

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

万志庭¹, 鲁梦¹, 吴沙沙¹, 米要磊^{2*}, 翟俊文^{1*}

(1. 福建农林大学园林学院/兰科植物保护与利用国家林业和草原局重点实验室, 福建 福州 350002;
2. 中国中医科学院中药研究所, 北京 100700)

摘要: MIKC 型 MADS-box 基因家族在植物花发育中发挥重要作用。本研究利用生物信息学方法从基因组数据中对大麻 (*Cannabis sativa* L.) MIKC 型基因进行鉴定, 并对基因序列特征、染色体定位、基因结构、系统发育树和组织表达模式等进行分析。结果表明, 共鉴定出 39 个 *CsMADS* 基因, 不均匀分布在 9 条染色体上, 编码的氨基酸数目介于 146~503, 等电点为 5.19~10.12, 分子质量为 16 739.35~57 070.56 Da; 亚细胞定位预测显示 *CsMADS* 基因均定位在细胞核里; 保守结构域表明 *CsMADS* 基因均含有 MADS 保守结构域; GO 功能注释分类共注释到 331 个 GO terms, 主要集中在分子功能中, 共聚类 178 个; 系统进化树显示 *CsMADS* 基因分为 14 个亚类, 按照 ABCDE 同源异型基因模型分类, 能够分别与拟南芥 (*Arabidopsis thaliana*) 中 SQUA/API、AP3、PI、GGM13、AG、SEP/AGL2 等类亚家族聚成一类; *CsMADS* 基因上游启动子区具有丰富的顺式作用元件, 其中主要为光调控元件; 实时荧光定量 PCR 验证了部分 *CsMADS* 在花和苞片中相对表达量较高, 具有组织表达特异性。本研究对大麻基因组中的 MIKC 型 MADS-box 家族进行鉴定和分析, 为进一步探究 MIKC 型基因在调控药用植物大麻生长发育以及培育优良品种等方面奠定理论基础。

关键词: 大麻; MIKC 型 MADS-box; 全基因组鉴定; 功能分析; 生物信息学

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2021)11-3173-11

Identification and expression analysis of the MIKC-type MADS-box gene family in *Cannabis sativa* L.

WAN Zhi-ting¹, LU Meng¹, WU Sha-sha¹, MI Yao-lei^{2*}, ZHAI Jun-wen^{1*}

(1. Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization / College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2. Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China)

Abstract: The MIKC-type MADS-box gene family plays an important role in flower development in plants. This study identifies members of the MIKC-type gene family in *Cannabis sativa* L. at the genome level, with the chromosomal location and linkage, evolutionary relationships, and identification of conserved motifs determined using bioinformatics tools. The results show that *C. sativa* contains 39 members of the MIKC-type MADS-box gene family (named *CsMADS1-CsMADS39*) unevenly distributed on nine chromosomes. The encoded proteins range in length from 146 to 503 amino acids, and the theoretical isoelectric points range from 5.19 to 10.12. Molecular weights range from 16 739.35 to 57 070.56 Da. The subcellular location of *CsMADS* genes is mainly in the nucleus.

收稿日期: 2021-06-18; 修回日期: 2021-07-07.

基金项目: 中央级公益性科研院所基本科研业务费专项 (ZXKT20020).

*通讯作者 Tel: 18650750580, E-mail: zhai-jw@163.com;

Tel: 86-10-64032658, E-mail: xiaomi20063@sina.com

DOI: 10.16438/j.0513-4870.2021-0892

The result of conserved domain analysis showed that all genes contain the MADS conserved domain. The analysis of GO showed that all genes were annotated to 331 GO terms, which were clustered by molecular function. A phylogenetic tree showed that *CsMADS* genes could be divided into 14 subclasses, according to ABCDE homologous gene model; *CsMADS* genes are clustered with SQUA/AP1, AP3, PI, GGM13, AG, SEP/AGL2 subfamilies in Arabidopsis. There are abundant *cis*-acting elements in the upstream promoter region of *CsMADS* genes, mainly light regulatory elements. The results of real-time fluorescent quantitative PCR showed that some *CsMADS* genes are highly expressed in flowers and bracts, with tissue-specific expression. This study identified and analyzed the MIKC-type MADS-box gene family in *C. sativa* at the genome level, and provides a theoretical basis for further exploration of the function of MIKC-type genes and their role in regulating the growth and development of medicinally important hemp.

Key words: *Cannabis sativa*; MIKC-type MADS-box; genome-wide identification; functional analysis; bioinformatics

MADS-box 基因普遍存在于动物、植物和真菌中, 是一类在植物成花诱导及花器官发育等方面具有重要调控作用的转录因子, 能够调控生长发育和信号转导^[1,2]。MADS-box 转录因子家族主要分为 I 和 II 两类, I 类具有 MADS (M)、Intervening (I)、Keratin-like (K) 和 Carboxyl-terminal (C) 4 个结构域, 故亦称 MIKC 基因^[3]。该基因家族具有高度保守的结构域, 其中 MADS 结构域对其功能至关重要, K 域是蛋白二聚化的关键结构域, 它们分别是 DNA 结合和蛋白复合物形成所不可缺少的^[4,5]。MIKC 型 MADS-box 基因可根据 I 区序列的不同分为 MIKC^c 和 MIKC* 两个亚家族, MIKC* 通常在雄配子发育中发挥重要作用, MIKC^c 基因则在花器官生长发育等方面相对于研究更多具有重要意义^[6,7]。在目前与植物花发育分子机制相关的 ABCDE 模型中, MADS-box 基因所涉及的同源二聚体和异源二聚体(花的四聚体), 它们能调节不同花序中的特定表达过程^[8]。MIKC 基因在花发育的各个阶段都发挥作用, 大多数 MIKC 型 MADS-box 成员均是花器官决定基因, 在花形态建成、调节开花时间等方面具有重要作用^[9]。

大麻 (*Cannabis sativa* L.) 是大麻科 (Cannabaceae) 大麻属 (*Cannabis*) 一年生草本植物, 栽培历史悠久, 在世界各地均有种植^[10]。大麻的用途十分广泛, 不但在各种工业产品、生物建筑、隔热材料等方面有应用^[11,12], 在药理特性方面如在治疗癫痫、破伤风、风湿病和失眠等方面同样具有良好的功效, 常常被称为天然的治疗草药^[13-15]。

在药用植物大麻中大麻素是一种具有精神活性的酚萜类物质, 它们主要在腺毛中合成和积累^[16]。随着大麻全基因组序列的公布, 考虑到 MADS-box 基因家族在植物开花调控中的重要作用, 以及该基因家族在药用开花植物中的基础信息缺乏, 对大麻 MIKC 型 MADS-box 基因家族全序列进行较为全面的研究, 有

助于更好地挖掘相关基因的生物信息学功能, 为进一步探究大麻 MIKC 型基因在花发育调控、形态建成中的作用提供理论基础。

材料与方法

材料 从 NCBI (<https://www.ncbi.nlm.nih.gov>) 网站上获取大麻 (GCA_900626175.1) 和拟南芥 (GCA_000005425.2) 的全基因组及注释文件, 品种为 CRBRx 的雌株。转录组数据分为两部分, 9 个不同品种腺毛的 RNA-seq 数据 (PRJNA498707) 获取于 NCBI 数据库^[17], 品种名称分别为 Black Berry Kush (BB)、Black Lime (BL)、Canna Tsu (CT)、Mama Thai (MT)、Valley Fire (VF)、Cherry Chem (CC)、Terple (T)、Sour Diesel (SD) 和 White Cookies (WC); 同一品种 Diku 植株的茎、叶、苞片和雌花的 RNA-seq 数据是由课题组通过测序获得, Diku 品种的全称为 Dinamed Kush CBD Autoflowering, 是 Purple Kush 和 Dinamed Autoflowering CBD 的杂交品种。

大麻 MIKC 型 MADS-box 家族成员的鉴定、蛋白特征分析及染色体定位 大麻 *CsMADS* 基因家族鉴定使用 TBtools^[18] 软件提取大麻蛋白序列, 利用 PlantTFDB 网站进行转录因子预测, 将预测得到的目的家族基因成员进行筛选, 去除可变剪切导致的重复转录本, 并利用 TBtools 软件获取大麻 *CsMADS* 的蛋白序列并解析大麻的注释文件, 同时使用在线生物信息学工具 ExPASy-ProSite (<http://web.expasy.org/protparam/>) 对 39 个 *CsMADS* 蛋白的理化性质进行预测。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。同时从基因组注释文件中筛选出基因家族成员 ID, 并利用 TBtools 软件确定每个 *CsMADS* 基因的位置信息并绘制其所对应的染色体物理位置图。

大麻 MIKC 型 MADS-box 基因保守结构域、基因结构分析与 GO 功能分析 利用 MEME (<http://meme-suite.org/tools/meme>) 对 39 个 *CsMADS* 蛋白质保守基序进行挖掘与分析, 目标保守基序数量设置为 10。利用基因组注释文件对大麻 *CsMADS* 基因进行解析, 最终通过 TBtools 软件对大麻 MIKC 家族的基因进行基因结构、保守结构域和保守基序三者可视化分析。利用 OmicsBox 分析工具对所有鉴定出的 *CsMADS* 转录因子序列进行氨基酸相似性比对、鉴定和筛选, 并对筛选出的转录因子进行 GO 功能分类。

大麻 MIKC 型 MADS-box 系统发育树构建及共线性分析 利用 39 个 *CsMADS* 的蛋白序列, 与 TAIR 数据库 (<https://www.arabidopsis.org/>) 搜索到的拟南芥 (*Arabidopsis thaliana*) 45 个 MIKC 家族基因蛋白序列, 应用 MEGA-X 比对工具采取 NJ (Neighbour-join) 方法构建分子进化树, 并进行了重复 1 000 次的 Bootstrap 检验, 所有的设定均采用默认值。使用 TBtools 软件对大麻和拟南芥共线性关系进行可视化, 获得两者的同源基因对。

大麻 MIKC 型 MADS-box 基因顺式作用元件预测及表达模式分析 使用 TBtools 软件提取 39 个 *CsMADS* 基因上游 2 000 bp 启动子序列, 利用 PlantCare (<http://www.plantcare.co.uk/>) 在线网站对其进行顺式作用元件预测, 通过 TBtools 软件对其进行可视化构图。基于实验室测序获得的一个杂交品种 Diku 的雌花、苞片、茎和叶的转录组数据以及 9 个不同栽培品种雌性花序的转录组数据分析基因的差异表达情况, 利用 TBtools 软件进行表达模式热图的绘制, 对 *CsMADS* 成员基因表达量 (FPKM) 使用 $\log_2(\text{FPKM}+1)$ 对数据进行标准化处理。

大麻 MIKC 型 MADS-box 基因实时荧光定量 PCR 分析 利用 Liu 等^[19]筛选的 *EFla* 作为内参基因, 对转录组数据中存在表达量差异的 *CsMADS4*、*CsMADS7*、*CsMADS12*、*CsMADS29* 进行实时荧光定量 PCR (quantitative real-time PCR, qRT-PCR) 验证, 利用 NCBI 设计引物 (表 1)。实验样品为大麻 (Diku) 的雌花、苞片、茎和叶, 每组试验 3 个重复, 采用 WARYONG 试剂盒提取样品 RNA, 用 TransGen 试剂盒进行 cDNA 合成, 以及 TransStart 试剂盒进行实时荧光定量 PCR 分

析。反应程序为: 95 °C 酶激活 30 s, 95 °C 变性 5 s, 56 °C 退火 15 s, 72 °C 延伸 10 s, 循环次数 40。 *CsMADS* 的相对表达量采用 $2^{-\Delta\Delta Ct}$ 法计算。

结果与分析

1 大麻 MIKC 型 MADS-box 基因的鉴定与染色体分布

从大麻基因组数据库筛选并鉴定出 63 个 MIKC 型 MADS-box 基因, 去除可变剪切导致的重复转录本和结构域确认后, 最终获得 39 个 MIKC 型 MADS-box 基因家族成员, 依次命名为 *CsMADS1* ~ *CsMADS39*。利用大麻 MIKC 家族的基因 ID, 使用 TBtools 软件获得 *CsMADS* 的蛋白序列。结果如表 2 所示, 其中氨基酸数量最少的是 *CsMADS34*, 包含 146 个氨基酸残基, 数量最多的为 *CsMADS38*, 由 503 个氨基酸组成。蛋白质等电点最小的为 *CsMADS26*, 其 PI 为 5.09, 最大的是 *CsMADS20*, 其 PI 为 10.12。其中所有蛋白的分子量在 16 739.35 ~ 57 070.56 Da 之间。亚细胞定位预测结果显示 39 个 *CsMADS* 蛋白均集中在细胞核定位。

CsMADS 基因家族 39 个基因不均匀地分布在 9 条染色体上 (NC_044370.1、NC_044371.1、NC_044372.1、NC_044373.1、NC_044374.1、NC_044375.1、NC_044377.1、NC_044378.1、NC_044379.1 分别分布了 6、3、5、5、3、4、1、1、11 个 *CsMADS* 基因) (图 1)。每条染色体排列的基因数目差别很大, 其中 NC_044379.1 染色体上 *CsMADS* 基因家族数目最多, 含有 11 个基因, 最少的是 NC_044377.1 和 NC_044378.1 均只分布 1 个基因。并发现 NC_044370.1 染色体上的 *CsMADS15* 和 *CsMADS16*, NC_044375.1 染色体上的 *CsMADS33* 和 *CsMADS35*、*CsMADS32* 和 *CsMADS34*; NC_044379.1 染色体上的 *CsMADS12*、*CsMADS7* 和 *CsMADS3* 出现了基因成簇排列现象。另外在同一条染色体中出现的多基因存在串联重复特征, 如 NC_044372.1 染色体上的 *CsMADS20* 和 *CsMADS22*、NC_044379.1 染色体上的 *CsMADS5* 和 *CsMADS8*。

2 大麻 MIKC 型 MADS-box 基因保守结构域与基因结构分析

基因家族成员是通过不断拷贝进化而来, 在序列上具有保守性和相关性^[20]。通过 MEME 在线网站分析

Table 1 Primer sequences of *CsMADS*

Gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
<i>CsMADS4</i>	TCCAGAACTCAAACAGGCAT	GTTTTCGAACCTCCATCTCCC
<i>CsMADS7</i>	TATGTGATGCTGAAGTCGCT	TTTTCCTTCTGTGCTCCAT
<i>CsMADS12</i>	CTTCGACGTACACAGAGGAA	TCAACATTCTTCTCCGCCCT
<i>CsMADS29</i>	GCTGAAGTCCAAGCTTGAAG	GGTTGTATGGAGTCCGGATT
<i>EFla</i>	ACCAAGATTGACAGGCGTTC	CCTTCTTCTCCACAGCCTTG

Table 2 The information of 39 *CsMADS* genes in *Cannabis sativa*

Locus name	Accession No.	Gene name	AA	PI	MW/Da	Subcellular location	MIKC-type Subclade
LOC115697576	rna-XM_030624644.1	<i>CsMADS1</i>	298	9.23	33 797.65	Nucleus	AG
LOC115697575	rna-XM_030624658.1	<i>CsMADS2</i>	250	9.02	28 894.13	Nucleus	AGL17
LOC115698549	rna-XM_030625630.1	<i>CsMADS3</i>	255	7.63	29 534.55	Nucleus	API
LOC115698701	rna-XM_030625845.1	<i>CsMADS4</i>	280	7.77	32 324.74	Nucleus	AG
LOC115699531	rna-XM_030626996.1	<i>CsMADS5</i>	204	7.8	23 610.12	Nucleus	PI
LOC115699728	rna-XM_030627267.1	<i>CsMADS6</i>	214	8.94	24 164.17	Nucleus	AGL12
LOC115700576	rna-XM_030628159.1	<i>CsMADS7</i>	259	8.1	29 085.9	Nucleus	AGL6
LOC115700653	rna-XM_030628272.1	<i>CsMADS8</i>	214	7.8	24 777.45	Nucleus	PI
LOC115700827	rna-XM_030628489.1	<i>CsMADS9</i>	257	8.73	29 622.47	Nucleus	SEP/AGL2
LOC115700828	rna-XM_030628492.1	<i>CsMADS10</i>	254	8.95	29 273.06	Nucleus	SEP/AGL2
LOC115700907	rna-XM_030628576.1	<i>CsMADS11</i>	251	5.75	28 913.85	Nucleus	GGM13
LOC115701144	rna-XM_030628850.1	<i>CsMADS12</i>	268	6.52	30 700.9	Nucleus	AGL6
LOC115701208	rna-XM_030628937.1	<i>CsMADS13</i>	317	6.76	35 912.35	Nucleus	MIKC*
LOC115701438	rna-XM_030629241.1	<i>CsMADS14</i>	263	9.44	30 090.31	Nucleus	API
LOC115702704	rna-XM_030630128.1	<i>CsMADS15</i>	318	6.49	36 768.27	Nucleus	MIKC*
LOC115702756	rna-XM_030630191.1	<i>CsMADS16</i>	318	6.49	36 709.25	Nucleus	MIKC*
LOC115704295	rna-XM_030631508.1	<i>CsMADS17</i>	184	9.92	21 400.84	Nucleus	SVP
LOC115706939	rna-XM_030634719.1	<i>CsMADS18</i>	236	9.69	27 041.82	Nucleus	SOC1
LOC115707234	rna-XM_030635120.1	<i>CsMADS19</i>	243	7.14	27 867.62	Nucleus	GGM13
LOC115708930	rna-XM_030636958.1	<i>CsMADS20</i>	181	10.12	21 092.38	Nucleus	SVP
LOC115709018	rna-XM_030637041.1	<i>CsMADS21</i>	251	8.56	28 557.3	Nucleus	SEP/AGL2
LOC115709034	rna-XM_030637066.1	<i>CsMADS22</i>	181	10.12	21 092.38	Nucleus	SVP
LOC115710309	rna-XM_030638677.1	<i>CsMADS23</i>	255	6.98	28 863.59	Nucleus	AGL15
LOC115712087	rna-XM_030640317.1	<i>CsMADS24</i>	244	8.78	27 934.64	Nucleus	SEP/AGL2
LOC115712534	rna-XM_030640823.1	<i>CsMADS25</i>	234	9.26	26 902.74	Nucleus	AGL17
LOC115713471	rna-XM_030641955.1	<i>CsMADS26</i>	345	5.19	38 374.17	Nucleus	MIKC*
LOC115714329	rna-XM_030642992.1	<i>CsMADS27</i>	235	6.35	26 509.02	Nucleus	SVP
LOC115714657	rna-XM_030643408.1	<i>CsMADS28</i>	305	6.77	34 599.12	Nucleus	AP3
LOC115716337	rna-XM_030645105.1	<i>CsMADS29</i>	226	7.84	26 372.5	Nucleus	AGL6
LOC115716986	rna-XM_030645911.1	<i>CsMADS30</i>	237	9.13	27 211.99	Nucleus	SOC1
LOC115717309	rna-XM_030646279.1	<i>CsMADS31</i>	218	9.49	25 581.53	Nucleus	SOC1
LOC115718840	rna-XM_030647651.1	<i>CsMADS32</i>	265	9.27	30 442.75	Nucleus	AGL17
LOC115720605	rna-XM_030649758.1	<i>CsMADS33</i>	228	9.01	26 267.98	Nucleus	SVP
LOC115720817	rna-XM_030649993.1	<i>CsMADS34</i>	146	9.59	16 739.35	Nucleus	AGL17
LOC115721274	rna-XM_030650540.1	<i>CsMADS35</i>	220	6.98	25 044.69	Nucleus	SVP
LOC115709360	rna-XM_030637447.1	<i>CsMADS36</i>	341	5.91	38 297.35	Nucleus	MIKC*
LOC115711048	rna-XM_030639386.1	<i>CsMADS37</i>	247	7.7	28 908.82	Nucleus	MIKC*
LOC115715550	rna-XM_030644207.1	<i>CsMADS38</i>	503	5.91	57 070.56	Nucleus	MIKC*
LOC115725163	rna-XM_030654597.1	<i>CsMADS39</i>	193	6.93	22 068.24	Nucleus	FLC

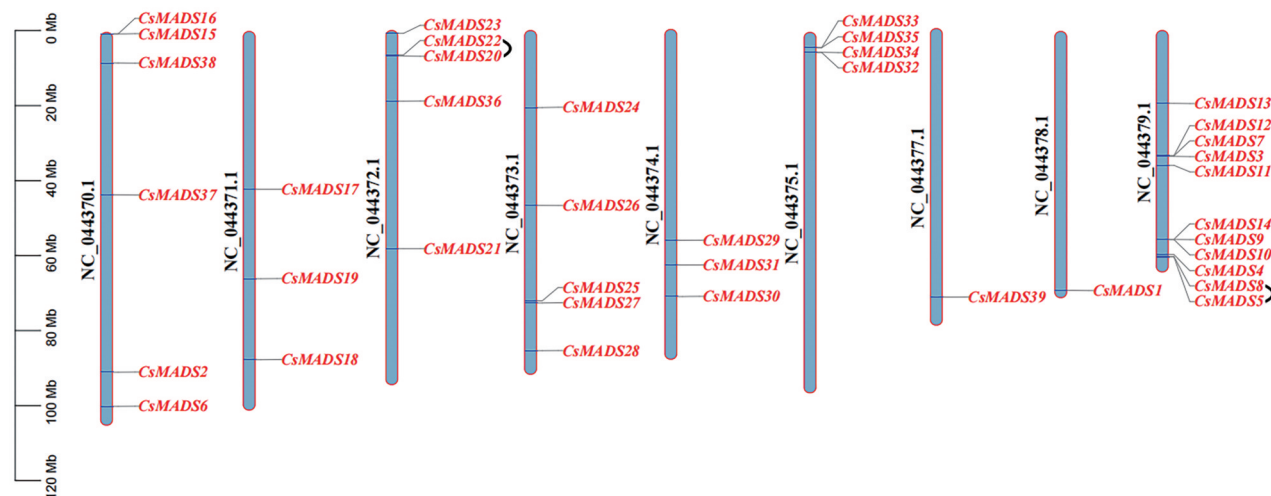


Figure 1 Distribution of 39 *CsMADS* genes on nine chromosomes in *Cannabis sativa*

基因序列, 获得 10 个相对保守性较高的基序 (图 2)。其中 motif 之间的位置具有一定规律, 在所鉴定的 39 个基因中均存在高度保守的 MADS 区 (motif 1), 而且除了 MIKC* 亚类、*CsMADS15* 和 *CsMADS16* 之外, 其余基因均存在较为保守的 K-box 区, 在 AGL15、AGL17、API 和 AGL6 亚类中具有 motif1-motif3-motif5-motif2-motif4 的相对稳定结构, 并未出现保守基序增加、替换或缺失现象, 但在不同亚类中保守基序数量存在不同, 说明在基因家族进化过程中结构特征存在差异。

通过分析 39 个 *CsMADS* 基因的非翻译区 (UTR, untranslated region)、内含子和外显子结构, 发现大多数基因 UTR 数量位于 0~2 之间, 其中 87.2% 的基因含有 2 个, 约 7.7% 的基因只含有 1 个; 外显子数量较为保守, 除了 *CsMADS26* 由 11 个外显子组成外, 其他大多数基因的外显子数量处于 6~8 之间; 大部分基因包含 3~9 个不等的内含子, 其中 *CsMADS2*、*CsMADS12* 以及 *CsMADS30* 的单个内含子长度较大, 但有 3 个 *CsMADS* 基因 (*CsMADS15*、*CsMADS16* 和 *CsMADS38*) 没有内含子, 只有一个外显子; 同一进化支上基因的结构较为保守, 例如 SEP/AGL2 (*CsMADS9*、*CsMADS10*、*CsMADS21* 和 *CsMADS24*) 和 AGL6 (*CsMADS7*、*CsMADS12*、*CsMADS29*) 进化支上的基因均由 8 个外显子和 7 个内含子组成 (图 3)。

3 大麻 MIKC 型 MADS-box 基因 GO 富集分析

利用 OmicsBox 软件进行 GO 功能注释分析, 结果

显示 39 个 *CsMADS* 基因共注释到 331 个 GO terms (图 4), 生物过程 (biological process, BP)、分子功能 (molecular function, MF) 和细胞组分 (cellular component, CC) 涉及的 GO terms 分别为: 178、77 和 76 个。其中 DNA binding 富集个数最多, 在所有基因中均有注释; 细胞组分中细胞结构组分和胞内组分均富集到 38 个转录因子, 生物过程中细胞过程也相对较多, 富集数目为 36 个。

4 大麻 MIKC 型 MADS-box 蛋白系统发育树的构建及家族进化特性

为探究大麻 MIKC 型 MADS-box 基因家族的进化关系, 使用 MEGA-X 软件对大麻 39 个和拟南芥 45 个蛋白序列构建 Neighbor-join (NJ) 树。大麻 39 个 *CsMADS* 基因明显分为 14 个进化支, 即 AG、AGL12、MIKC*、AGL15、AGL17、API、AGL6、SEP/AGL2、PI、SVP、AP3、FLC、SOC1 和 GGM13, FLC 进化支在拟南芥中共有 6 个基因, 是春化开花调控途径的关键基因, 但是在大麻中该进化支上只有 1 个同源基因, 即 *CsMADS39* (图 5)。不同进化支上大麻分布的基因数量略有差异, 如在调控开花时间的 SVP 进化支中, 大麻的基因数目明显多于拟南芥, 在拟南芥中有 2 个基因, 在大麻中却有 6 个同类基因; AGL17 和 SEP/AGL2 进化支中, 拟南芥和大麻基因数量相同, 均有 4 个基因; 而在 PI 进化支中, 大麻的基因数量较少, 只有 2 个, 即 *CsMADS5* 和 *CsMADS8*, 这类基因均为控制花形态

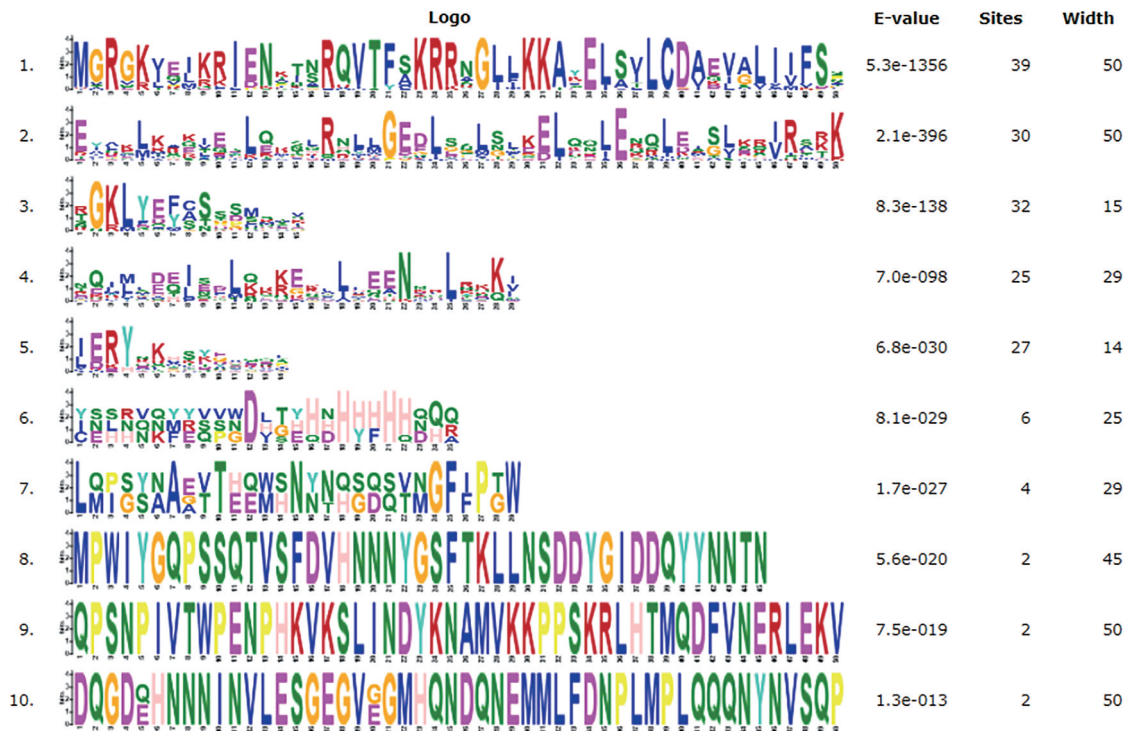


Figure 2 The conservation domain of 39 *CsMADS* genes in *Cannabis sativa*

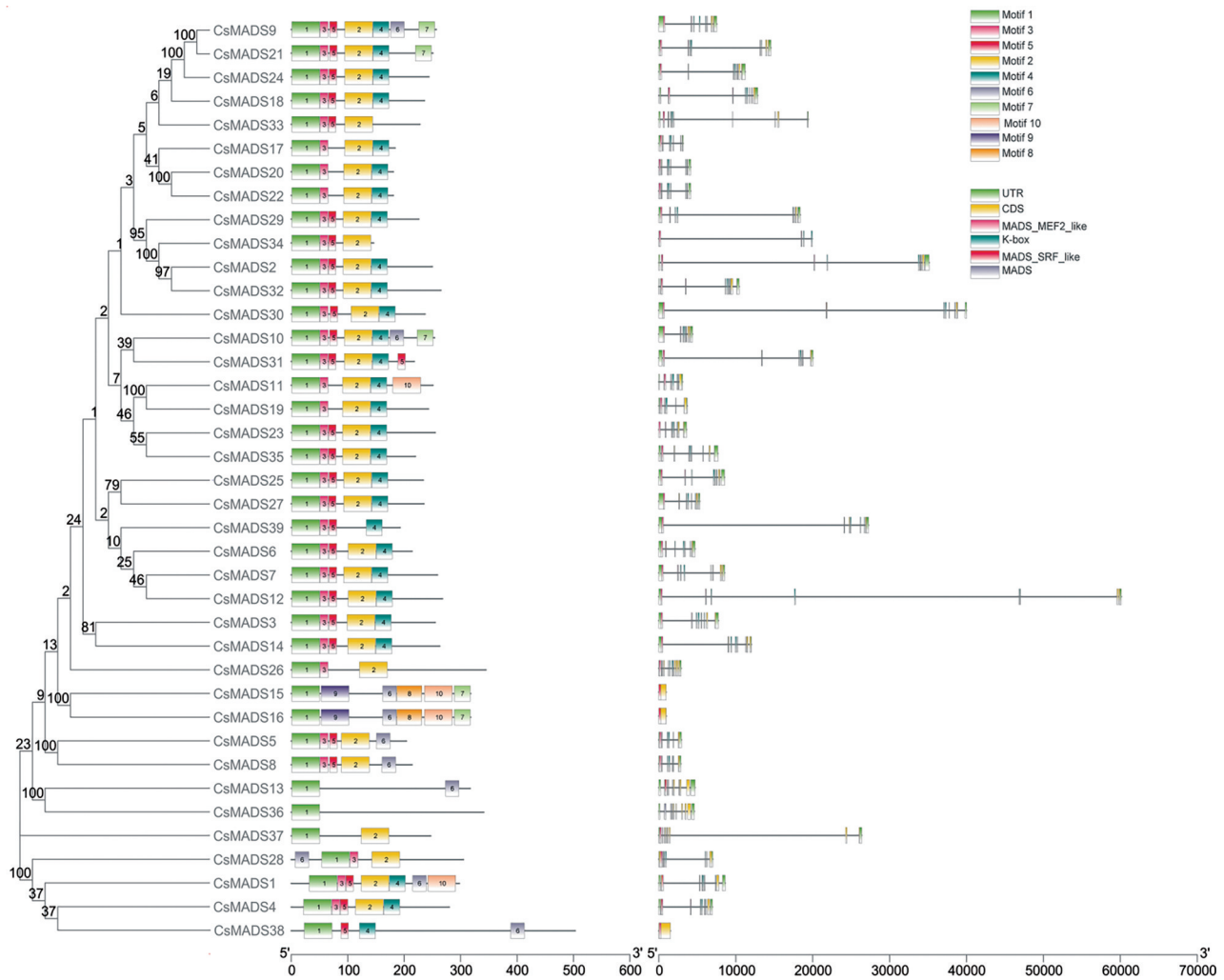


Figure 3 Conserved motifs and gene structure analyses of 39 *CsMADS* genes of *Cannabis sativa*

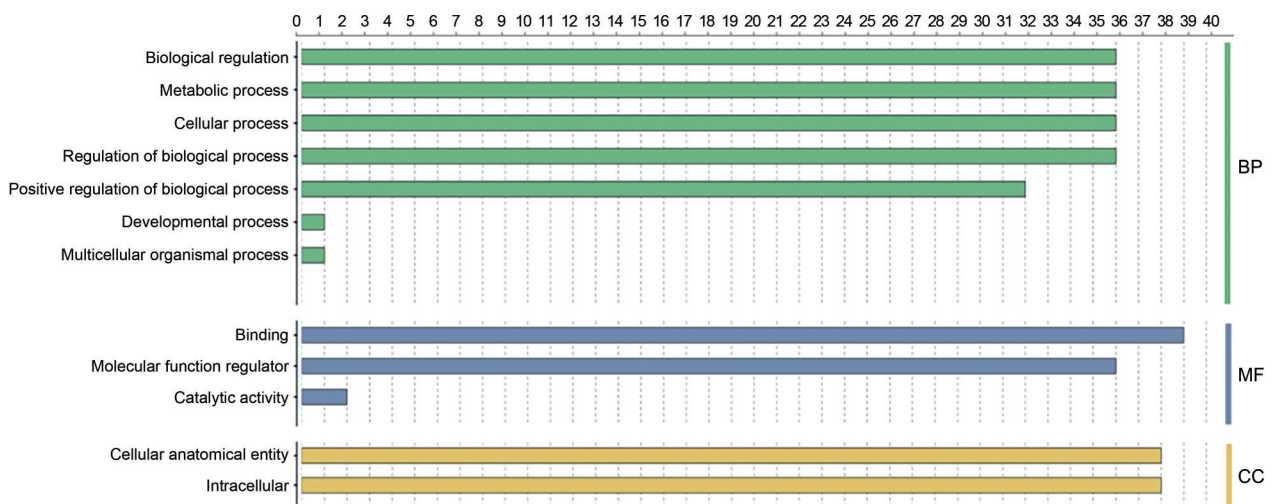


Figure 4 GO categorization and annotation of 39 *CsMADS* genes

建成的ABCDE模型中的同源异型基因,推测大麻在器官形态建成中可能存在更加复杂和精细的调控机制。

5 大麻MIKC型MADS-box基因家族共线性分析

利用共线性原理,探究大麻*CsMADS*基因和拟南芥基因家族成员的共线性关系并可视化,结果显示,大

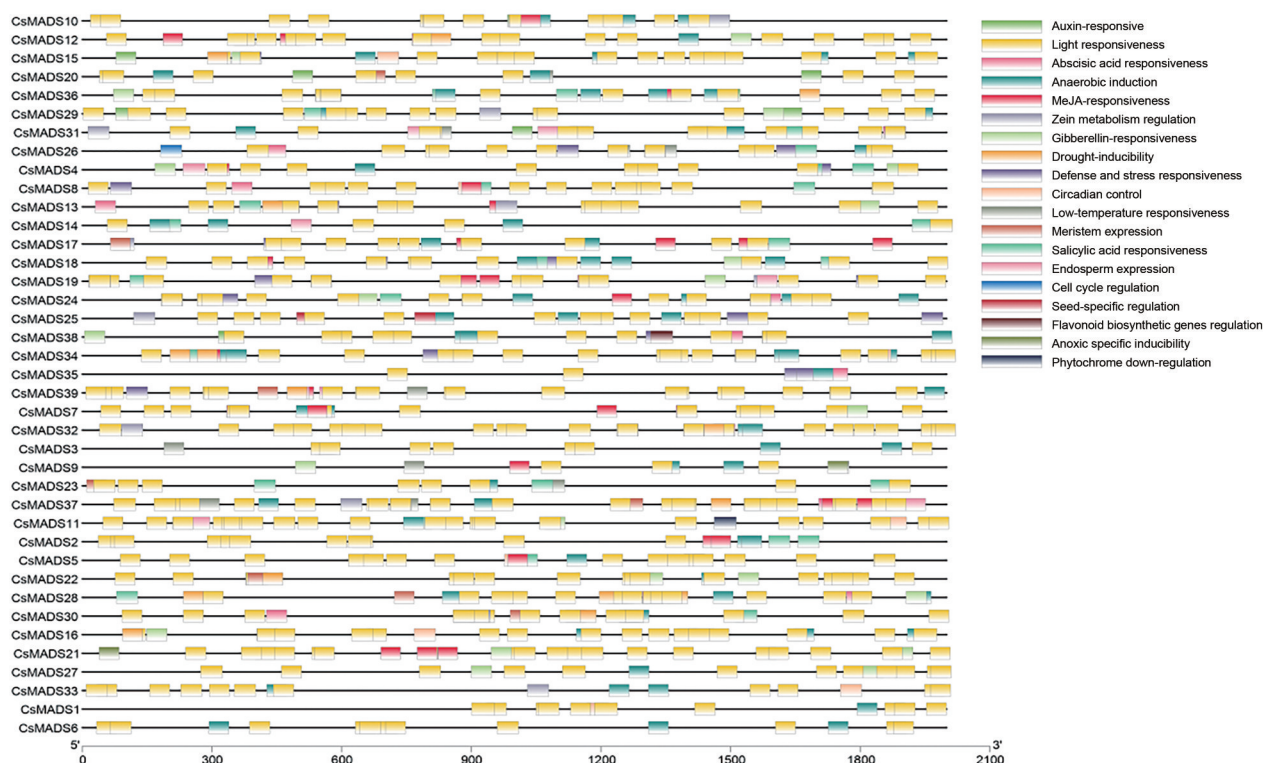


Figure 7 cis-Acting elements within the promoter of 39 *CsMADS* genes in *Cannabis sativa*

7 大麻 MIKC 型 MADS-box 家族基因表达模式分析

为揭示 *CsMADS* 基因的表达情况, 将 9 个大麻品种的腺毛以及大麻 (Diku) 的花、苞片、茎和叶的转录组数据, 绘制基因表达热图, 并将表达模式进行聚类。在不同品种中, *CsMADS29* 和 *CsMADS14* 在 SD、CT、BL、VF、CC、WC、BB、MT 和 T 品种中表达最高, 而 *CsMADS17* 和 *CsMADS20* 则在 Diku 品种中表达最高

(图 8A)。在同一品种的不同部位中, *CsMADS* 表达模式存在差异, *CsMADS29* 和 *CsMADS7* 在花和苞片中表达上调 (图 8B), *CsMADS20* 和 *CsMADS22* 在茎中表达较高, 而 *CsMADS34*、*CsMADS38*、*CsMADS11*、*CsMADS26*、*CsMADS28*、*CsMADS33* 在花、苞片、茎和叶中均没有表达。表达模式分析结果表明大麻 *CsMADS* 基因成员在不同组织器官和品种中表达均存在差异

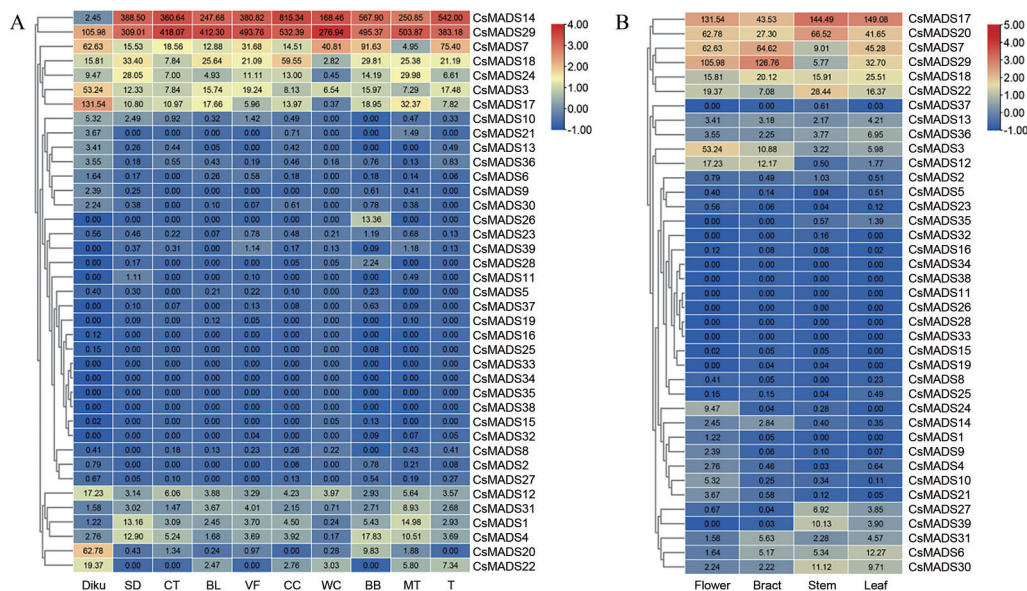


Figure 8 Expression pattern of MIKC-type MADS-box genes in *Cannabis sativa*. A: Indicates the expression pattern of *CsMADS* in female flowers of ten breeds; B: Indicates the expression pattern of *CsMADS* in different organs of same breed

性, 暗示其功能存在分化。

8 *CsMADS* 基因家族成员的实时荧光定量 PCR 分析

大麻腺毛位于雌花苞片的表面, 是富含大麻素和萜类物质的树脂生物合成和储存的主要场所, 而 *CsMADS4*、*CsMADS7*、*CsMADS12*、*CsMADS29* 正是在雌花或苞片中明显上调的基因, 利用 qRT-PCR 检测这 4 个基因在大麻 (Diku) 不同组织部位的表达情况, 结果显示, *CsMADS7*、*CsMADS12*、*CsMADS29* 在大麻 Diku 品种各组织部位的表达情况与转录组趋势略有不同, 但整体来说这 3 个基因在雌花和苞片中的表达量高于在茎和叶中的表达, 且 *CsMADS12* 和 *CsMADS29* 在花和苞片中表达量极高; *CsMADS4* 具有明显的组织差异性, 只在花中极显著表达, 在其他部位几乎无表达 (图 9)。

讨论

在植物中, MIKC 型 MADS-box 转录因子包含 MADS-域和 K-box 域, 在植物发育和生长等多种生物过程中发挥着关键作用^[21,22]。近年来, MIKC 型 MADS-box 基因家族的鉴定和进化分析在许多植物中均有报道, 例如在大豆 (*Glycine max*)^[23]、葡萄 (*Vitis vinifera*)^[24]、石榴 (*Punica granatum*)^[25] 中, 分别鉴定到 57、38、36 个 MIKC 基因。本研究共鉴定了大麻 39 个 *CsMADS* 候选基因, 分布于 9 条染色体上, 基因之间存在串联重复现象, 编码的氨基酸大小位于 146~503 氨基酸之间, 理化性质存在差异, GO 功能分析中, 主要富集于生物过程调控, 亚细胞定位均预测在细胞核中, 说明该家族可能参与大麻遗传代谢调控。

MIKC 型 MADS-box 基因家族在进化过程中具有较高的保守性, 在大麻和拟南芥共线性分析中, 39 个 *CsMADS* 基因共有 14 对旁系同源基因, 进一步与拟南

芥 MIKC 型 MADS-box 家族成员共同构建系统进化树, 可分成 14 个亚族, 其中 *CsMADS3* 和 *CsMADS14* 属于 AP1 亚家族, 根据该亚族在小麦 (*Triticum aestivum*) 中具有调控营养和生殖结构功能^[26], 进一步推测该同源基因可能在调控大麻开花时间方面发挥重要作用; *CsMADS5* 和 *CsMADS8* 基因属于 PI 亚家族, 该亚族在兰科 (*Orchidaceae*) 植物花器官、萼片、花瓣、唇瓣、合蕊柱和未发育子房中均发挥作用^[27]; AG 亚家族基因对于子房授粉后胚珠发育具有调控作用^[28], 在大麻中 *CsMADS1* 和 *CsMADS4* 属于该亚家族, 进一步推测可能对雄蕊、雌蕊和柱头的形态发育具有一定的影响。

39 个 *CsMADS* 基因的保守序列分析表明, 同一亚家族成员的基序数目和序列相似, 保守基序 1 编码的 MADS 结构域, 是最保守的结构域, 其次, K-box 结构域是第二保守的结构域。基因结构分析表明, 同一亚家族具有一定的相似性, *CsMADS* 基因普遍包含 6~7 个内含子, 0~2 个外显子。顺式作用元件存在于基因启动子中, 与转录因子形成特异性结合, 在调控靶基因表达中发挥重要作用^[29]。*CsMADS* 基因启动子序列存在种类丰富和功能多样的顺式作用元件, 其中富含光响应元件, 进一步证明最佳光质量、光量和光周期对于大麻生物量的旺盛生长是非常重要的, 光的质量和数量对光合作用有深刻的影响, 最终影响植物的生长发育^[30], 另外 *CsMADS* 基因能够响应脱落酸、水杨酸、茉莉酸甲酯等激素的调控, 表明这些激素可能在大麻生长发育和应对非生物胁迫中发挥重要作用。据文献^[16]记载大麻腺毛位于雌花苞片的表面, 是富含 7 种大麻素和萜类物质的树脂生物合成和储存的主要场所, 转录组表达谱聚类分析显示 *CsMADS* 基因具有明显的组织特异性, *CsMADS7*、*CsMADS12* 和 *CsMADS29* 在花和

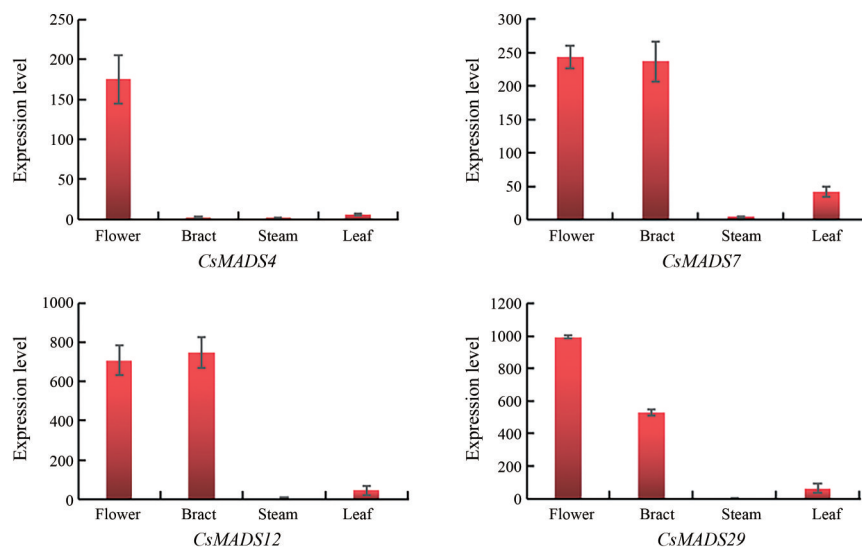


Figure 9 The expression levels of the *CsMADS* genes in different tissues

苞片中表达上调, *CsMADS4* 只在雌花中显著表达, 在其他部位几乎无表达, 说明其可能在植物开花或苞片的生长发育过程中发挥关键作用, 推测其可能调控大麻素生物合成; *CsMADS34*、*CsMADS38*、*CsMADS11*、*CsMADS26*、*CsMADS28*、*CsMADS33* 在花、苞片、茎、叶和种子中均没有表达, 推测以上未表达的基因在该品种某一特定的生长发育过程中未发挥调控作用。通过 RT-qPCR 进行验证, *CsMADS7*、*CsMADS12* 和 *CsMADS29* 在花和苞片中相对表达较高, 且进化树分析都聚类在一起, 均属于 AGL6 亚家族, 根据拟南芥相关功能推测 *CsMADS7*、*CsMADS12* 和 *CsMADS29* 可能在大麻开花期间促进了侧枝器官发育和苞片的形成^[31]。

综上所述, 本研究基于大麻全基因组数据, 筛选和鉴定出 39 个大麻 MIKC 型 MADS-box 基因, 对其进行全面的生物信息学分析, 包括序列特征、染色体定位、GO 富集分类、保守结构域、基因结构、系统发育、共线性分析、顺势元件预测以及表达模式的分析和验证, 能够初步了解大麻 MIKC 型 MADS-box 基因数量、功能分类以及表达趋势, 对后续进一步开展 MIKC 型 MADS-box 基因对药用植物大麻生长发育调控研究具有一定理论价值。

作者贡献: 万志庭负责文章撰写及修改; 鲁梦负责 qRT-PCR 实验部分; 吴沙沙指导文章撰写并提出修改意见; 米要磊负责实验设计及指导撰写思路; 翟俊文负责文章的选题、指导文章撰写及修改。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Elena R, Alvarez B, Soraya P, et al. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals [J]. Proc Natl Acad Sci U S A, 2000, 97: 5328-5333.
- [2] Annette B, Günter T. The major clades of MADS-box genes and their role in the development and evolution of flowering plants [J]. Mol Phylogenet Evol, 2003, 29: 464-489.
- [3] Ma H, Yanofsky MF, Meyerowitz EM. AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes [J]. Genes Dev, 1991, 5: 484-495.
- [4] Kaufmann K, Melzer R, Theissen G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants [J]. Gene, 2005, 347: 183-198.
- [5] Tang QL, Li NZ, Song M, et al. Identification of acting domains mediating the protein interactions between SVP and FLC in *Brassica juncea* Coss [J]. Acta Hort Sin (园艺研究), 2012, 39: 2395-2403.
- [6] Gramzow L, Theissen G. A hitchhiker's guide to the MADS world of plants [J]. Gen Biol, 2010, 11: 214-225.
- [7] Kwantes M, Liebsch D, Verelst W. How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophyte [J]. Mol Biol Evol, 2012, 29: 293-302.
- [8] Soltis DE, Ma H, Frohlich MW, et al. The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression [J]. Trends Plant Sci, 2007, 12: 358-367.
- [9] Guo P. The Phylogenetic Analysis of MADS-box Gene Family in Brassicaceae Plants (十字花科 MADS-box 基因家族系统进化分析) [D]. Fuzhou: Fujian Agriculture & Forestry University, 2018.
- [10] McPartland JM, Guy GW, Hegman W. Cannabis is indigenous to Europe and cultivation began during the Copper or Bronze age: a probabilistic synthesis of fossil pollen studies [J]. Veg Hist Archaeobot, 2018, 27: 635-648.
- [11] William L, Zhang PZ, Ying DY, et al. Extrusion improves the phenolic profile and biological activities of hempseed (*Cannabis sativa* L.) hull [J]. Food Chem, 2021, 346: 128606.
- [12] Irakli M, Tsaliki E, Kalivas A, et al. Effect of genotype and growing year on the nutritional, phytochemical, and antioxidant properties of industrial hemp (*Cannabis sativa* L.) seeds [J]. Antioxidants, 2019, 8: 491.
- [13] ElSohly MA, Sladea D. Chemical constituents of marijuana: the complex mixture of natural cannabinoids [J]. Life Sci, 2005, 78: 539-548.
- [14] Flores-Sanchez JJ, Verpoorte R. Secondary metabolism in cannabis [J]. Phytochem Rev, 2008, 7: 615-639.
- [15] Giacoppo S, Mandolino G, Galuppo M, et al. Cannabinoids: new promising agents in the treatment of neurological diseases [J]. Molecules, 2014, 19: 18781-18816.
- [16] Bai YJ, Zhou XY, Yuan Y, et al. Origin of medicinal *Cannabis sativa* and its early spread [J]. Chin Tradit Herb Drugs (中草药), 2019, 50: 5071-5079.
- [17] Zager JJ, Lange I, Srividya N, et al. Gene networks underlying cannabinoid and terpenoid accumulation in Cannabis [J]. Plant Physiol, 2019, 180: 1877-1897.
- [18] Chen CJ, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data [J]. Mol Plant, 2020, 13: 1194-1202.
- [19] Liu MQ, Sun W, Meng XX, et al. Identification and expression analysis of the C2H2 gene family in *Cannabis sativa* L [J]. Acta Pharm Sin (药学报), 2021, 56: 1486-1496.
- [20] Ren ZY, Yu DQ, Yang ZE, et al. Genome-wide identification of the MIKC-type MADS-box gene family in *Gossypium hirsutum* L. unravels their roles in flowering [J]. Front Plant Sci, 2017, 8: 384-400.
- [21] de Folter S, Shchennikova AV, Franken J, et al. A B_{sister} MADS-box gene involved in ovule and seed development in petunia and Arabidopsis [J]. Plant J, 2006, 47: 934-946.
- [22] Mondragón-Palomino M, Theissen G. Conserved differential

- expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the lowers of Orchidaceae: refining the 'orchid code' [J]. *Plant J*, 2011, 66: 1008-1019.
- [23] Hu RB, Fan CM, Li HY, et al. Analysis of MIKC-type MADS-box genes in Soybean (*Glycine max*) [J]. *Mol Plant Breeding (分子植物育种)*, 2009, 7: 429-436.
- [24] Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, et al. Genome-wide analysis of MIKCC-type MADS-box genes in grapevine [J]. *Plant Physiol*, 2009, 149: 354-369.
- [25] Zhao YJ, Zhao HL, Wang YY, et al. Genome-wide identification and expression analysis of MIKC-Type MADS-Box gene family in *Punica granatum* L. [J]. *Agronomy*, 2020, 10: 1-10.
- [26] Susanne S, Alice K, Sirui P, et al. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization [J]. *New Phytol*, 2020, 225: 511-529.
- [27] He CM, Si C, Jaime A, et al. Genome-wide identification and classification of MIKC-type MADS-box genes in Streptophyte lineages and expression analyses to reveal their role in seed germination of orchid [J]. *BMC Plant Biol*, 2019, 19: 223.
- [28] Li CR, Dong N, Li XP, et al. A review of MADS-box genes, the molecular regulatory genes for floral organ development in Orchidaceae [J]. *Acta Hort Sin (园艺学报)*, 2020, 47: 2047-2062.
- [29] Riechmann JL, Heard J, Martin G, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes [J]. *Science*, 2000, 290: 2105-2110.
- [30] Chandra S, Lata H, Khan IA, et al. Photosynthetic response of *Cannabis sativa* L. to variations in photosynthetic photon flux densities, temperature and CO₂ conditions [J]. *Physiol Mol Biol Plants*, 2008, 14: 299-306.
- [31] Shen GX, Yang CH, Shen CY, et al. Origination and selection of ABCDE and AGL6 subfamily MADS-box genes in gymnosperms and angiosperms [J]. *Biol Res*, 2019, 52: 25.