 GRAS 基因家族分为 10 类。GRAS 基因家族的代表基因是 DELLA 基因,它是环境与激素,激素与激素信号途径之间的联系枢纽,如负调节赤霉素 (gibberellin, GA) 信号通路、脱落酸信号传导和光信号传导^[8]等,同时还参与植物次生代谢产物的合成调控,例如在缺氮条件下,DELLA 蛋白促进花青素的合成和积累^[9]。王立儒^[10]研究发现丹参 SmDELLA1 蛋白可能是丹参总酚酸和总黄酮生物合成途径中的正调控因子。

大麻 *Cannabis sativa* L. 又名火麻,是大麻科 (Cannabinaceae) 大麻属 *Cannabis* L. 一年生草本植物^[11],具有很高的药用价值和商业价值。我国是最早利用大麻的国家,也是大麻原产地之一^[12]。大麻以合成大麻素而闻名,大麻素是具有多种药理活性的萜类物质,主要积累在雌花的腺毛中^[13]。大麻二酚 (cannabidiol, CBD) 具有阻断某些多酚对人体神经系统的影响,并且有止痛抗炎、抗癫痫、抗焦虑、抗类风湿性关节炎和抗失眠等一系列生理活性功能,对治疗多发性硬化症具有良好的效果^[14]。随着高纯度大麻二酚液体制剂 Epidiolex® 药物在美国上市,并且成为治疗小儿癫痫的“孤儿药”,大麻素的市场需求量迅速增加。具有精神活性的四氢大麻酚 (tetrahydrocannabinol, THC) 还具有抗癌、抗菌、抗呕吐、利尿等作用,但四氢大麻酚也会导致焦虑、认知功能障碍和降低免疫力,是被严格管控的物质^[15]。大麻环萜酚 (cannabichromen, CBC) 具有消炎等作用,是大麻幼苗期主要的大麻素成分,但随着植株的逐渐成熟大麻

环萜酚的含量迅速降低,以至于可以忽略不计^[16]。李秋实等^[12]明确了大麻种质资源管理三级分类体系,根据 THC 的含量将大麻分为工业大麻、药用大麻和医用大麻 3 类。THC<0.3% 时为药用大麻和工业大麻,其中 CBD 含量高为药用大麻, CBD 含量低为工业大麻,用于获取纤维和种子; THC>0.3% 时为医用大麻,严禁种植和使用。

大麻素的前体物质来源于甲基赤藓糖醇-4-磷酸 (methylerythritol 4-phosphate, MEP) 和脂肪酸 2 个代谢途径。首先,在 MEP 途径,顺式异戊烯己基二磷酸 (cis-isopentenyl diphosphate, IPP) 在 IPP 异构酶以及 GPP 合成酶作用下形成了香叶基二磷酸 (geranyl diphosphate, GPP)。同时,脂肪酸类合成途径中丙二酰辅酶 A (malonyl-CoA) 在橄榄酸环化酶 (olivetolic acid cyclase, OAC) 和橄榄醇合成酶 (olivetol synthase, OLS) 的作用下形成 2,4-二羟基-6-戊基苯甲酸 (olivetolic acid, OA)。然后,底物 OA 和 GPP 在大麻醇酸合成酶 (cannabigerolic acid synthase, CBGAS) 的作用下合成了大麻素生物合成的主要前体物质大麻醇酸 (cannabigerolic acid, CBGA), CBGA 在四氢大麻酸合成酶 (tetrahydrocannabinolic acid synthase, THCAS)、大麻二酸合成酶 (cannabidiolic acid synthase, CBDAS) 和大麻黄酸合成酶 (cannabichromenic acid synthase, CBCAS) 的催化作用下分别形成了四氢大麻酚酸 (tetrahydrocannabinolic acid, THCA)、大麻二酚酸 (cannabidiolic acid, CBDA) 和大麻色烯酸 (cannabichromenic acid, CBCA)。最后在光和热的物理作用下发生脱羧基反应分别生成了四氢大麻酚、大麻二酚和大麻环萜酚。

随着大麻全基因组序列的公布,对整个基因家族的综合分析变得便捷。由于 GRAS 蛋白在调控植物次生代谢和逆境胁迫方面起重要作用,但是该基因家族在大麻中的信息缺乏,因此,本研究首次对大麻中 GRAS 基因家族全序列进行了研究。本研究鉴定了 44 个 CsGRAS 基因,并对它们的基因分类、染色体定位、系统发育和外显子-内含子分布进行了分析。此外,采用数据库及本课题组完成的转录组数据,分析大麻 GRAS 基因家族的表达模式,推测参与大麻素生物合成的相关基因。本研究提供了 GRAS 基因家族的详细信息,为进一步研究 GRAS 基因在大麻中的功能奠定了基础。

1 材料与方法

1.1 大麻 GRAS 家族成员的获取和染色体定位

从 NCBI (<https://www.ncbi.nlm.nih.gov>) 网站上获取大麻 (GCA_900626175.1)、拟南芥 (GCA_000005425.2) 和水稻 (GCA_000005425.2) 的全基因组及注释文件。采用的大麻品种为 CRBRx, 性别为雌株。应用 TBtools 软件^[17]提取的大麻蛋白序列, 放入 PlantTFDB 网站进行转录因子预测, 预测得到的所有 GRAS 家族成员去除重复转录本数据, 得到大麻 GRAS 家族的 gene ID, 利用 TBtools 软件获得大麻 GRAS 家族的 gene ID 所对应的蛋白序列。使用在线生物信息学工具 ExPASy-ProSite (<http://web.expasy.org/protparam/>) 对大麻 GRAS 蛋白的分子量和等电点进行预测。利用 Cell-PLoc (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>) 网站进行亚细胞定位预测。使用 TBtools 软件解析大麻的注释文件, 确定每个 GRAS 基因的位置信息并绘制其所对应的染色体物理位置图。

1.2 大麻、水稻和拟南芥 GRAS 家族共线性分析

使用从 NCBI 网站上下载的大麻、拟南芥及水稻全基因组及注释文件, 利用 MEGA-X 软件, 采取邻接法 (neighbor-joining, NJ) 法, 对拟南芥和大麻 GRAS 蛋白构建无根系统发育树。使用 TBtools 软件对其三者共线性关系进行可视化, 并找到 3 者 GRAS 基因家族的同源基因对。

1.3 大麻 GRAS 蛋白系统发育树的构建及 motif 发掘与分析

使用 MEGA-X 通过 NJ 方法构建了基于比对的系统发育树, 并进行了重复 1000 次的 Bootstrap 检验, 所有的设定均采用默认值。大麻 GRAS 蛋白质 motif 的发掘与位置分布均采用 MEME (<http://meme-suite.org/tools/meme>) 蛋白质序列分析在线程序, 并设定目标 motif 数量为 10。通过使用 TBtools 软件对大麻的 GRAS 注释文件的解析, 对大麻 GRAS 基因家族进行了外显子-内含子、保守结构域和 motif 3 者可视化处理, 并使用 DNAMAN 软件进行保守结构域蛋白序列的比对分析。

1.4 大麻 GRAS 基因家族顺式作用元件分析

使用 PlantCare (<http://www.plantcare.co.uk/>) 网站对从 TBtools 软件提取的大麻 GRAS 基因上游 2000 bp 序列进行顺式作用元件预测, 并使用 TBtools 软件对其可视化构图。

1.5 大麻 GRAS 基因的表达模式分析

利用本课题组 ZYS 品种 (由 Purple Kush 和 Dinamed Autoflowering CBD 杂交获得的高 CBD、低 THC 品种) 大麻花、茎和叶的转录组数据, 使用 TBtools 软件绘制大麻 GRAS 家族基因表达模式热图, GRAS 成员基因表达量 (FPKM) 经 row scale 均一化处理, 并进行聚类 and 差异表达分析。

2 结果与分析

2.1 大麻 GRAS 基因的鉴定与染色体分布

从 NCBI (<https://www.ncbi.nlm.nih.gov>) 网站上获取 CRBRx 品种的大麻全基因组及注释文件, 将 TBtools 软件提取的大麻蛋白序列, 放入 PlantTFDB 网站进行转录因子的预测, 预测得到 54 个 GRAS, 去除 10 个重复转录本数据后, 最终得到 44 个 GRAS 家族成员, 命名为 CsGRAS1 ~ CsGRAS44。利用大麻 GRAS 家族的 gene ID, 使用 TBtools 软件获得大麻 GRAS 家族的 gene ID 所对应的蛋白序列。大麻 GRAS 蛋白长度最小的是 CsGRAS8, 由 436 个氨基酸残基组成, 长度最大的蛋白质为 CsGRAS21, 包含 757 个氨基酸残基。大麻 GRAS 蛋白质等电点介于 4.77~7.24, 相对分子质量介于 49 164.54~85 748.52 (表 1)。

GRAS 基因在染色体上的位置分布如图 1 所示, 该家族 44 个基因不均匀地分布在大麻的 10 条染色体上 (NC_044370.1、NC_044371.1、NC_044372.1、NC_044373.1、NC_044374.1、NC_044375.1、NC_044376.1、NC_044377.1、NC_044378.1、NC_044379.1 分别分布了 3、6、3、6、5、7、5、4、1、4 个 GRAS 基因), 其中 NC_044375.1 染色体上 GRAS 基因家族数目最多, 含有 7 个基因, 其次是 NC_044371.1 和 NC_044373.1 都含有 6 个基因, 最少的是 NC_044378.1 只含有 1 个基因。最长的 CsGRAS33 基因在 NC_044374.1 染色体下臂, 最短的 CsGRAS7 在 NC_044376.1 染色体上臂。NC_044371.1 染色体上的 CsGRAS34 和 CsGRAS35, CsGRAS43 和 CsGRAS44; NC_044372.1 染色体上的 CsGRAS2 和 CsGRAS3; NC_044373.1 染色体上的 CsGRAS38 和 CsGRAS40; NC_044374.1 染色体上的 CsGRAS12 和 CsGRAS13; NC_044375.1 染色体上的 CsGRAS20、CsGRAS21、CsGRAS22、CsGRAS23、CsGRAS24、CsGRAS25 和 CsGRAS26; NC_044376.1 染色体上的 CsGRAS6、CsGRAS7、CsGRAS8 和 CsGRAS9 成簇排列, 序列比

表 1 大麻 GRAS 基因家族基本信息及特征

Table 1 Information and characteristics of CsGRAS genes

基因号	转录本登录号	基因名称	大小/aa	等电点	相对分子质量	亚细胞定位
LOC115710262	rna-XM_030638621.1	CsGRAS2	514	4.93	57 833.13	细胞核
LOC115711337	rna-XM_030639675.1	CsGRAS3	467	5.82	52 554.66	叶绿体
LOC115712712	rna-XM_030641050.1	CsGRAS4	613	5.08	67 527.83	细胞核
LOC115696991	rna-XM_030623887.1	CsGRAS5	458	4.77	52 006.96	叶绿体
LOC115724160	rna-XM_030653624.1	CsGRAS6	636	5.20	73 211.22	叶绿体
LOC115723500	rna-XM_030652996.1	CsGRAS7	454	6.00	51 420.59	叶绿体
LOC115723133	rna-XM_030652562.1	CsGRAS8	436	5.58	49 164.54	叶绿体
LOC115722004	rna-XM_030651120.1	CsGRAS9	577	4.94	66 178.24	叶绿体
LOC115711473	rna-XM_030639814.1	CsGRAS10	549	6.13	61 322.10	细胞核
LOC115715793	rna-XM_030644465.1	CsGRAS11	628	5.08	71 316.90	细胞核
LOC115716239	rna-XM_030644994.1	CsGRAS12	617	6.03	68 009.17	细胞核
LOC115716409	rna-XM_030645194.1	CsGRAS13	616	6.19	67 882.02	细胞核
LOC115725491	rna-XM_030655030.1	CsGRAS14	756	6.16	85 229.91	细胞核
LOC115724919	rna-XM_030654291.1	CsGRAS15	548	5.14	61 008.73	细胞核
LOC115702418	rna-XM_030629883.1	CsGRAS16	528	5.24	57 541.56	细胞核
LOC115724710	rna-XM_030654047.1	CsGRAS17	577	6.05	64 503.06	细胞核
LOC115710247	rna-XM_030638634.1	CsGRAS18	582	4.92	64 716.73	细胞核
LOC115717012	rna-XM_030645940.1	CsGRAS19	697	5.36	77 185.61	细胞核
LOC115719730	rna-XM_030648898.1	CsGRAS20	698	5.68	78 688.43	细胞核
LOC115719656	rna-XM_030648776.1	CsGRAS21	757	5.10	85 748.52	细胞核
LOC115719316	rna-XM_030648302.1	CsGRAS22	749	5.03	84 965.93	细胞核
LOC115718811	rna-XM_030647633.1	CsGRAS23	645	6.73	73 634.61	叶绿体
LOC115718867	rna-XM_030647676.1	CsGRAS24	621	5.69	71 491.35	叶绿体
LOC115718866	rna-XM_030647675.1	CsGRAS25	630	7.24	72 454.10	叶绿体
LOC115718865	rna-XM_030647673.1	CsGRAS26	746	5.80	85 106.83	细胞核
LOC115700662	rna-XM_030628281.1	CsGRAS27	455	5.66	49 596.06	细胞核
LOC115723037	rna-XM_030652434.1	CsGRAS28	575	5.14	64 688.92	细胞核
-LOC115698980	rna-XM_030626193.1	CsGRAS29	494	5.15	54 901.55	质膜
LOC115701004	rna-XM_030628687.1	CsGRAS30	454	6.68	50 178.21	细胞核
LOC115698988	rna-XM_030626201.1	CsGRAS31	724	5.99	81 047.24	细胞核
LOC115712444	rna-XM_030640719.1	CsGRAS32	497	5.28	56 735.16	叶绿体
LOC115716431	rna-XM_030645221.1	CsGRAS33	752	6.29	82 128.85	叶绿体
LOC115705815	rna-XM_030633255.1	CsGRAS34	722	5.72	78 817.91	细胞核
LOC115705915	rna-XM_030633390.1	CsGRAS35	721	5.74	78 730.89	细胞核
LOC115705736	rna-XM_030633154.1	CsGRAS36	722	5.77	78 760.91	细胞核
LOC115712525	rna-XM_030640813.1	CsGRAS37	504	5.68	56 355.13	质膜
LOC115714408	rna-XM_030643106.1	CsGRAS38	600	5.95	66 439.36	质膜
LOC115706185	rna-XM_030633743.1	CsGRAS39	539	5.51	59 352.41	细胞核
LOC115714404	rna-XM_030643094.1	CsGRAS40	622	5.33	68 325.01	细胞核
LOC115716331	rna-XM_030645098.1	CsGRAS41	700	5.99	76 887.41	细胞核
LOC115712875	rna-XM_030641289.1	CsGRAS42	508	6.30	56 609.53	质膜
LOC115703632	rna-XM_030630867.1	CsGRAS43	559	6.21	62 215.78	细胞核
LOC115707313	rna-XM_030635227.1	CsGRAS44	544	6.33	61 258.62	细胞核

对分析发现这些成簇的基因之间存在串联重复现象。

2.2 大麻 GRAS 家族进化的特性

利用 MEGA-X (图 2) 软件, 采取 NJ 法, 对拟南芥 34 个和大麻 44 个 GRAS 蛋白进行分析, 构建无根系统发育树。根据系统发育树, GRAS 蛋白可分为 10 个亚家族: HAM、LS、DLT、SCR、DELLA、SCL3、LISCL、SHR、PAT1 和 HAMlike。HAMlike

亚家族包含 5 个 CsGRAS 蛋白 (CsGRAS6、CsGRAS7、CsGRAS8、CsGRAS9、CsGRAS11), 与 HAM 属于同一分支却不包含拟南芥 GRAS 蛋白。HAM 亚家族有 8 个 CsGRAS 蛋白 (CsGRAS5、CsGRAS16、CsGRAS32、CsGRAS34、CsGRAS35、CsGRAS36、CsGRAS38、CsGRAS41)。有 5 个 CsGRAS 蛋白 (CsGRAS2、CsGRAS4、CsGRAS10、CsGRAS31、CsGRAS39) 属于 DELLA 亚家族。

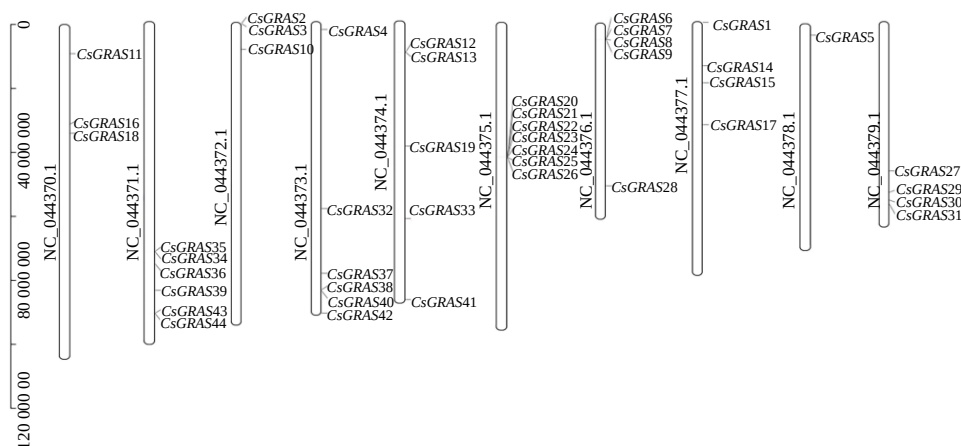


图 1 大麻 44 个 GRAS 基因在 10 条染色体的分布

Fig. 1 Distribution of 44 GRAS genes on 10 chromosomes in *C. sativa*

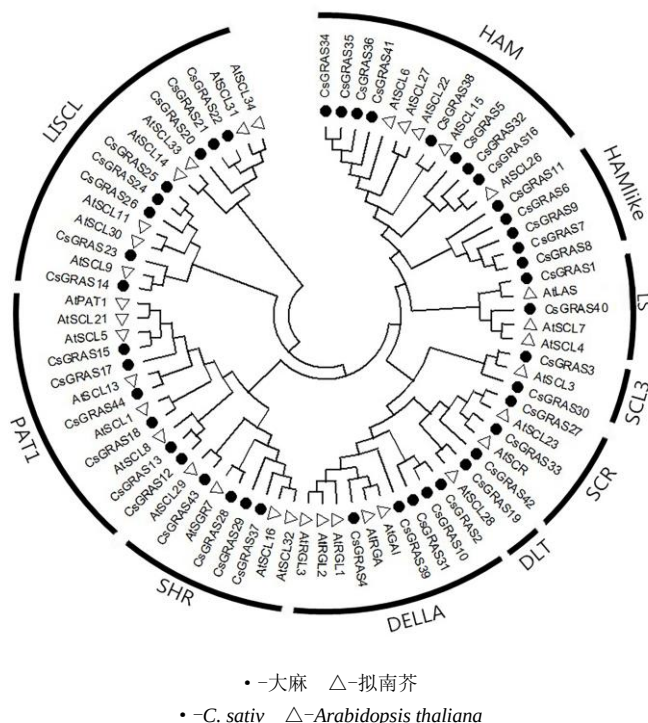


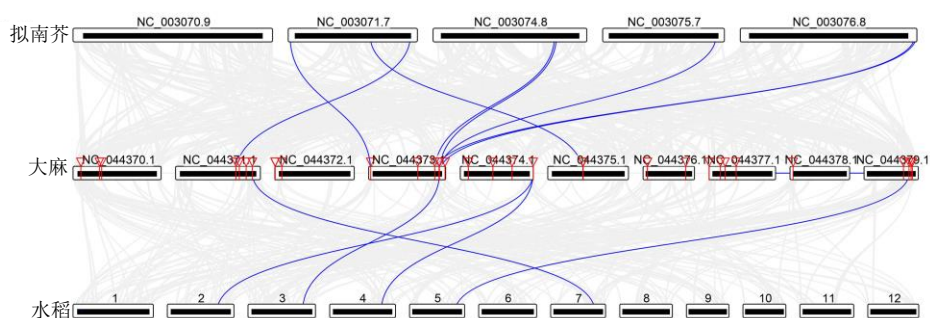
图 2 大麻和拟南芥 GRAS 转录因子系统发育树

Fig. 2 Phylogenetic trees of GRAS transcription factors in *C. sativa* and *Arabidopsis thaliana*

PAT1 亚家族包括来自工业大麻的 6 个成员 (CsGRAS12、CsGRAS13、CsGRAS15、CsGRAS17、CsGRAS18、CsGRAS44)。SHR 和 SCR 亚家族分别拥有 4 个蛋白 (CsGRAS28、CsGRAS29、CsGRAS37、CsGRAS43) 和 3 个蛋白 (CsGRAS27、CsGRAS33、CsGRAS42)，推测这些蛋白可能与植物的发育有关^[18]。SCL3 亚家族包含 2 个蛋白 (CsGRAS3、CsGRAS30)，在拟南芥中同家族 AtSCL3 蛋白通过整合多种信号来调节根细胞的伸长^[19]。CsGRAS19 是 DLT 亚家族中唯一的成员，它通过调节油菜素甾体信

号参与控制株高^[20]。LS 有 2 个成员 CsGRAS1 和 CsGRAS40。其余 8 个蛋白 (CsGRAS14、CsGRAS20、CsGRAS21、CsGRAS22、CsGRAS23、CsGRAS24、CsGRAS25、CsGRAS26) 属于 LISCL 家族。

利用水稻、拟南芥和大麻的同源关系，运用比较基因组学的方法找到大麻基因组、拟南芥基因组和水稻基因组之间的共线性关系，图 3 中一共有 12 条亮线，代表有 12 个 GRAS 同源基因对，大麻和拟南芥中，一共发现 7 对同源基因，分别为 *CsGRAS34/AtSCL23*、*CsGRAS4/AtRGA*、



蓝色线段表示大麻 GRAS 基因家族与拟南芥和水稻 GRAS 基因家族共线性关系, 红色标识表示大麻 GRAS 基因家族在大麻染色体的位置

Blue line segment indicates the collinear relationship between cannabis GRAS gene family and Arabidopsis and rice GRAS gene family, and the red mark indicates the position of cannabis GRAS gene family on cannabis chromosome

图 3 拟南芥、大麻和水稻共线性关系

Fig. 3 Collinear relationship between *A. thaliana*, *C. sativa*, and *Oryza sativa*

CsGRAS37/AtSCL22、*CsGRAS37/AtSCL27*、*CsGRAS38/AtSCL16*、*CsGRAS40/AtSCL2* 和 *CsGRAS40/AtSCL3*, 与大麻和拟南芥构建的系统发育树结果一致; 在大麻和水稻中, 一共发现了 5 对同源基因, 分别为 *CsGRAS44/Os07t0583600-01*、*CsGRAS40/Os03t0723000-02*、*CsGRAS41/Os02t0662700-01*、*CsGRAS41/Os04t0555000-01* 和 *CsGRAS29/Os05t0500600-01*, 根据大麻与拟南芥和水稻的同源基因可以推测其可能具有相似的生物学功能。

2.3 大麻 GRAS 家族基因结构和蛋白保守基序分析

为了了解大麻 GRAS 转录因子的保守性及相关性, 对大麻 GRAS 蛋白构建进化树, 并整合了 domain 和 motif 信息, 如图 4 所示。基因家族成员是由同一个基因通过复制重组不断演化而来。因此, 基因家族成员在序列上具有高度的保守性, 包含了一些高度保守性的 motif。大麻 GRAS 家族中共鉴定出 10 个保守基序, *CsGRAS8* 包含的保守基序种类最少, 为 8 个。结合图 4 分析发现大麻 GRAS 家族蛋白的 domain 和 motif 的分布相对保守。大麻 GRAS 每个亚家族所包含的基序数量也有所差别, 保守基序在一定程度上反映了大麻 GRAS 基因家族在进化过程中的结构特点, 为研究基因家族的功能提供了理论基础和研究依据。大麻 GRAS 家族基因结构相对比较简单, 内含子数量在 0~2, 其中 81.8% 基因不含有内含子, 15.9% 基因 (*CsGRAS5*、*CsGRAS6*、*CsGRAS8*、*CsGRAS19*、

CsGRAS26、*CsGRAS37*、*CsGRAS39*) 含有 1 个内含子, 只有 2.3% 基因 (*CsGRAS2*) 含有 2 个内含子。

通过 MEME 网站的分析, 大麻 GRAS 家族一共发现了 10 个 motif (图 5), 其中有 5 个最保守的 motif 分别为 VHIID、PFYRE、LRI、SAW、LRII。经过分析发现 motif 1 包含了 GRAS 家族中最保守的 VHIID 结构域, 其在靠近 N 端的部位出现 5 个保守的位点。motif 3 为 PFYRE 区; motif 4 为 LRI 区; motif 5 为 SAW 区, 在 C 端末尾处存在保守的 3 个氨基酸残基; motif 10 为 LRII 区。

使用 Jalview 软件对 44 个大麻 GRAS 蛋白保守结构域进行多重序列比对, 在图 6 中, 比对结果显示, 所有蛋白序列中 VHIID 基序中的 H 位点绝对保守; 不同序列中的 LRI 和 PFYRE 基序存在较大的差异, 但所有的 LRI 基序均呈现亮氨酸富集状态, PFYRE 基序保守性不如 VHIID 基序, 但也表现出高度的相似性; LRII 基序的保守性比 LRI 基序高; 在 SAW 基序内检测出了 S-A-W 3 个保守氨基酸残基。以上结果表明, 大麻 GRAS 蛋白在 C 端存在高度的保守性。

2.4 大麻 GRAS 基因家族顺式元件分析

对大麻 GRAS 基因家族成员顺式作用元件进行预测。从图 7 可以看出, 大麻 GRAS 家族主要包含光响应元件、厌氧诱导元件、脱落酸响应元件、赤霉素响应元件和茉莉酸甲酯响应元件等。光响应元件涵盖了所有 *CsGRAS* 家族成员, 元件数量也是

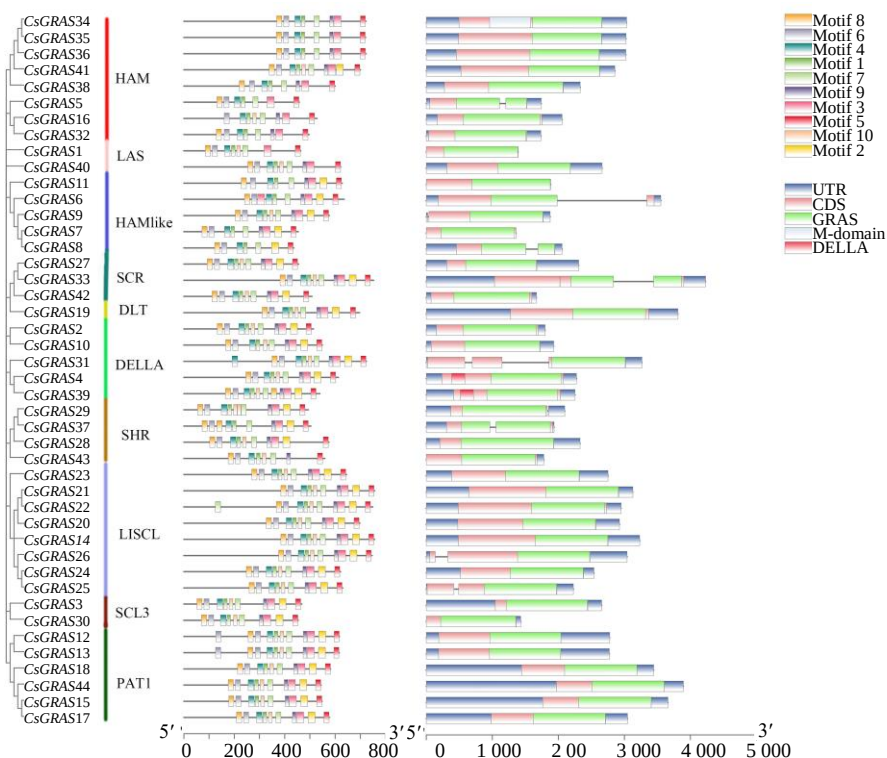


图 4 大麻 GRAS 家族进化树、保守基序及基因结构

Fig. 4 Phylogenetic, conserved motifs, and gene structure analyses of GRAS transcription factor family of *C. sativa*

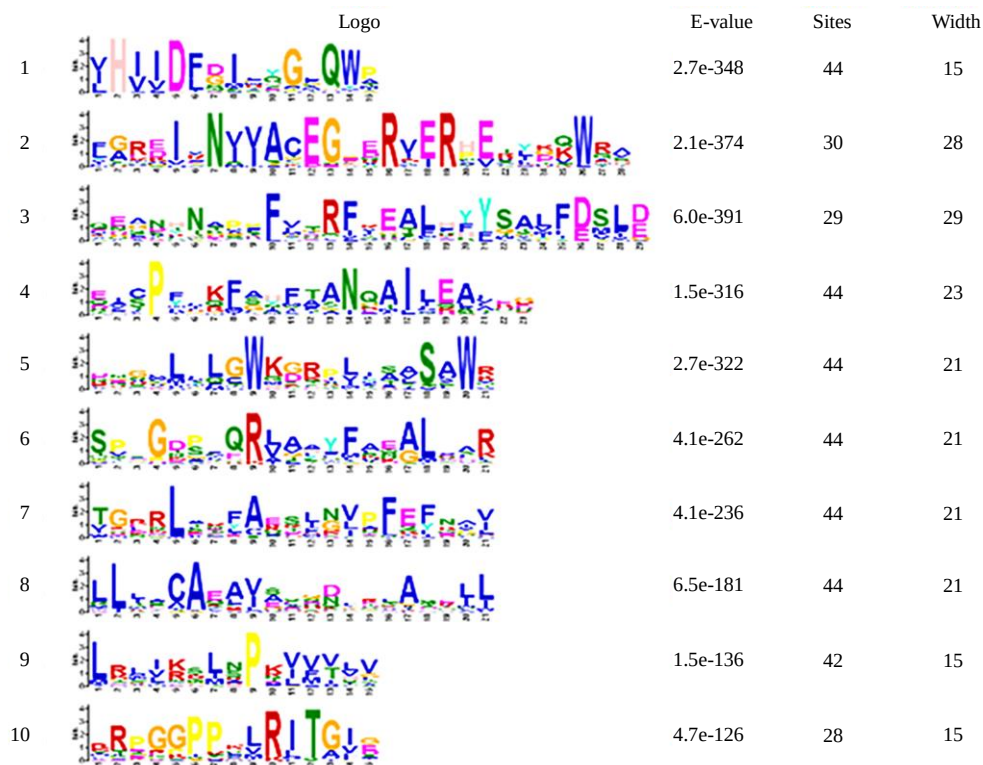


图 5 MEME 预测的 10 个 motifs LOGO

Fig. 5 Logos of ten motifs predicted by MEME online software

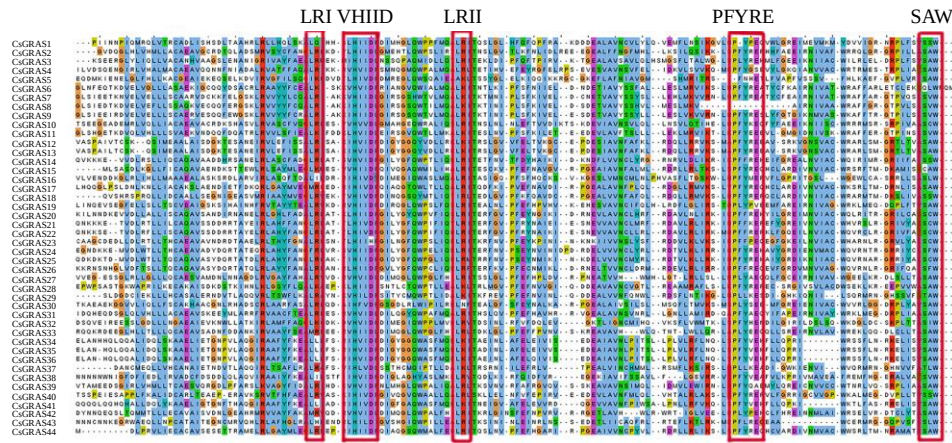
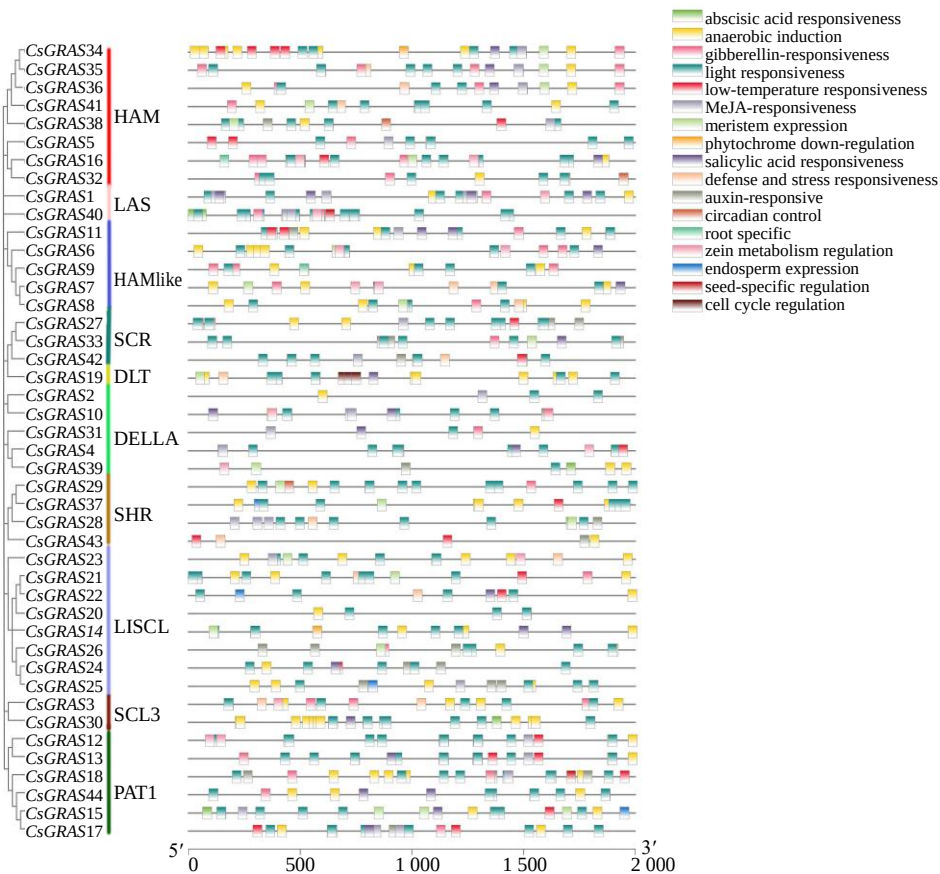


图6 大麻 GRAS 家族保守结构域序列比对

Fig. 6 Sequence alignment of conserved domains of GRAS family proteins in *C. sativa*

顺式元件分布从上到下依次为响应脱落酸的顺式作用元件、厌氧诱导的顺式作用元件、参与光响应的顺式作用元件、参与低温反应的顺式作用元件、响应茉莉酸甲酯的顺式作用元件、分生组织表达相关的顺式作用调控元件、光敏色素下调表达的顺式作用元件、参与水杨酸反应性顺式作用元件、参与防御和应激反应的顺式作用元件、参与生长素反应顺式作用元件、参与昼夜节律调控顺式作用元件、根特异性顺式作用调控元件、玉米醇溶蛋白代谢调控中的顺式作用元件、参与赤霉素反应的顺式作用元件、参与胚乳的表达顺式调控元件、种子特异性调控中的顺式作用元件、参与细胞周期调控的顺式作用元件

Distribution of CIS components is from top to bottom: abscisic acid responsiveness, anaerobic induction, light responsiveness, low-temperature responsiveness, MeJA-responsiveness, meristem expression, phytochrome down-regulation, salicylic acid responsiveness, defense and stress responsiveness, auxin-responsive, circadian control, root specific, zein metabolism regulation, gibberellin-responsiveness, endosperm expression, seed-specific regulation, cell cycle regulation

图7 大麻 GRAS 基因家族顺式作用元件预测图

Fig. 7 Prediction of cis-elements of hemp GRAS genes

最多, 共有 284 个。其次, 预测到厌氧诱导的顺式作用元件为 104 个, 响应脱落酸的顺式作用元件为 88 个, 响应茉莉酸甲酯的顺式作用元件为 72 个, 含量最少的是光敏色素下调表达的顺式作用元件和根特异性顺式作用调控元件, 都只含有 2 个, 分别存在于 *CsGRAS14*、*CsGRAS34* 和 *CsGRAS9*、*CsGRAS16* 启动子区。

2.5 大麻 GRAS 家族基因表达模式分析

为探究大麻 *GRAS* 基因的表达情况, 利用课题组的 1 个大麻品种花、茎和叶的转录组数据, 绘制大麻 *GRAS* 基因表达热图 (图 8), 并将表达模式进行聚类。 *GRAS* 基因在大麻 *ZYS* 品种花、茎和叶中的表达谱分析表明, 在花中, 除 *CsGRAS42* 不表达外, 其他基因均有不同程度表达, 其中 *CsGRAS4*、*CsGRAS13*、*CsGRAS15*、*CsGRAS17* 和 *CsGRAS44* 相对其他 38 个基因的表达量更高, 其中最高的基因是 *CsGRAS17*, *CsGRAS2*、*CsGRAS9*、*CsGRAS11*、*CsGRAS23* 和 *CsGRAS31* 基因表达量极低; 在叶中, 除 *CsGRAS8*、*CsGRAS9*、*CsGRAS18*、*CsGRAS23* 和 *CsGRAS24* 不表达, 其他基因均有不同程度表达, 其中 *CsGRAS4* 和 *CsGRAS39* 相对其他 37 个基因的表达量更高, 其中最高的基因是 *CsGRAS4*, *CsGRAS1*、*CsGRAS5*、*CsGRAS11*、*CsGRAS31*、*CsGRAS32* 和 *CsGRAS42* 基因表达量极低; 在茎中, 除 *CsGRAS9*、*CsGRAS18* 和 *CsGRAS24* 不表达, 其他基因均有不同程度表达, 其中 *CsGRAS4*、*CsGRAS39*、*CsGRAS40* 和 *CsGRAS44* 相对其他 37 个基因的表达量更高, 其中最高的基因是 *CsGRAS40*, *CsGRAS8*、*CsGRAS23*、*CsGRAS29*、*CsGRAS31* 和 *CsGRAS37* 基因表达量极低。 *GRAS* 基因在 *ZYS* 品种不同器官中的表达量存在显著差异, 其中 *CsGRAS17* 在花中显著高表达。

3 讨论

本研究从大麻基因组中筛选到 44 个 *GRAS* 基因家族成员, *GRAS* 转录因子在杨树、拟南芥、水稻、番茄、蓖麻、百脉根、大麻黄等作物中分别有 106、34、60、53、48、18、11 个家族成员^[421-24], *GRAS* 基因家族在不同物种中数量差异较大。

大麻与拟南芥 *GRAS* 家族成员进行系统进化分析, 根据结果可分成 10 个亚族, 与 Tian 等^[25-27] 的系统发育树一致, 基因结构域分析发现大麻 *GRAS* 基因家族在 C 端出现多个保守结构域。大麻 *GRAS* 基因家族成员 81.8% 都不含有内含子, 在梅花、番茄、拟南芥、水稻和杨树中无内含子

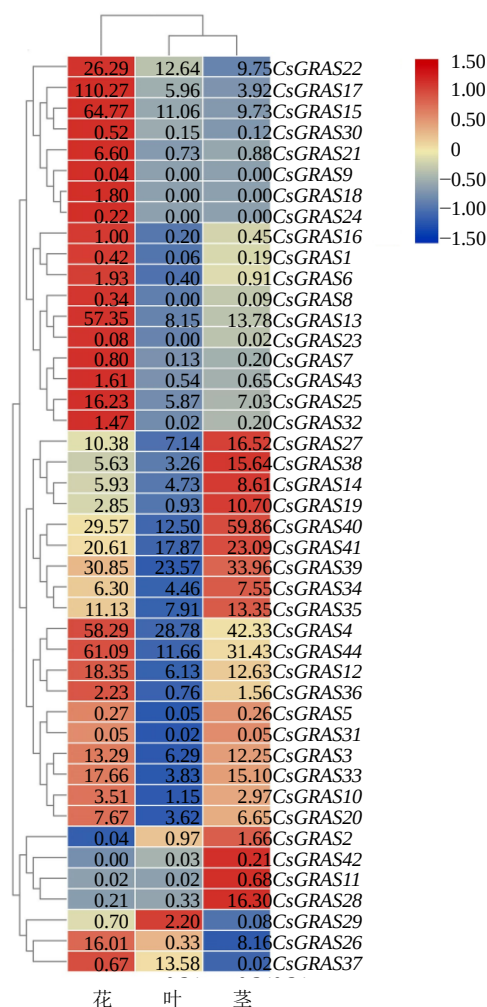


图 8 *ZYS* 品种不同部位的 *CsGRAS* 基因表达模式

Fig. 8 *CsGRAS* gene expression pattern of different organs of *ZYS*

GRAS 蛋白的比例分别占 82.2%、77.4%、67.6%、55.0%、54.7%^[4-5,28-30]。说明了大麻 *GRAS* 家族基因结构相对简单, 植物中无内含子的 *GRAS* 基因所占比例较高。

很多 *GRAS* 转录因子的功能已在模式植物中得到鉴定, 但该家族在药用植物中的功能仍有待进一步研究。在本研究中, 通过系统进化树的划分, 根据已报道基因的功能^[31], 能帮助了解和预测分在同一支中大麻 *GRAS* 基因的功能。如 *CsGRAS15*、*CsGRAS17* 与 *AtPAT1* 基因分在一支, 通过查询 *AtPAT1* 的功能, 推测 *CsGRAS15* 和 *CsGRAS17* 可能参与了光敏色素 phyA 信号转导^[31]。大麻素是一种具有精神活性的酚萜类物质。除了种子, 大麻植物其他部位都含有大麻素, 它们主要在花中合成和积累^[13], 其

生物合成涉及酚类和萜类化合物。花青素属于酚类化合物,与大麻素具有相同的上游代谢途径。在拟南芥中 RGA、RGL1 和 RGL2 这 3 种 DELLA 蛋白对拟南芥的花瓣、花蕊及花药的发育具有重要的调控作用^[32],GA 调控成花的作用部位为叶片和茎尖,而 DELLA 蛋白在长日照下可沉默叶片中的 *miR172* 基因和茎尖中的 *MADS* 基因,参与 GA 信号转导途径抑制开花^[33],PAP1 蛋白是参与花青素合成的转录因子,GAI 和 RGA 蛋白能与 PAP1 蛋白相互作用,从而促进花青素合成^[34-35]。而在缺氮条件下,DELLA 蛋白又能通过与 PAP1 蛋白相互作用,实现花青素的积累^[36],而 DELLA 蛋白又可能是酚酸和黄酮生物合成途径中的正调控因子^[10]。根据以上信息,推测大麻中的 5 个 DELLA 蛋白 CsGRAS2、CsGRAS4、CsGRAS10、CsGRAS32 和 CsGRAS40 可能调控大麻的花期,参与大麻素的合成。在系统发育树中,CsGRAS4 蛋白与 AtGAI 和 AtRGA 蛋白聚为一支,在共线性关系分析中也显示为同源基因对,CsGRAS15 和 CsGRAS17 与 AtPAT1 聚为一支,其中 CsGRAS17 在花中显著高表达,推测 CsGRAS4 蛋白可能降低大麻素生物合成相关的蛋白表达水平,抑制大麻素合成,而 CsGRAS15 和 CsGRAS17 蛋白可能与 CsGRAS4 蛋白相互作用,从而促进大麻素合成^[34]。CsGRAS3 与 AtSCL3 蛋白是同源基因,推测 CsGRAS3 蛋白可能正向调控 GA 的信号通路^[37],促进大麻素合成。

在大麻 GRAS 基因家族中发现很多参与光响应的顺式作用元件、根特异性顺式作用调控元件和光敏色素下调表达的顺式作用元件,表明该基因家族广泛地参与了大麻的生长发育过程,在种子萌发和腋芽及根的发育^[38]等方面发挥重要的调控作用;而赤霉素、脱落酸、茉莉酸甲酯等激素诱导类顺式作用元件的发现,可以推测出大麻 GRAS 基因受到多种激素信号分子的诱导^[39]。对大麻 GRAS 成员的鉴定、系统发育和表达分析将有助于全面研究 GRAS 基因家族的功能,更好地理解它们在大麻的生长发育、大麻素的生物合成、生物和非生物胁迫中的作用。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] Paz-Ares J, Ghosal D, Wienand U, et al. The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural

similarities to transcriptional activators [J]. *Embo J*, 1987, 6(12): 3553-3558.

- [2] Pysh L D, Wysocka-Diller J W, Camilleri C, et al. The GRAS gene family in Arabidopsis: Sequence characterization and basic expression analysis of the SCARECROW-LIKE genes [J]. *Plant J*, 1999, 18(1): 111-119.
- [3] Bolle C. The role of GRAS proteins in plant signal transduction and development [J]. *Planta*, 2004, 218(5): 683-692.
- [4] Liu X, Widmer A. Genome-wide comparative analysis of the GRAS gene family in populus, *Arabidopsis* and rice [J]. *Plant Mol Biol Rep*, 2014, 32(6): 1129-1145.
- [5] Song X M, Liu T K, Duan W K, et al. Genome-wide analysis of the GRAS gene family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) [J]. *Genomics*, 2014, 103(1): 135-146.
- [6] 包飞, 丁安琪, 程塘仁, 等. WRKY 基因家族成员的全基因组分析及其冷胁迫响应 [J]. *基因*, 2019, 10(11): 911.
- [7] Abarca D, Pizarro A, Hernández I, et al. The GRAS gene family in pine: Transcript expression patterns associated with the maturation-related decline of competence to form adventitious roots [J]. *BMC Plant Biol*, 2014, 14: 354.
- [8] Achard P, Gong F, Cheminant S, et al. The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins via its effect on gibberellin metabolism [J]. *Plant Cell*, 2008, 20(8): 2117-2129.
- [9] Weiss D, Ori N. Mechanisms of cross talk between gibberellin and other hormones [J]. *Plant Physiol*, 2007, 144(3): 1240-1246.
- [10] 王立儒. 丹参 GRAS 转录因子家族鉴定及 DELLA 亚家族基因功能研究 [D]. 西安: 陕西师范大学, 2018.
- [11] Kovalchuk I, Pellino M, Rigault P, et al. The genomics of *Cannabis* and its close relatives [J]. *Annu Rev Plant Biol*, 2020, 71: 713-739.
- [12] 李秋实, 孟莹, 陈士林. 药用大麻种质资源分类与研究策略 [J]. *中国中药杂志*, 2019, 44(20): 4309-4316.
- [13] 白云俊, 周新郢, 袁媛, 等. 药用大麻起源及其早期传播 [J]. *中草药*, 2019, 50(20): 5071-5079.
- [14] 张际庆, 陈士林, 尉广飞, 等. 高大麻二酚(CBD)含量药用大麻的新品种选育及生产 [J]. *中国中药杂志*, 2019, 44(21): 4772-4780.
- [15] 杨柳秀, 李超然, 高雯. 大麻化学成分及其种属差异研究进展 [J]. *中国中药杂志*, 2020, 45(15): 3556-3564.
- [16] 陈璇, 杨明, 郭鸿彦. 大麻植物中大麻素成分研究进展

- [J]. 植物学报, 2011, 46(2): 197-205.
- [17] Chen C, Chen H, Zhang Y, *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. *Mol Plant*, 2020, 13(8): 1194-1202.
- [18] Cui H, Levesque M P, Vernoux T, *et al.* An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants [J]. *Science*, 2007, 316(5823): 421-425.
- [19] Zhang Z L, Ogawa M, Fleet C M, *et al.* SCARECROW-LIKE 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in *Arabidopsis* [J]. *PNAS*, 2011, 108(5): 2160-2165.
- [20] Tong H, Jin Y, Liu W, *et al.* DWARF AND LOW-TILLERING, a new member of the GRAS family, plays positive roles in brassinosteroid signaling in rice [J]. *Plant J*, 2009, 58(5): 803-816.
- [21] 韩雯毓, 李国瑞, 风兰, 等. 蓖麻 GRAS 转录因子家族的全基因组分析及逆境胁迫响应 [J]. 植物遗传资源学报, 2020, 21(1): 252-259.
- [22] 牛义岭, 姜秀明, 许向阳. 番茄 GRAS 转录因子家族的生物信息学分析 [J]. 中国农学通报, 2016, 32(21): 106-116.
- [23] Zhang L, Li Q, Chen J F, *et al.* Computational identification and systematic classification of novel GRAS genes in *Isatis indigotica* [J]. *Chin J Nat Med*, 2016, 14(3): 161-176.
- [24] Xu W, Chen Z, Ahmed N, *et al.* Genome-wide identification, evolutionary analysis, and stress responses of the GRAS gene family in *Castor* beans [J]. *Int J Mol Sci*, 2016, 17(7): 263-271.
- [25] Tian C, Wan P, Sun S, *et al.* Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis* [J]. *Plant Mol Biol*, 2004, 54(4): 519-532.
- [26] 郭华军. 拟南芥转录因子 GRAS 家族 SCL15 基因对干旱胁迫的响应分析 [D]. 杨凌: 西北农林科技大学, 2009.
- [27] 徐龙. 柑橘 GRAS 基因家族全基因组分析 [D]. 武汉: 华中农业大学, 2016.
- [28] Huang W, Xian Z Q, Kang X, *et al.* Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato [J]. *BMC Plant Biol*, 2015, 15: 209.
- [29] Lu J, Wang T, Xu Z, *et al.* Genome-wide analysis of the GRAS gene family in *Prunus mume* [J]. *Mol Genet Genomics*, 2015, 290(1): 303-317.
- [30] Abarca D, Pizarro A, Hernández I, *et al.* The GRAS gene family in pine: Transcript expression patterns associated with the maturation-related decline of competence to form adventitious roots [J]. *BMC Plant Biol*, 2014, 14: 354.
- [31] Bolle C. The role of GRAS proteins in plant signal transduction and development [J]. *Planta*, 2004, 218(5): 683-692.
- [32] Achard P, Vriezen W H, Van Der Straeten D, *et al.* Ethylene regulates *Arabidopsis* development via the modulation of DELLA protein growth repressor function [J]. *Plant Cell*, 2003, 15(12): 2816-2825.
- [33] Cheng X F, Wang Z Y. Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in *Arabidopsis thaliana* [J]. *Plant J*, 2005, 43(5): 758-768.
- [34] Zhang Y, Liu Z, Liu J, *et al.* GA-DELLA pathway is involved in regulation of nitrogen deficiency-induced anthocyanin accumulation [J]. *Plant Cell Rep*, 2017, 36(4): 557-569.
- [35] Xie Y, Tan H, Ma Z, *et al.* DELLA proteins promote anthocyanin biosynthesis via sequestering MYBL2 and JAZ suppressors of the MYB/bHLH/WD40 complex in *Arabidopsis thaliana* [J]. *Mol Plant*, 2016, 9(5): 711-721.
- [36] Bassel G W, Mullen R T, Bewley J D. ProcerA is a putative DELLA mutant in tomato (*Solanum lycopersicum*): Effects on the seed and vegetative plant [J]. *J Exp Bot*, 2008, 59(3): 585-593.
- [37] 董丽莉. 毛竹 GRAS 转录因子家族全基因组分析及 PeSCR 和 PeSCL3 的功能研究 [D]. 北京: 中国林业科学研究院, 2015.
- [38] Zimmermann R, Sakai H, Hochholdinger F. The gibberellic acid Stimulated-Like gene family in maize and its role in lateral root development [J]. *Plant Physiol*, 2010, 152(1): 356-365.
- [39] Wang L, Wang Z, Xu Y, *et al.* OsGSR1 is involved in crosstalk between gibberellins and brassinosteroids in rice [J]. *Plant J*, 2009, 57(3): 498-510.

[责任编辑 时圣明]