

药用植物大麻 *C2H2* 基因家族鉴定与表达分析

刘美琦¹, 孙伟², 孟祥霄², 万会花², 刘廷霞¹, 孙嘉莹¹, 王震¹,
米要磊^{2*}, 马伟^{1*}

(1. 黑龙江中医药大学药学院, 黑龙江 哈尔滨 150040; 2. 中国中医科学院中药研究所, 北京 100700)

摘要: *C2H2* 型转录因子在植物生长发育及次生代谢调控中发挥重要作用, 本文利用生物信息学从基因组层面鉴定大麻 (*Cannabis sativa* L.) *C2H2* 基因家族成员, 基于大麻基因组数据及转录组数据, 使用 TBtools、MEGA 软件以及 NCBI、PlantTFDB、ExPASy、HMMSCAN、MEME、WoLFPSORT、PlantCARE 等在线网站对其基因信息、染色体定位、进化关系、保守基序、共线性关系等进行分析。结果显示, 大麻中含有 30 个 *C2H2* 型基因家族成员 (*CsC2H2-1*~*CsC2H2-30*), 分布在大麻的 9 条染色体上, 氨基酸大小介于 138~635 个氨基酸, 理论等电点介于 5.85~9.52, 蛋白质分子质量介于 15 909.48~68 445.53 Da。转录组结果分析显示, *CsC2H2* 在大麻 Diku 品种的雌花、苞片、叶、茎及不同品种雌花中表达量均存在差异。实时荧光定量 PCR 结果验证了 *CsC2H2-1*、*CsC2H2-5*、*CsC2H2-19* 在 Diku 品种的雌花及苞片中表达显著, 为深入研究 *C2H2* 基因家族功能及大麻优质品种的选育提供理论参考。

关键词: 大麻; *C2H2* 基因家族; 生物信息学; 系统进化; 功能预测

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2021)05-1486-11

Identification and expression analysis of the *C2H2* gene family in *Cannabis sativa* L.

LIU Mei-qi¹, SUN Wei², MENG Xiang-xiao², WAN Hui-hua², LIU Ting-xia¹, SUN Jia-ying¹,
WANG Zhen¹, MI Yao-lei^{2*}, MA Wei^{1*}

(1. College of Pharmaceutical Sciences, Heilongjiang University of Chinese Medicine, Harbin 150040, China;

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

Abstract: *C2H2* transcription factors play an important role in plant growth, development and the regulation of secondary metabolism. This article identifies members of the *C2H2* gene family in *Cannabis sativa* L. at the genome level. Chromosomal location and linkage, evolutionary relationships, and identification of conserved motifs was determined from the *C. sativa* genome and transcriptome data using bioinformatics tools and on-line websites such as TBtools, MEGA software, NCBI, PlantTFDB, ExPASy, HMMSCAN, MEME, WoLFPSORT and PlantCARE. The results show that *C. sativa* contains 30 members of the *C2H2* gene family (named *CsC2H2-1*-*CsC2H2-30*) distributed on 9 chromosomes. The encoded proteins range in length from 138 to 635 amino acids, and the theoretical isoelectric points range from 5.85 to 9.52. Molecular weights range from 15 909.48 to 68 445.53 Da. Transcriptome analysis showed that *CsC2H2* was differentially expressed in the female flowers, bracts, leaves, and stems of the Diku variety and female flowers of nine different varieties of *C. sativa*. Quantitative real-time PCR verified that *CsC2H2-1*, *CsC2H2-5*, and *CsC2H2-19* were significantly expressed in the female flowers and bracts of the Diku variety. This provides a theoretical basis for in-depth study of the function of the *C2H2* gene family and the breeding

收稿日期: 2020-12-11; 修回日期: 2021-01-29.

基金项目: 中央级公益性科研院所基本科研业务费专项资金资助 (ZKKT19019); 国家中医药管理局全国中药资源普查项目 (GZY-KJS-2018-004).

*通讯作者 Tel: 86-451-87266988, E-mail: 88788891@qq.com;

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

DOI: 10.16438/j.0513-4870.2020-1902

of high-quality *C. sativa* varieties.

Key words: *Cannabis sativa*; *C2H2* gene family; bioinformatics; phylogeny; functional prediction

C2H2 型基因家族是目前研究最为详尽的锌指家族之一, 在真核生物中分布广泛并能调控基因的表达^[1]。*C2H2* 型锌指蛋白通常含有 X2-Cys-X(2-4)-Cys-X12-His-X(3-5)-His 构成的保守结构域 (Cys: 半胱氨酸, His: 组氨酸, X: 任意氨基酸), 根据氨基酸残基数量和位置的差异, 可以将锌指蛋白转录因子分为 *C2H2*、*C2HC*、*C2HC5*、*C2C2*、*CCCH*、*C3HC4*、*C4*、*C4HC3*、*C6* 和 *C8* 等类型^[2-4]。Takatsuji 等^[5]首次在矮牵牛中发现植物 *C2H2* 型锌指蛋白 EPF1。根据研究报道, 目前已在多种植物中进行了 *C2H2* 基因家族成员的分析, 如拟南芥中有 176 个 *C2H2* 基因, 烟草、番茄、黄瓜、水稻、杨树中分别含有 118、92、120、189、109 个 *C2H2* 家族成员^[6-11]。植物 *C2H2* 型转录因子在植物生长发育、干旱胁迫、盐胁迫、冷胁迫、外源激素胁迫等方面发挥作用, 也在 DNA 识别、RNA 包装、转录激活、凋亡调控以及蛋白质折叠和组装中发挥重要的作用^[1,12]。拟南芥中的 *RBE* 基因参与花原基的早期分化, 影响花瓣的生长

发育过程^[13]。水稻中的 *OsMSR15* 基因诱导其生长发育过程中温度胁迫及干旱胁迫下的基因表达^[14]。番茄中的 *SICZFP1* 基因调控冷胁迫下的基因表达, 是提高作物耐寒性的一个关键基因; *SIZF2* 基因通过影响番茄中花青素和丙二醛的积累, 进而改变叶和花的形态^[15,16]。

大麻 (*Cannabis sativa* L.) 是大麻科 (Cannabaceae) 大麻属 (*Cannabis*) 一年生草本植物, 是一种重要的药用经济作物^[17]。大麻的种植已有数千年的历史, 被一些亚洲和欧洲国家广泛栽培, 在治疗慢性神经性疼痛等方面发挥重要作用^[18,19]。大麻中含有许多酚萜类、萜类以及黄酮类等药用成分^[20]。大麻素为大麻特有的酚萜类化合物, 主要通过脂肪酸途径和萜类途径 (图 1) 生成大麻素生物合成的主要前体大麻萜酚酸 (cannabigerolic acid, CBGA), 再经酶催化作用合成终产物大麻二酚 (cannabidiol, CBD) 和四氢大麻酚 (tetrahydrocannabinol, THC)^[21]。

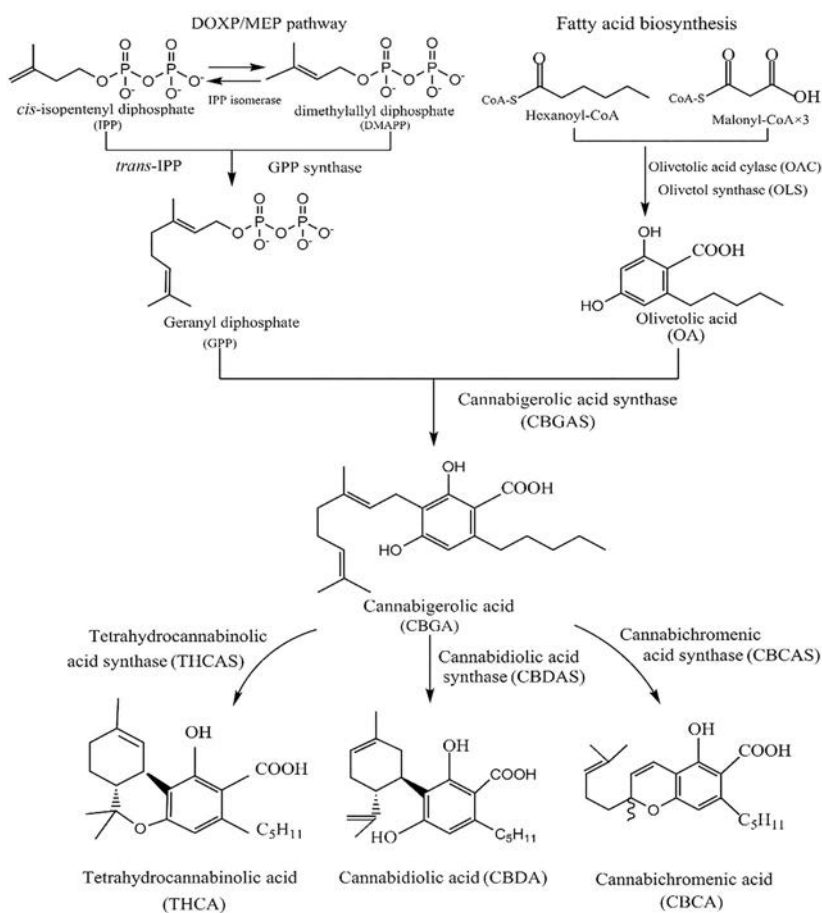


Figure 1 Synthetic pathways of cannabinoids

THC具有较强的致瘾性和致幻性,具有抗癫痫、抗炎及镇痛的作用,但也会引发焦虑、抑郁等精神疾病^[22-26]。CBD类大麻素主要富集在雌性大麻的花及苞片的有柄腺毛的囊腔中,可以抑制高剂量THC引发的焦虑,降低认知障碍和不适感,并有治疗精神分裂症的潜力^[27-30]。

大麻种植历史悠久,具有较高的经济价值和药用价值^[31]。因其含有非致幻性成分CBD等和致幻性成分THC等,不同行业及学者对其进行不同程度的划分,Li等^[32]明确了大麻种质资源管理三级分类体系,即医用大麻(THC > 0.3%)、药用大麻(THC < 0.3%, CBD含量高)和工业大麻(THC < 0.3%, CBD含量低),其中我国法律明令禁止种植THC > 0.3%的大麻。以基因家族角度解析大麻次生代谢产物合成途径,挖掘功能基因对大麻素生物合成的调控机制,继而培育低THC高CBD含量的大麻品种是现阶段生物信息学研究热点。本文基于大麻的基因组数据及转录组数据,对其C2H2转录因子的基因信息、染色体定位情况、进化关系、保守基序、共线性关系等进行分析,探讨C2H2型转录因子对次生代谢的调控作用,并对可能参与大麻素合成的基因进行了预测,为大麻与C2H2型转录因子的后续研究及培育高CBD低THC含量的优质大麻品种提供理论依据。

材料与方法

材料 本研究所应用的大麻的基因组数据(GCA_900626175.1)来源于NCBI (<https://www.ncbi.nlm.nih.gov/>) 数据库^[33], 品种为CRBRx的雌株。拟南芥C2H2基因的数据来自Englbrecht等^[6]的研究,拟南芥基因组数据(GCA_000005425.2)来自NCBI数据库。9个不同品种雌花同一批次的转录组数据(PRJNA498707)获取于NCBI数据库^[34],这9个品种分别为Mama Thai (MT)、White Cookies (WC)、Canna Tsu (CT)、Black Lime (BL)、Terple (T)、Cherry Chem (CC)、Black Berry Kush (BB)、Sour Diesel (SD)和Valley Fire (VF)。大麻品种Diku(名称为Dinamed Kush CBD Autoflowering)是Purple Kush和Dinamed Autoflowering CBD杂交而得的雌性植株,其苞片、茎、花、和叶的转录组数据(NCBI Accession No: 苞片 : SAMN16122880~SAMN16122882; 茎 : SAMN16122883~SAMN16122885; 花 : SAMN16122886~SAMN16122888; 叶 : SAMN16122889~SAMN16122891)来自本课题组。番茄SIZF2转录因子数据来自NCBI数据库(HQ738351)。

C2H2基因家族成员获取及理化性质分析 利用TBtools软件^[35]提取大麻全部蛋白序列,将蛋白序列文件在PlantTFDB网站进行C2H2基因家族成员的预测,并通过HMMSCAN (<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>) 网站去除不含C2H2锌指结构域的成员,通过TBtools软件得到大麻C2H2基因家族的全部蛋白序列。利用ExPASy (<http://web.expasy.org/protparam/>) 在线工具对大麻C2H2锌指蛋白的分子质量、氨基酸大小、等电点、不稳定系数、蛋白质疏水性等理化性质进行预测。采用WoLFPSORT网站 (<https://wolfpsort.hgc.jp/>) 在线预测大麻C2H2锌指蛋白的亚细胞定位状况。利用TBtools软件确定不同的C2H2基因家族成员在染色体上的分布情况,并进行可视化分析。

C2H2基因家族成员基因结构及Motif分析 利用NCBI (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) 网站的CD-search功能对大麻C2H2基因家族成员的保守结构域进行分析。通过MEME (<http://meme-suite.org/tools/meme>) 网站进行保守基序(Motif)分析,识别的Motif数设置为10。将大麻C2H2基因家族成员的进化树、Motif以及保守结构域利用TBtools软件合并处理并可视化分析。

C2H2基因家族成员系统进化分析及共线性分析 利用MEGA-X软件采用邻接法(Neigh-Joining, NJ), Bootstrap值设置为1 000,将大麻与拟南芥的C2H2型锌指蛋白合并构建NJ树,并用Evolview (<https://www.evolgenius.info/evolview>) 在线网站对其美化。利用TBtools软件对大麻、拟南芥的共线性关系进行可视化分析。

C2H2基因家族成员顺式作用元件分析 利用在线网站PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 对大麻C2H2基因启动子上游2 000 bp的序列进行顺式作用元件分析,并将结果用TBtools可视化。

C2H2基因家族成员表达模式分析 利用课题组自测的杂交品种Diku的雌花、叶、苞片、茎的转录组数据以及9个不同品种雌花的数据^[34]进行大麻C2H2基因的表达模式分析,并利用TBtools软件绘制热图,进行可视化分析。

实时荧光定量PCR分析 本实验对转录组数据中表达量差异的CsC2H2-1、CsC2H2-5、CsC2H2-19、CsC2H2-24进行实时荧光定量PCR(quantitative real-time PCR, qRT-PCR)验证,内参基因选择EF1a,利用NCBI设计引物(表1)。实验样品为大麻Diku品种的花、苞片、叶和茎,每组实验3个重复,采用WARYONG

Table 1 Quantitative primers of *CsC2H2s*

Gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
<i>CsC2H2-1</i>	CACGTCTCGAACACCTCCA	AAAGGCGAGAACTCCTCCG
<i>CsC2H2-5</i>	CATCCGAAGGCATAGGGTCC	CAGGCGGATCTTTTGAGCG
<i>CsC2H2-19</i>	TGGCGGTCATATGAGGAGAC	ATAATTCCTCCGGGAGGAGCCA
<i>CsC2H2-24</i>	TTATGGCCCTACCGCTACT	GCGAAAATCGGTCCCAAAGG
<i>EF1a</i>	ACCAAGATTGACAGGCGTTC	CCTTCTTCCACAGCCTTG

试剂盒提取样品 RNA, 用 TransGen 试剂盒进行 cDNA 合成, 以及 FOREVER STAR 试剂盒进行实时荧光定量 PCR 分析。反应程序为: 95 °C 酶激活 5 min, 95 °C 变性 30 s, 56 °C 退火 20 s, 72 °C 延伸 1 s, 循环次数 40。仪器为 Rotor-Gene Q (QIAGEN, 德国)。CsC2H2s 的相对表达量采用 $2^{-\Delta\Delta C_t}$ 法计算。

结果与分析

1 大麻 *C2H2* 基因鉴定及理化性质分析

通过 PlantTFDB 数据库和 HMMSCAN 网站的对比及结果分析, 共从大麻基因组中鉴定出 30 个大麻 *C2H2* 基因家族成员。根据 TBtools 软件将大麻 *C2H2* 基因家族成员进行染色体分布可视化分析 (图 2), 并根据基因在染色体上的分布情况, 依次将 30 个大麻 *C2H2* 基因家族成员命名为 *CsC2H2-1*~*CsC2H2-30*。观察发现, 经鉴定出的大麻 *C2H2* 基因家族成员在大麻的 9 条染色体上 (NC1~9) 呈现出不均匀分布的情况, 其中在 NC1 号染色体上分布的成员较多, 共包含 6 个基因, 在 NC5 和 NC8 号染色体上分布最少, 仅有一个家族成员。有 5 组基因簇呈现在大麻的 4 条染色体上, 依次为 NC1 号中的 *CsC2H2-3* 和 *CsC2H2-4*, NC2 号中的 *CsC2H2-7* 和 *CsC2H2-8*, NC3 号中的 *CsC2H2-11* 和 *CsC2H2-12*, NC7 号中的 *CsC2H2-22* 和 *CsC2H2-23* 以及 *CsC2H2-25* 和 *CsC2H2-26*。

大麻 *C2H2* 基因家族的信息特征及理化性质分析 (表 2、3) 显示, 氨基酸大小介于 138~635 个氨基酸之间, 理论等电点介于 5.85~9.52, 蛋白质分子质量介于 15 909.48~68 445.53 Da。不稳定系数均大于 40, 且 GRAVY 均为负值, 该结果表明大麻的 *C2H2* 蛋白均为不稳定的亲水蛋白。亚细胞定位分析显示, 所有成员均定位在细胞核中, 但一些成员有可能也在细胞质、细胞外基质、叶绿体以及过氧化物酶体中发挥作用。

2 大麻 *C2H2* 基因结构及 Motif 分析

结合大麻 *C2H2* 基因家族的 Motif、UTR (untranslated region, 非编码区)、CDS、内含子以及进化关系, 对基因结构进行可视化分析 (图 3)。其中, 所有 *CsC2H2* 成员中均存在 Motif 1, 其次分别有 22 个、16 个和 15 个成员中存在 Motif 2、Motif 5 和 Motif 7。*CsC2H2-16*、*CsC2H2-22*、*CsC2H2-23*、*CsC2H2-24* 只含有 Motif 1, *CsC2H2-13* 所含保守基序最多。*CsC2H2-2*、*CsC2H2-6*、*CsC2H2-20*、*CsC2H2-25* 和 *CsC2H2-26* 所含保守基序相似, 以上 5 个成员在系统进化关系上 (图 4) 同属 A1a 亚家族, 这可能与行使某些特定的功能有关。在被研究的 30 个大麻 *C2H2* 基因家族成员中, 10 个 *CsC2H2* 成员含有内含子, 内含子数目为 1~5, 内含子数目较少。

通过 MEME 网站对 30 个大麻 *C2H2* 锌指蛋白保守基序进行研究, 将 10 个 Motif 区块进行可视化分析

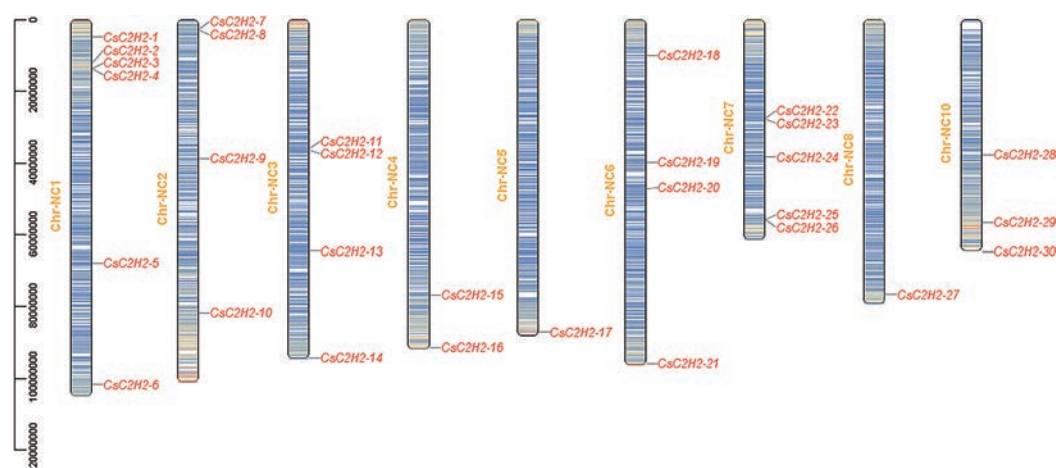


Figure 2 Chromosome locations of *CsC2H2* genes. NC1, 2, 3, 4, 5, 6, 7, 8 and 10 of chromosomes represent NC_044370.1, NC_044371.1, NC_044372.1, NC_044373.1, NC_044374.1, NC_044375.1, NC_044376.1, NC_044377.1, NC_044379.1

Table 2 Information of *CsC2H2* gene family

Gene name	Locus name	Accession No.	Chr (NC)
<i>CsC2H2-1</i>	gene-LOC115703078	rna-XM_030630564.1	1
<i>CsC2H2-2</i>	gene-LOC115723701	rna-XM_030653198.1	1
<i>CsC2H2-3</i>	gene-LOC115716789	rna-XM_030645693.1	1
<i>CsC2H2-4</i>	gene-LOC115716796	rna-XM_030645703.1	1
<i>CsC2H2-5</i>	gene-LOC115713627	rna-XM_030642115.1	1
<i>CsC2H2-6</i>	gene-LOC115700551	rna-XM_030628118.1	1
<i>CsC2H2-7</i>	gene-LOC115703664	rna-XM_030630904.1	2
<i>CsC2H2-8</i>	gene-LOC115703667	rna-XM_030630907.1	2
<i>CsC2H2-9</i>	gene-LOC115706081	rna-XM_030633597.1	2
<i>CsC2H2-10</i>	gene-LOC115706013	rna-XM_030633507.1	2
<i>CsC2H2-11</i>	gene-LOC115710655	rna-XM_030639012.1	3
<i>CsC2H2-12</i>	gene-LOC115708962	rna-XM_030636985.1	3
<i>CsC2H2-13</i>	gene-LOC115709909	rna-XM_030638164.1	3
<i>CsC2H2-14</i>	gene-LOC115711191	rna-XM_030639517.1	3
<i>CsC2H2-15</i>	gene-LOC115714922	rna-XM_030643680.1	4
<i>CsC2H2-16</i>	gene-LOC115713903	rna-XM_030642383.1	4
<i>CsC2H2-17</i>	gene-LOC115717123	rna-XM_030646068.1	5
<i>CsC2H2-18</i>	gene-LOC115719844	rna-XM_030649029.1	6
<i>CsC2H2-19</i>	gene-LOC115721051	rna-XM_030650282.1	6
<i>CsC2H2-20</i>	gene-LOC115720754	rna-XM_030649931.1	6
<i>CsC2H2-21</i>	gene-LOC115719653	rna-XM_030648772.1	6
<i>CsC2H2-22</i>	gene-LOC115723725	rna-XM_030653203.1	7
<i>CsC2H2-23</i>	gene-LOC115723168	rna-XM_030652600.1	7
<i>CsC2H2-24</i>	gene-LOC115723817	rna-XM_030653286.1	7
<i>CsC2H2-25</i>	gene-LOC115723093	rna-XM_030652511.1	7
<i>CsC2H2-26</i>	gene-LOC115722487	rna-XM_030651704.1	7
<i>CsC2H2-27</i>	gene-LOC115695523	rna-XM_030622581.1	8
<i>CsC2H2-28</i>	gene-LOC115699729	rna-XM_030627264.1	10
<i>CsC2H2-29</i>	gene-LOC115698836	rna-XM_030626032.1	10
<i>CsC2H2-30</i>	gene-LOC115701170	rna-XM_030628882.1	10

(图5), 其中 Motif 1 含有 CX2CX3FX5LX2HX3H 结构域, 分布于大麻 C2H2 锌指家族的所有成员中, 表明 Motif 1 是 *CsC2H2* 锌指结构域发挥作用的关键, 其次 Motif 2、Motif 5 和 Motif 7 保守基序分布较多, 丰富了 *CsC2H2* 锌指结构域的种类。

3 大麻 C2H2 型锌指蛋白系统进化分析及共线性分析

通过对大麻和拟南芥的 C2H2 家族成员的系统进化研究, 利用 MEGA-X 软件以邻接法 (neighbor joining, NJ) 构建系统进化树 (图4), 根据 Englbrecht 等^[9]的分类方法进行分类, 系统进化关系显示大麻的 C2H2 家族可分为 10 个亚家族, 分别为 A1a、A1b、A1c、A1d、C1a、C1b、C1-1i、C1-2i、C1-3i 和 C1-3iD, 其中 A1a、A1c、C1a、C1b、C1-1i、C1-2i 和 C1-3i 中分别含有 5、1、2、2、8、8、4 个大麻 C2H2 家族成员, 其他亚家族中无 *CsC2H2* 成员。大麻 C2H2 型转录因子与拟南芥 C2H2 型转录因子在系统进化上的同源关系表明, 两者可能在某些生物学功能上起着相同或相似的作用。番茄 SIZF2 蛋白与 *CsC2H2-1*、*CsC2H2-5* 同源关系较近 (图6)。

大麻和拟南芥中, 共发现 13 条存在共线性关系的

Table 3 Physical and chemical properties of *CsC2H2* gene family.

N indicates nucleus, C indicates cytoplasm, E indicates extracell, H indicates chloroplast, and P indicates peroxisome

Gene name	AA	PI	MW /Da	Instability index	GRAVY	Subcellular location
<i>CsC2H2-1</i>	273	8.71	28 810.79	76.55	-0.728	N
<i>CsC2H2-2</i>	601	8.69	65 789.02	53.18	-0.825	N
<i>CsC2H2-3</i>	165	8.91	17 705.17	54.5	-0.483	N, H, E
<i>CsC2H2-4</i>	165	8.91	17 705.17	54.5	-0.483	N, H, E
<i>CsC2H2-5</i>	264	8.87	28 367.54	56.35	-0.645	N
<i>CsC2H2-6</i>	539	9.02	56 238.16	53.25	-0.33	N
<i>CsC2H2-7</i>	274	8.99	31 140.52	51.89	-0.711	N, E
<i>CsC2H2-8</i>	212	8.66	24 111.93	55.45	-1.074	N
<i>CsC2H2-9</i>	366	9.13	42 191.41	55.07	-0.811	N
<i>CsC2H2-10</i>	464	6.97	51 112.83	68.49	-0.997	N
<i>CsC2H2-11</i>	138	8.36	15 909.48	66.04	-1.628	N
<i>CsC2H2-12</i>	357	7.71	39 125.65	48.82	-0.727	N
<i>CsC2H2-13</i>	556	8.77	61 741.4	58.69	-1.013	N
<i>CsC2H2-14</i>	149	9.37	16 403.73	43.01	-0.425	N
<i>CsC2H2-15</i>	356	5.85	39 711.97	47.8	-0.478	N
<i>CsC2H2-16</i>	287	7.72	32 273.8	66.3	-0.93	N
<i>CsC2H2-17</i>	335	8.31	37 094.64	55.36	-0.736	N
<i>CsC2H2-18</i>	333	9.52	36 418.6	55.65	-0.862	N
<i>CsC2H2-19</i>	192	8.47	21 415.37	60.51	-0.66	N, P
<i>CsC2H2-20</i>	558	8.49	60 912.44	49.82	-0.719	N
<i>CsC2H2-21</i>	374	6.07	42 994.72	59.15	-1.106	N, C
<i>CsC2H2-22</i>	209	5.85	22 882.49	61.42	-1.087	N
<i>CsC2H2-23</i>	365	6.35	41 101.21	61.74	-0.994	N
<i>CsC2H2-24</i>	271	6.8	30 374.44	51.36	-0.931	N
<i>CsC2H2-25</i>	627	8.67	68 445.53	50.42	-0.889	N, C
<i>CsC2H2-26</i>	635	9.2	67 621.1	54.45	-0.818	N, C
<i>CsC2H2-27</i>	312	8.95	34 189.54	45.15	-0.846	N
<i>CsC2H2-28</i>	493	8.41	54 380.14	43.65	-0.725	N
<i>CsC2H2-29</i>	570	6.85	63 518.01	45.03	-0.744	N
<i>CsC2H2-30</i>	277	6.95	30 948.4	72.89	-0.774	N

基因 (图7), 分别为 *CsC2H2-1/AT3G19580*、*CsC2H2-3/AT2G17180*、*CsC2H2-3/AT4G35280*、*CsC2H2-6/AT5G66730*、*CsC2H2-10/AT3G10470*、*CsC2H2-10/AT5G04390*、*CsC2H2-12/AT5G03510*、*CsC2H2-16/AT3G23130*、*CsC2H2-17/AT4G16610*、*CsC2H2-19/AT5G59820*、*CsC2H2-25/AT1G14580*、*CsC2H2-25/AT2G02070* 和 *CsC2H2-29/AT5G15480*。大麻与拟南芥在共线性分析中有同源关系的基因, 在系统发育中同样为近缘关系, 表明基因在进化的过程中具有一定的保守性, 其可能以相同或相似的遗传形式作用于植物体。

4 C2H2 基因家族成员顺式作用元件分析

对基因家族成员的启动子序列 (基因上游 2 000 bp) 进行顺式作用元件预测, 结果 (图8) 显示, 大麻 C2H2 基因家族 30 个成员中全部含有光响应元件, 且光响应元件数目最多, 数目最少的为缺氧特异性诱导响应元件, 只存在于 *CsC2H2-20*, 其次为类黄酮生物合成基因调控元件, 存在于 *CsC2H2-7* 和 *CsC2H2-8*。该基因家



Figure 3 Genetic structure and conserved motif analysis of *CsC2H2*

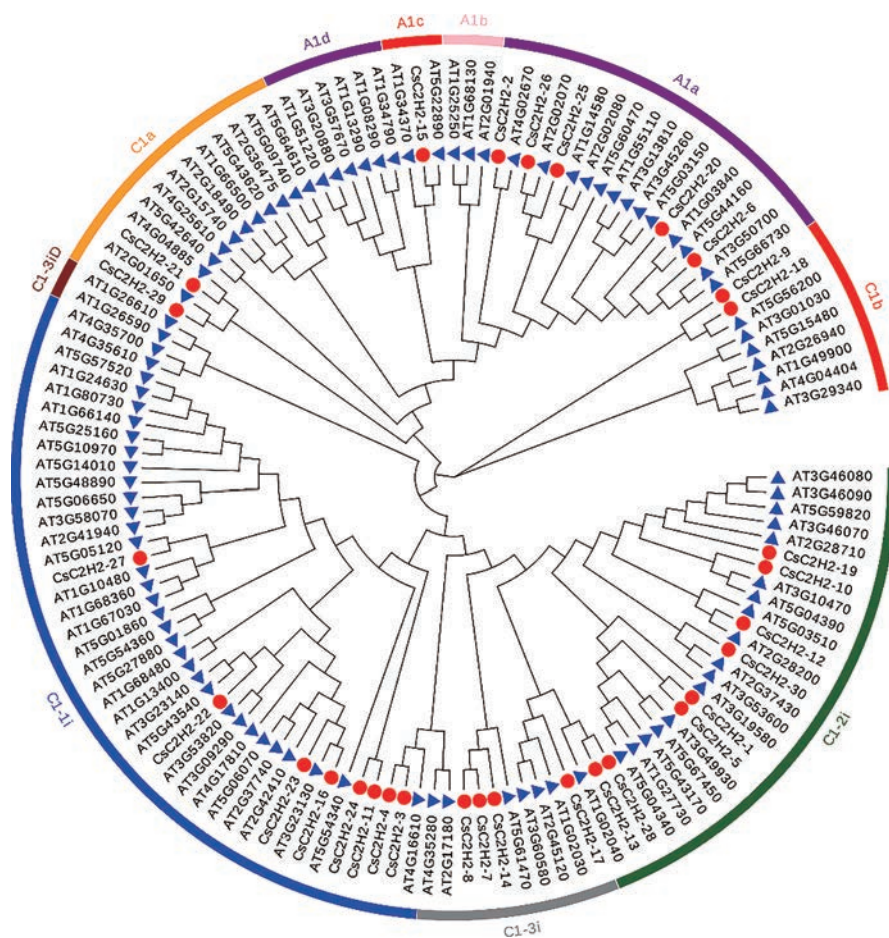


Figure 4 Phylogenetic tree of the C2H2 zinc finger protein of *C. sativa* (●) and *Arabidopsis thaliana* (▲)

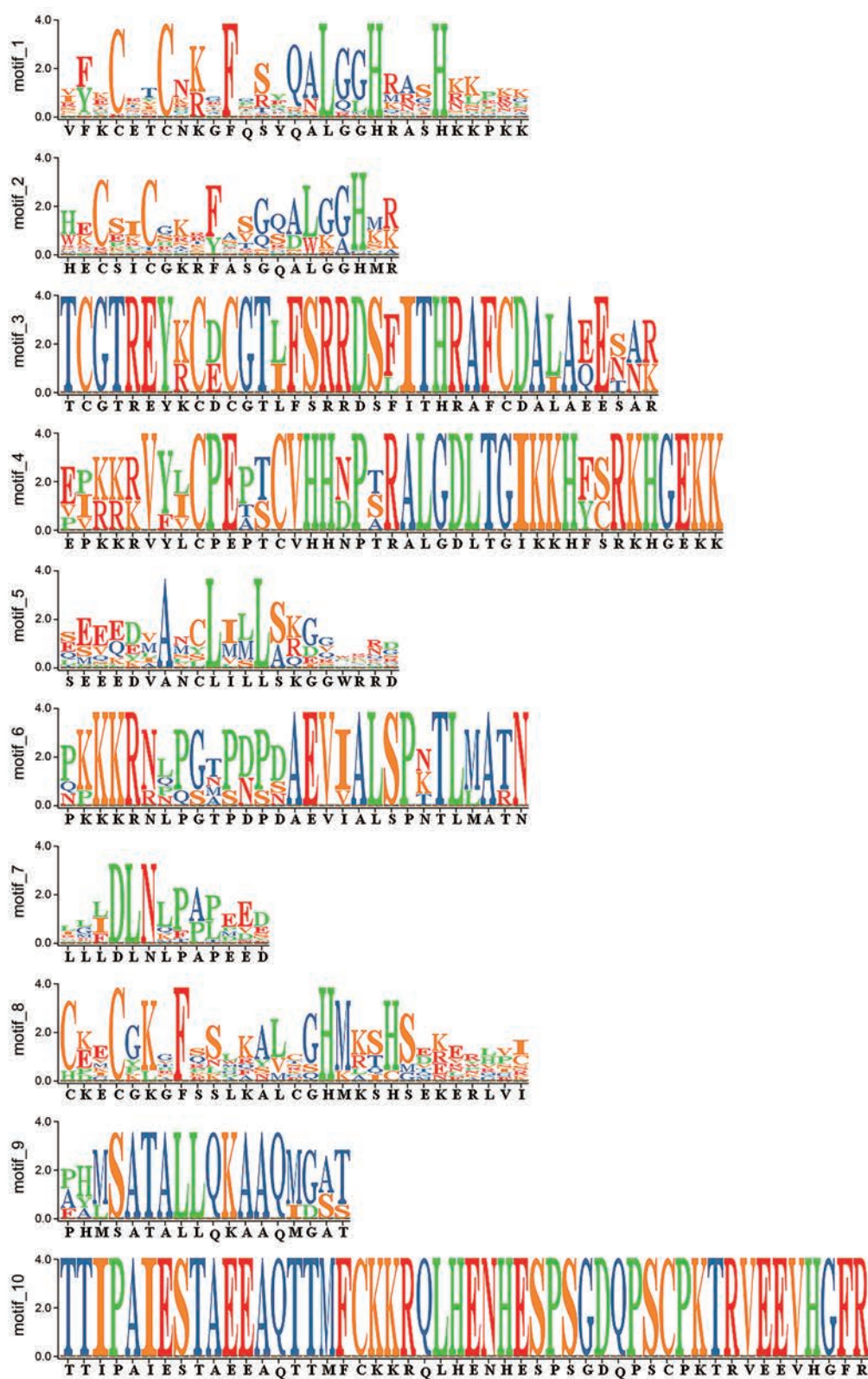


Figure 5 Ten conserved motifs of the CsC2H2 zinc finger protein

族还含有茉莉酸甲酯诱导响应元件、赤霉素响应元件、脱落酸响应元件等多种顺式作用元件,可见该基因家族顺式作用元件种类繁多、功能多样。

5 C2H2 基因家族成员表达模式分析

通过对课题组自测的 Diku 品种以及源自数据库中的 9 个不同品种雌花的转录组数据进行聚类

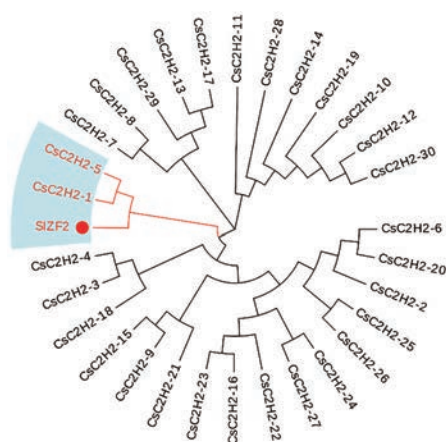


Figure 6 Unrooted neighbor-joining phylogenetic tree of SIZF2 and CsC2H2

分析, 构建大麻中 *C2H2* 基因表达模式的热图 (图 9A、B), 在 Diku 品种的聚类分析中 (图 9A), *CsC2H2-1*、*CsC2H2-5* 和 *CsC2H2-19* 表达量高于其他基因, 说明以上三者可能参与了 Diku 品种的生长发育过程, 并且其在苞片中表达量均为最高, 说明其可能在 Diku 品种苞片的生长发育过程中发挥关键作用。*CsC2H2-24* 在 Diku 品种的叶中特异性表达, 说明该基因可能只

调控叶的生长发育。*CsC2H2-3*、*CsC2H2-4*、*CsC2H2-7*、*CsC2H2-8*、*CsC2H2-11*、*CsC2H2-16*、*CsC2H2-18*、*CsC2H2-23*、*CsC2H2-27* 在花、叶、苞片和茎中均没有表达, 推测以上未表达的基因可能是环境依赖型转录因子或者在 Diku 品种某一特定的生长发育过程中参与表达。*CsC2H2* 在 9 个不同品种及大麻 Diku 品种雌花的表达中 (图 9B), *CsC2H2-1*、*CsC2H2-5* 和 *CsC2H2-19* 同样表达量最高; *CsC2H2-3*、*CsC2H2-4*、*CsC2H2-11*、*CsC2H2-18*、*CsC2H2-27* 未参与表达, 其他基因表达量相对低水平呈现。

6 *CsC2H2* 基因家族成员的实时荧光定量 PCR 分析

利用 qRT-PCR 检测 *CsC2H2-1*、*CsC2H2-5*、*CsC2H2-19*、*CsC2H2-24* 在大麻 Diku 品种不同组织部位的表达情况, 结果 (图 10) 显示, *CsC2H2-1*、*CsC2H2-5*、*CsC2H2-19* 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, 但整体来说这三个基因在花、苞片和茎中的表达量高于在叶中的表达, 且 *CsC2H2-1*、*CsC2H2-5*、*CsC2H2-19* 在花和苞片中表达量极高, *CsC2H2-24* 表达量极低甚至不表达, 证明 qRT-PCR 结果与转录组结果一致。

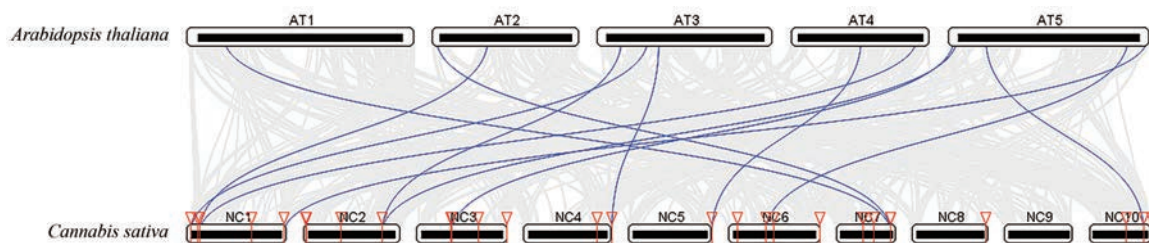


Figure 7 The collinear relationship between *C. sativa* and *Arabidopsis thaliana*. Red indicates the distribution of *CsC2H2* genes on chromosomes, and blue indicates the collinearity between *C. sativa* and *Arabidopsis thaliana*

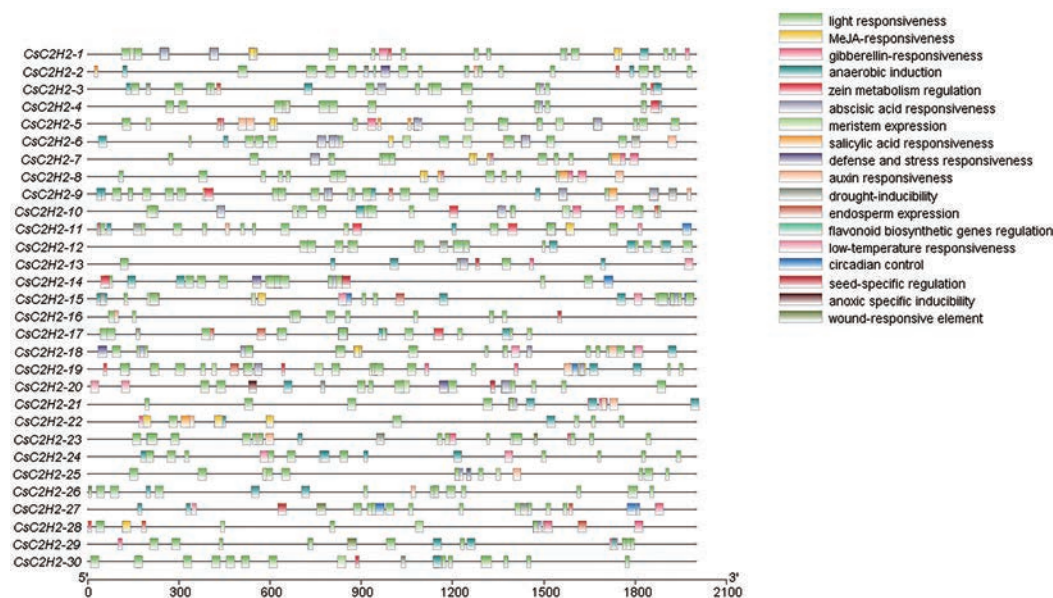


Figure 8 The *cis*-acting elements within the promoters of *CsC2H2*

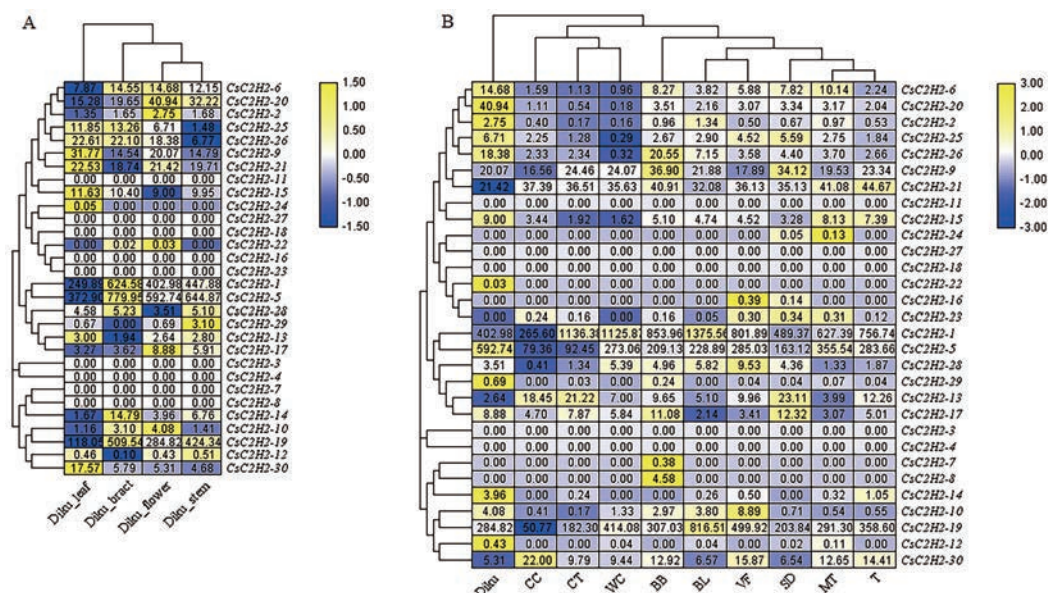


Figure 9 Expression analysis of *CsC2H2* genes. A: *CsC2H2* gene expression patterns in leaf, bract, flower, and stem of Diku variety; B: Expression of *CsC2H2* in female flowers of ten different varieties. (Diku: Dinamed Kush CBD Autoflowering; CC: Cherry Chem; CT: Canna Tsu; WC: White Cookies; BB: Black Berry Kush; BL: Black Lime; VF: Valley Fire; SD: Sour Diesel; MT: Mama Thai; T: Terple)

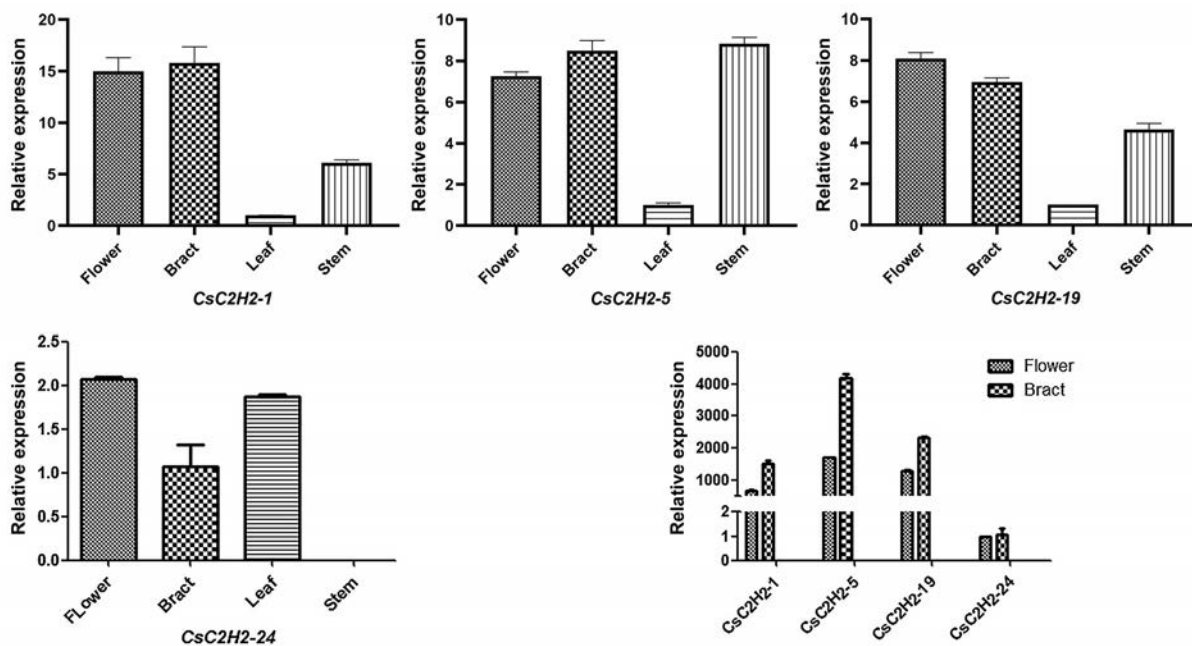


Figure 10 The relative expression levels of the *CsC2H2s* in different tissues

讨论

本研究基于生物信息学分析,对大麻 *C2H2* 基因家族的染色体定位情况、进化关系、保守基序、共线性关系等可视化构图并分析。*C2H2* 型转录因子广泛存在于真核生物中,研究表明, *C2H2* 基因可能在植物的生长发育、水分胁迫及次生代谢等方面发挥重要的作用,基于大麻基因组数据分析,共鉴定出 30 个 *CsC2H2* 基因家族成员,分布在大麻的 9 条染色体上,各成员之

间理化性质存在差异。保守基序分析显示, Motif 1 在各个成员中均有体现,在大麻 Motif 1 及 Motif 2 结构中,存在 QALGGH 基序,该基序主要存在于植物蛋白中,在 *C2H2* 锌指结构域中相对较为保守, QALGGH 基序对 DNA 结合活性至关重要,并且也在植物的生长发育过程中发挥关键作用^[4,36,37]。

在大麻染色体中,有 5 组基因簇串联重复,表明了 *C2H2* 基因在进化过程中自身发生了复制。大麻与拟

南芥间的共线性分析显示, 大麻的某些 C2H2 基因在进化的过程中与拟南芥的基因存在着同源关系, 可见 C2H2 基因在进化的过程中存在一定的功能保守性。AZF2 (AT3G19580) 基因与拟南芥水分胁迫反应有关, 可见其同源基因 CsC2H2-1 也可能在大麻中发挥相似作用^[38]。

转录组结果分析显示, CsC2H2-1、CsC2H2-5 和 CsC2H2-19 在 Diku 品种及 9 个不同品种雌花中均表达显著, 表明其为大麻生命活动过程中的关键转录因子, 可能参与次生代谢产物的合成过程。qRT-PCR 结果验证了以上三者在花及苞片中表达显著, 其他未参与表达的基因可能是环境依赖型转录因子或者在大麻某一特定的生长发育过程中参与表达, 但具体表达形式还有待进一步分析。大麻素属于酚萜类化合物, 其生物合成途径包括酚类化合物与萜类化合物。酚类化合物中的黄酮类及其衍生物具有抗肿瘤、抗衰老、提高机体免疫等药理作用^[39,40]。花青素是酚类中的类黄酮物质, 是广泛存在于植物体内的一类次级代谢产物^[41]。拟南芥中的 ZAT6 (AT5G04340) 基因^[42]正向调节花青素和总黄酮的浓度; SIZF2 基因^[16]在番茄的生长发育进程中广泛表达, 易受盐胁迫及干旱胁迫的诱导, 该 C2H2 基因通过影响花青素和丙二醛的积累, 进而改变叶和花的形态。在系统进化关系上, CsC2H2-1、CsC2H2-5 与 ZAT6 及 SIZF2 同源关系较高且在大麻苞片的腺毛中表达量最高, 由此表明 CsC2H2-1、CsC2H2-5 可能参与大麻次生代谢过程中酚类化合物的合成, 进而影响大麻素的合成。SIZF2 基因亦有通过脱落酸生物信号传导影响植株发育的作用, 且对基因成员上游 2 000 bp 的启动子序列的顺式作用元件预测显示, CsC2H2-1、CsC2H2-5 均含有脱落酸响应元件, 说明大麻可能受脱落酸信号介导次生代谢产物的合成^[16]。

本研究利用生物信息学对大麻的 C2H2 基因家族进行分析, 验证了 CsC2H2-1、CsC2H2-5、CsC2H2-19 在大麻 Diku 品种花及苞片中表达显著, 证明以上基因在植物的生长发育过程及次生代谢调控中发挥重要作用, 有助于更加全面地了解基因家族的生物学功能, 为培育高 CBD、低 THC 的优质大麻品种及 C2H2 型基因家族的功能研究提供理论支撑。

作者贡献: 刘美琦负责文章撰写及修改; 孙伟负责文章的选题; 孟祥霄、万会花负责实验材料的收集; 刘廷霞负责 qRT-PCR 实验部分; 孙嘉莹、王震指导文章撰写并提出修改意见; 米要磊负责实验设计及论文审阅; 马伟指导论文写作及修改。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Zhang J, Liu JF, Zhao TT, et al. Research progress of C2H2 zinc finger protein in plant [J]. Mol Plant Breed (分子植物育种), 2018, 16: 427-433.
- [2] Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys2 His2 zinc finger proteins [J]. Annu Rev Biochem, 2001, 70: 313-340.
- [3] Moore M, Ullman C. Recent developments in the engineering of zinc finger proteins [J]. Brief Funct Genomic Proteomic, 2003, 1: 342-355.
- [4] Berg JM, Shi Y. The galvanization of biology: a growing appreciation for the roles of zinc [J]. Science, 1996, 271: 1081-1085.
- [5] Takatsuji H, Mori M, Benfey PN, et al. Characterization of a zinc finger DNA-binding protein expressed specifically in Petunia petals and seedlings [J]. EMBO J, 1992, 11: 241-249.
- [6] Englbrecht CC, Schoof H, Böhm S. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome [J]. BMC Genomics, 2004, 5: 39.
- [7] Yang ML, Chao JT, Wang DW, et al. Genome-wide identification and expression profiling of the C2H2-type zinc finger protein transcription factor family in tobacco [J]. Hereditas (遗传), 2016, 38: 337-349.
- [8] Hu JH, Wang DW, Yang ML, et al. Identification and bioinformatics analysis of C2H2 zinc finger protein in tomato [J]. Jiangsu Agric Sci (江苏农业科学), 2018, 46: 23-27.
- [9] Zhang S. Genome-wide Identification and Characterization of Zinc Finger Proteins in Cucumber (黄瓜基因组内锌指蛋白的鉴定及特征分析) [D]. Beijing: Chinese Academy of Agricultural Sciences, 2017.
- [10] Agarwal P, Arora R, Ray S, et al. Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis [J]. Plant Mol Biol, 2007, 65: 467-485.
- [11] Liu Q, Wang Z, Xu X, et al. Genome-wide analysis of C2H2 zinc-finger family transcription factors and their responses to abiotic stresses in Poplar (*Populus trichocarpa*) [J]. PLoS One, 2015, 10: e0134753.
- [12] Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity [J]. Curr Opin Struct Biol, 2001, 11: 39-46.
- [13] Takeda S, Matsumoto N, Okada K. Rabbit ears, encoding a superman-like zinc finger protein, regulates petal development in *Arabidopsis thaliana* [J]. Development, 2004, 131: 425-434.
- [14] Zhang X, Zhang B, Li MJ, et al. OsMSR15 encoding a rice C2H2-type zinc finger protein confers enhanced drought tolerance in transgenic *Arabidopsis* [J]. J Plant Biol, 2016, 59: 271-281.
- [15] Zhang X, Guo X, Lei C, et al. Overexpression of SICZFP1, a novel TFIIIA-type zinc finger protein from tomato, confers enhanced cold tolerance in transgenic *Arabidopsis* and rice [J]. Plant Mol Biol Rep, 2011, 29: 185-196.

- [16] Hichri I, Muhovski Y, Žižková E, et al. The solanum lycopersicum zinc finger2 cysteine-2/ histidine-2 repressor-like transcription factor regulates development and tolerance to salinity in tomato and Arabidopsis [J]. Plant Physiol, 2014, 164: 1967-1990.
- [17] Li J, Zhu XW, Wan HH, et al. Progress on chemical constituents and analytical methods of cannabinoids in *Cannabis sativa* [J]. Chin Tradit Herb Drugs (中草药), 2020, 51: 6414-6425.
- [18] Schultes RE, Klein WM, Plowman T. Cannabis: an example of taxonomic neglect [J]. Bot Museum Leaflets, Harvard Univ, 1974, 23: 337-367. http://www.jstor.org/stable/41762285?seq=1&cid=pdf-reference#references_tab_contents.
- [19] Ware MA, Wang T, Shapiro S, et al. Smoked cannabis for chronic neuropathic pain: a randomized controlled trial [J]. CMAJ, 2010, 182: E694-E701.
- [20] Flores-Sanchez IJ, Choi YH, Verpoorte R. Metabolite analysis of *Cannabis sativa* L. by NMR spectroscopy [J]. Methods Mol Biol, 2012, 815: 363-375.
- [21] Preedy V. The Biosynthesis of Cannabinoids [M]/Handbook of Cannabis and Related Pathologies. Chapter 2. San Diego: Academic Press, 2017: 13-23.
- [22] Chang XW, Chen WQ, Sun Y, et al. Potential medical use and risks of cannabis [J]. Chin J Drug Depend (中国药物依赖性杂志), 2020, 29: 161-168.
- [23] Rácz I, Nent E, Erxlebe E, et al. CB1 receptors modulate affective behaviour induced by neuropathic pain [J]. Brain Res Bull, 2015, 114: 42-48.
- [24] Romero-Sandoval EA, Kolano AL, Alvarado-Vázquez PA. Cannabis and cannabinoids for chronic pain [J]. Curr Rheumatol Rep, 2017, 19: 67.
- [25] Cristino L, Bisogno T, Di Marzo V. Cannabinoids and the expanded endocannabinoid system in neurological disorders [J]. Nat Rev Neurol, 2019, 16: 9-29.
- [26] Chopra K, Arora V. An intricate relationship between pain and depression: clinical correlates, coactivation factors and therapeutic targets [J]. Expert Opin Ther Targets, 2014, 18: 159-176.
- [27] Zhang JQ, Chen SL, Wei GF, et al. Cultivars breeding and production of non-psychoactive medicinal cannabis with high CBD content [J]. China J Chin Mater Med (中国中药杂志), 2019, 44: 4772-4780.
- [28] Rodriguez de Fonseca F, Cebeira M, Hernández ML, et al. Changes in brain dopaminergic indices induced by perinatal exposure to cannabinoids in rats [J]. Dev Brain Res, 1990, 51: 237-240.
- [29] Bergamaschi MM, Queiroz RH, Chagas MH, et al. Cannabidiol reduces the anxiety induced by simulated public speaking in treatment-naive social phobia patients [J]. Neuropsychopharmacology, 2011, 36: 1219-1226.
- [30] Levin R, Peres FF, Almeida V, et al. Effects of cannabinoid drugs on the deficit of prepulse inhibition of startle in an animal model of schizophrenia: the SHR strain [J]. Front Pharmacol, 2014, 5: 10.
- [31] Luo X, Reiter MA, d'Espaux L, et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast [J]. Nature, 2019, 567: 123-126.
- [32] Li QS, Meng Y, Chen SL. A new *Cannabis* germplasm classification system and research strategies of non-psychoactive medicinal cannabis [J]. China J Chin Mater Med (中国中药杂志), 2019, 44: 4309-4316.
- [33] Grassa CJ, Wenger JP, Dabney C, et al. A complete *Cannabis* chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content [J]. BioRxiv, 2018. DOI: 10.1101/458083.
- [34] Zager JJ, Lange I, Srividya N, et al. Gene networks underlying cannabinoid and terpenoid accumulation in *Cannabis* [J]. Plant Physiol, 2019, 180: 1877-1897.
- [35] Chen C, Xia R, Chen H, et al. TBtools, a toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface [J]. BioRxiv, 2018. DOI: 10.1101/289660.
- [36] Lee BM, Xu J, Clarkson BK, et al. Induced fit and "lock and key" recognition of 5 S RNA by zinc fingers of transcription factor IIIA [J]. J Mol Biol, 2006, 357: 275-291.
- [37] Takatsuji H. Zinc-finger transcription factors in plants [J]. Cell Mol Life Sci, 1998, 54: 582-596.
- [38] Sakamoto H, Araki T, Meshi T, et al. Expression of a subset of the *Arabidopsis* Cys2/His2-type zinc-finger protein gene family under water stress [J]. Gene, 2000, 248: 23-32.
- [39] Li XP, Yu MX, Kuang TR, et al. Research progress on the anti-tumor effect of flavonoid derivatives [J]. Acta Pharm Sin (药学报), 2021, 56: 913-923.
- [40] Li MR, Zhou YZ, Du GH, et al. Research progress about the anti-aging effect and mechanism of flavonoids from traditional Chinese medicine [J]. Acta Pharm Sin (药学报), 2019, 54: 1382-1391.
- [41] Harborne JB, Williams CA. Anthocyanins and other flavonoids [J]. Nat Prod Rep, 2001, 18: 310-333.
- [42] Shi H, Liu G, Wei Y, et al. The zinc-finger transcription factor ZAT6 is essential for hydrogen peroxide induction of anthocyanin synthesis in *Arabidopsis* [J]. Plant Mol Biol, 2018, 97: 165-176.