

工业大麻PP2C基因家族成员鉴定及表达分析

蔡晓雪¹, 王思凡², 米要磊², 万会花², 曹雪², 孙伟², 苏畅³,
陈士林², 徐艳琴^{1*}, 陈伟强^{2*}

- (1. 江西中医药大学药学院, 南昌 330004;
2. 中国中医科学院中药研究所中药鉴定与安全性评估北京市重点实验室, 北京 100700;
3. 深圳市药品检验研究院, 广东深圳 518057)

[摘要] 目的: 2C类蛋白磷酸酶(PP2C)参与植物多种信号转导途径。其通过负调节丝裂原活化蛋白激酶(MAPK)信号通路参与植物多种胁迫响应和代谢产物合成的调控。对工业大麻PP2C(CsPP2Cs)基因家族进行全基因组鉴定及表达分析, 以为研究CsPP2Cs在工业大麻生长发育过程中的生物学功能提供参考。方法: 利用MEGA-X构建系统发育树, 利用ExPASy、WoLF PSORT、MEME、Batch-CD-Search、PlantCare、TBtools分别对CsPP2Cs蛋白理化性质、亚细胞定位、保守基序、蛋白结构域、CsPP2Cs基因启动子顺式作用元件和与拟南芥PP2Cs(*AtPP2Cs*)基因的共线性进行预测和分析。通过转录组数据和实时荧光定量聚合酶链式反应(Real-time PCR)对CsPP2Cs基因表达进行分析和验证。结果: 从工业大麻全基因组中共鉴定出52个具有保守结构域的PP2Cs, 基因编码蛋白长度244~1 089个氨基酸不等, 相对分子质量在26.76~122.53 kDa, 亚细胞定位主要分布于细胞核、细胞质和叶绿体。系统进化树将52个CsPP2Cs分为10个亚族和5个未分组成员。共线性分析发现, 工业大麻与拟南芥存在7对同源基因。顺式元件预测显示光响应元件和脱落酸元件居多。基因表达热图显示, 相同基因在各组织中存在差异表达。对部分基因进行Real-time PCR验证, 进一步证实了转录组数据的准确性。同时, 可变剪切分析表明部分CsPP2Cs基因在进化过程中存在可变剪切本。结论: 从全基因组层面对CsPP2Cs进行了分析和预测, 提示CsPP2Cs可能广泛参与工业大麻的多种生物学过程, 为进一步研究CsPP2Cs功能奠定了基础。

[关键词] 工业大麻; 2C类蛋白磷酸酶(PP2C)基因家族; 成员鉴定; 功能预测; 表达分析

[中图分类号] R284.2; R289; R22; R2-031; R33 **[文献标识码]** A **[文章编号]** 1005-9903(2022)19-0162-11

[doi] 10.13422/j.cnki.syfjx.20221111

[网络出版地址] <https://kns.cnki.net/kcms/detail/11.3495.R.20220624.1554.004.html>

[网络出版日期] 2022-06-27 7:00

Identification and Expression Analysis of PP2C Gene Family Members in *Cannabis sativa*

CAI Xiaoxue¹, WANG Sifan², MI Yaolei², WAN Huihua², CAO Xue², SUN Wei², SU Chang³,
CHEN Shilin², XU Yanqin^{1*}, CHEN Weiqiang^{2*}

- (1. School of Pharmacy, Jiangxi University of Chinese Medicine, Nanchang 330004, China;
2. Institute of Chinese Materia Medica, Chinese Academy of Chinese Medical Science, Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Beijing 100700, China;
3. Shenzhen Institute for Drug Control, Shenzhen 518057, China)

[Abstract] **Objective:** The type 2C protein phosphatases (PP2C) are involved in numerous plant signal transduction pathways. They mainly participate in plant stress response and regulate second metabolites biosynthesis via negatively regulating MAPK signaling pathway. Herein, we were to identify and analyze PP2C

[收稿日期] 2022-03-21

[基金项目] 中国中医科学院科技创新工程项目(CI2021A04113); 国家自然科学基金项目(82060684)

[第一作者] 蔡晓雪, 在读硕士, 从事中药资源研究, E-mail: alinaa777@163.com

[通信作者] * 陈伟强, 助理研究员, 从事中药资源、中药生物技术研究, E-mail: wqchen@icmm.ac.cn;

* 徐艳琴, 教授, 从事中药资源研究, E-mail: yqxutcm@163.com

(CsPP2C) gene family from hemp genome, in hope of providing comprehensive insights for studying CsPP2C function during the development of hemp. **Method:** Molecular Evolutionary Genetics Analysis (MAGA)-X was used to construct phylogenetic tree. Expert Protein Analysis System (ExpASy), WoLF PSORT, Multiple EM for Motif Elicitation (MEME), Batch Conserved Domain Search (Batch-CD-Search), PlantCare, and TBtools were used, respectively, to predict CsPP2C physicochemical properties, subcellular localization, conserved motifs, protein structure, *cis*-element in promoter and collinearity with *Arabidopsis PP2C*. *Cannabis sativa* transcriptome and Real-time polymerase chain reaction (Real-time PCR) were used to analyze and verify gene expressions, respectively. **Result:** Fifty-two CsPP2C with conserved domains were identified from the entire genome of hemp, encoding proteins ranging from 244 to 1 089 aa in length and with molecular weights ranging from 26.76 to 122.53 kDa. Those genes were mainly distributed in the nucleus, cytoplasm and chloroplast. The 47 CsPP2C were divided into 10 subfamilies, and the remaining 5 were not clustered. Seven pairs of homologous genes between hemp and *Arabidopsis thaliana* were identified according to collinear analysis. The light-responsive elements and abscisic acid elements are most abundant in the prediction. The gene expression heat map showed varied expression pattern of *CsPP2C* in different tissues. Real-time PCR results of three *CsPP2C* were consistent with transcriptome data. Moreover, alternative splicing analysis showed that some *CsPP2C* had alternative-splicing genes during evolution. **Conclusion:** We predicted and analyzed CsPP2C gene family in genomic scale and showed that CsPP2C are involved in many biological processes, whereby provides foundation for CsPP2C functional study.

[Keywords] hemp; type 2C protein phosphatases (PP2C) gene family; gene identification; functional prediction; expression analysis

大麻 (*Cannabis sativa*) 是大麻科 (Cannabinaceae) 大麻属 (*Cannabis*) 一年生草本植株, 多雌雄异株, 生长适应能力强, 各国均有分布^[1]。大麻具有广泛的药用价值, 食用价值和工业价值。火麻仁为成熟的大麻种子, 是中医常用的润下药, 始载于《神农本草经》“补中益气, 久服肥健”, 对消化系统、心血管系统、中枢神经系统有不同程度的药理作用, 可用于治疗肠燥便秘、高血压、高血脂、风湿性疼痛等^[2-3]。此外, 火麻仁含有丰富的不饱和脂肪酸、蛋白质、氨基酸、维生素等成分。食用火麻油、火麻酒等火麻仁食品, 对人体具有一定的保健作用^[4]。大麻化学成分复杂, 可以分为大麻素和其他类化合物^[5]。大麻素为具有 C21 骨架的次生代谢产物, 主要存在于大麻的花和苞片中, 具有镇痛、降眼压、抗肿瘤、抗炎等药理作用, 对于治疗癌症效果显著, 可通过抑制细胞活性和诱导细胞凋亡, 治疗乳腺癌、前列腺癌、肝细胞癌、胰腺癌等^[6-7]。大麻富含植物纤维, 韧皮纤维是最好的天然纤维之一, 是上等的纺织和造纸原材料^[8]。大麻纤维所制成的衣物具有良好的透气性、吸汗排湿, 导热性能出色, 同时因其含有酚类物质及其衍生物, 具有较好的抗菌性, 深受消费者喜爱^[9-10]。根据大麻不同的使用价值可以将大麻分为用于收获种子、纤维及非精神活

性大麻二酚的工业大麻或汉麻, 具有精神活性物质四氢大麻酚的毒品大麻、娱乐大麻或药用大麻两类^[11]。大麻最早在新石器时代时在东亚地区被驯化, 目前所有的工业大麻和毒品品种都是从中国的野生植物和地方品种代表的祖先基因库中分化出来的^[12]。

2C 类蛋白磷酸酶 (PP2C) 为丝氨酸/苏氨酸残基蛋白磷酸酶, 以单体的形式存在, 其活性依赖于镁离子或锰离子^[13]。PP2C 为多功能蛋白磷酸酶, 其结构特殊, C 端为保守的催化区, N 端为长短不一的延伸区。延伸区域的差异导致 PP2C 具有多种生物学功能^[13-15]。PP2C 基因家族成员参与植物多种生理途径, 包括脱落酸 (ABA) 信号途径、调控植物生长发育、植物创伤信号途径、植物抗病信号途径等^[16]。PP2C 基因家族 A 亚族参与 ABA 信号转导途径。例如, 拟南芥 PP2C (AtPP2C) 基因家族 A 亚族 9 个成员启动子区域均含有 ABA 响应元件, 其中 *ABII*、*PP2CA*、*HAB1*、*HAB2* 已被证实对 ABA 信号转导途径具有负调控作用^[17]。黄花蒿中, AaPP2C1 与 ABA 受体蛋白 AaPYLs 相互作用, 在拟南芥中过表达 *AaPP2C1* 后, ABA 响应基因在 ABA 处理后表达下调, 从而降低植株对于 ABA 的敏感性, 促进种子发芽生根^[18]。多个 PP2C 基因家族 B 亚族成员负调

节丝裂原活化蛋白激酶(MAPK)途径,而蛋白质磷酸化和去磷酸化几乎涉及植物所有的生理过程^[19]。MAPK信号途径促进表皮扁平细胞的分化并抑制气孔细胞的形成,这一空间上特殊的细胞发育过程是通过*AP2C3*基因特异性地负调节*MPK3*,*MPK4*和*MPK6*基因的表达实现的^[20]。在苜蓿中,机械损伤可以激活盐胁迫诱导的MAPK(SIMK)蛋白,而PP2C基因家族B亚族的MP2C通过去磷酸化SIMK使MAPK途径失活,负调控外界损伤响应^[21]。烟草毛状根中过表达*NtMPK4*或叶片瞬时表达*NtMPK4*基因可以提高尼古丁的生物合成,而*NtPP2C2b*通过去磷酸化活化*NtMPK4*而减少尼古丁的积累^[22],表明PP2C可能参与植物次生代谢产物合成的调控。

目前,许多园艺植物和药用植物的PP2C基因家族均已进行了鉴定和分析,其中拟南芥和水稻分别含有76个^[23]和78个PP2C基因家族成员^[24],谷子80个^[25],毛竹125个^[26]、甘蓝128个^[27]、蒺藜苜蓿94个^[28]、丹参83个^[29]等。大麻基因组测序已完成,但尚未有关于大麻PP2C基因家族的研究报道。本研究基于工业大麻基因组和转录组数据,通过生物信息学技术对CsPP2C基因家族进行鉴定和分析,以期后续研究CsPP2C基因功能提供参考依据,为工业大麻分子育种提供研究思路。

1 材料

本项目使用的工业大麻CRBRx基因组数据及其他品种Cherry Chem、Canna Tsu、White Cookies、Black Berry Kush、Black Lime、Valley Fire、Sour Diesel、Mama Thai和Terple腺毛转录组数据均来源于美国国家生物信息中心(NCBI, <https://www.ncbi.nlm.nih.gov>),拟南芥基因组和AtPP2C基因家族数据来源于TAIR数据库(<https://www.arabidopsis.org>)。工业大麻Dinamed Kush(Diku)的叶、花、苞片和种子转录组数据由本课题组测序获得。

2 方法

2.1 CsPP2Cs基因家族成员鉴定及染色体定位

使用TBtools软件^[30]提取工业大麻所有蛋白序列与AtPP2C成员文件进行双向Blast,并结合结构域获取可能的CsPP2C基因家族成员,去除重复转录本后获取CsPP2C蛋白序列。利用ExPASy(<https://web.expasy.org/protparam>)网站对CsPP2C蛋白理化性质、分子量、等电点和氨基酸数目进行预测。采用WoLF PSORT(<https://wolfsort.hgc.jp>)网站,根据结果分值预测CsPP2C蛋白亚细胞定位。

2.2 进化树系统分类 将CsPP2C蛋白序列和AtPP2C蛋白序列共同导入MEGA X软件中进行多序列比对,利用邻接法(NJ)构建系统进化树(Bootstrap值1 000),以AtPP2C为分类依据对CsPP2C成员进行分类。

2.3 基因结构、保守基序与系统进化分析 将CsPP2C与AtPP2C蛋白序列共同导入MEGA-X软件中进行多序列比对,利用邻接法构建系统进化树(Bootstrap值1 000)。运用MEME(<https://meme-suite.org/meme/tools/meme>)网站预测CsPP2C蛋白保守基序(Motif),保守基序参数设为10,其他值默认。利用NCBI的Batch-CD-Search(<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)对CsPP2C蛋白结构域进行预测。最后,将CsPP2C进化树文本、全基因组注释文件、CsPP2C保守基序、CsPP2C保守结构域及重命名文件在TBtools软件中进行可视化作图。

2.4 共线性分析 从NCBI网站上下载拟南芥、大麻的全基因组数据,在TBtools软件中导入二者的基因组序列文件,同时导入CsPP2C的ID,对二者的共线性关系进行可视化。

2.5 顺式作用元件预测 通过TBtools软件获取CsPP2C基因上游2 kbp序列文件,利用PlantCare(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>)网站预测启动子区的顺式作用元件,在TBtools软件中将其可视化。

2.6 表达模式分析 将已报道的9个大麻栽培品种(Cherry Chem, Canna Tsu, White Cookies, Black Berry Kush, Black Lime, Valley Fire, Sour Diesel, Mama Thai和Terple)腺毛转录组以及工业大麻Diku品种的叶、花、苞片和种子的转录组数据中CsPP2C基因的表达量数据导入TBtools软件中绘制热图,对该基因在不同组织中差异性表达和基因在不同品种的表达进行聚类分析。

2.7 实时荧光定量聚合酶链式反应(Real-time PCR)分析 根据转录组数据的差异性表达,对*CsPP2C31*、*CsPP2C33*、*CsPP2C41*进行Real-time PCR验证分析。利用WARYONG试剂盒提取工业大麻Diku品种花、苞片、叶和种子的RNA,后用TransGen试剂盒反转录获取cDNA。利用NCBI设计引物,用FOREVER STAR试剂盒进行Real-time PCR分析,反应程序为95℃预变性3 min,95℃变性30 s,56℃退火20 s,72℃延伸15 s,循环40次。*CsPP2C31*上游5'-GAAATTGAGGACGACGCCG-

3', 下游 5'-GCAAATGTTTCGGCGGTGAT-3'; *CsPP2C33* 上游 5'-AGAACCAGAACTTGCTCCGG-3', 下游 5'-GTCTTCCATCTCCTCTCGCG-3'; *CsPP2C41* 上游 5'-GGCGGTTCTTGTGTGTGAC-3', 下游 5'-TGTTAGAGCCTCAGCAGCAC-3'; 内参基因 *EFla* 上游 5'-ACCAAGATTGACAGGCGTTC-3', 下游 5'-CCTTCTTCTCCACAGCCTTG-3'。

2.8 可变剪切分析 利用 IGV 软件对工业大麻可变剪切文件和工业大麻基因组文件进行处理, 获取大麻部分基因家族成员可变剪切图谱。将 Diku 品种的叶、花、苞片和种子的转录组数据中部分 *CsPP2C* 成员的表达量数据导入 TBtools 软件中绘制热图。

3 结果与分析

3.1 *CsPP2C* 基因家族成员鉴定和进化树分析

用 TBtools 软件将工业大麻基因组蛋白序列与 AtPP2C 蛋白序列进行双向 Blast 比对, 获得 52 个 *CsPP2C* 成员。*CsPP2C* 数目明显少于拟南芥、水稻和丹参等。生物信息学分析发现, *CsPP2C* 的蛋白氨基酸数目在 244~1 089; 相对分子质量 26.76~122.53 kDa, 相对分子质量相差较大; 等电点 4.74~9.60, 见表 1。利用 WoLF PSORT 对 *CsPP2C* 进行亚细胞定位预测发现, 52 个 *CsPP2C* 可能分布在叶绿体、细胞核、细胞质、线粒体、液泡和质膜。90% 以上成员分布于细胞核、细胞质和叶绿体。利用 TBtools 分析发现, *CsPP2C* 基因不均匀地分布在工业大麻的 10 条染色体上, 其中 Chr.NC_044370.1 染色体上 *PP2Cs* 数目最多, 达 13 个, 且该染色体上含有明显的基因簇, 如 *CsPP2C10-13*, 而 Chr.NC_044372.1 仅含有 1 个 *CsPP2C23*。见图 1。

表 1 工业大麻 *PP2C* 基因家族基本信息及特征

Table 1 Information and characteristics of *CsPP2C*

基因号	转录本登录号	名称	染色体	大小/aa	等电点	相对分子质量/Da	亚细胞定位
LOC115696549	rna-XM_030623445.1	<i>CsPP2C1</i>	NC_044378.1	499	5.35	55 717.73	细胞核
LOC115696847	rna-XM_030623732.1	<i>CsPP2C2</i>	NC_044378.1	624	5.59	70 158.13	细胞核
LOC115698277	rna-XM_030625488.1	<i>CsPP2C3</i>	NC_044378.1	658	5.52	72 660.58	细胞质
LOC115699256	rna-XM_030626601.1	<i>CsPP2C4</i>	NC_044370.1	244	6.59	26 762.13	叶绿体
LOC115699343	rna-XM_030626696.1	<i>CsPP2C5</i>	NC_044379.1	471	5.48	52 144.63	叶绿体
LOC115699509	rna-XM_030626965.1	<i>CsPP2C6</i>	NC_044370.1	438	5.48	47 590.23	叶绿体
LOC115699513	rna-XM_030626970.1	<i>CsPP2C7</i>	NC_044379.1	283	7.80	30 913.34	细胞质
LOC115700505	rna-XM_030628049.1	<i>CsPP2C8</i>	NC_044379.1	385	8.90	42 672.04	细胞质
LOC115701251	rna-XM_030628999.1	<i>CsPP2C9</i>	NC_044379.1	376	6.77	41 948.55	叶绿体
LOC115702601	rna-XM_030630013.1	<i>CsPP2C10</i>	NC_044370.1	381	5.81	41 658.01	质膜
LOC115702862	rna-XM_030630309.1	<i>CsPP2C11</i>	NC_044370.1	387	5.18	42 360.82	细胞质
LOC115702900	rna-XM_030630358.1	<i>CsPP2C12</i>	NC_044370.1	397	8.78	44 093.95	细胞质
LOC115703037	rna-XM_030630519.1	<i>CsPP2C13</i>	NC_044370.1	377	8.66	42 099.88	叶绿体
LOC115704787	rna-XM_030631996.1	<i>CsPP2C14</i>	NC_044371.1	373	6.98	40 208.62	细胞核
LOC115705123	rna-XM_030632362.1	<i>CsPP2C15</i>	NC_044371.1	436	5.46	47 584.81	细胞质
LOC115705630	rna-XM_030633026.1	<i>CsPP2C16</i>	NC_044370.1	351	5.45	38 657.54	细胞核
LOC115706627	rna-XM_030634341.1	<i>CsPP2C17</i>	NC_044370.1	542	4.74	58 969.51	叶绿体
LOC115707022	rna-XM_030634825.1	<i>CsPP2C18</i>	NC_044371.1	541	4.74	57 955.86	线粒体
LOC115707309	rna-XM_030635226.1	<i>CsPP2C19</i>	NC_044370.1	320	7.22	35 319.18	细胞核
LOC115707407	rna-XM_030635364.1	<i>CsPP2C20</i>	NC_044371.1	1 073	4.98	119 330.50	细胞核
LOC115707545	rna-XM_030635555.1	<i>CsPP2C21</i>	NC_044370.1	289	4.85	31 269.59	细胞质
LOC115707957	rna-XM_030636077.1	<i>CsPP2C22</i>	NC_044371.1	350	5.21	38 417.25	叶绿体
LOC115709373	rna-XM_030637460.1	<i>CsPP2C23</i>	NC_044372.1	395	8.45	43 902.83	细胞质
LOC115712168	rna-XM_030640419.1	<i>CsPP2C24</i>	NC_044373.1	427	8.43	45 739.24	叶绿体
LOC115712666	rna-XM_030640986.1	<i>CsPP2C25</i>	NC_044373.1	395	8.21	44 084.20	细胞核

续表 1

基因号	转录本登录号	名称	染色体	大小/aa	等电点	相对分子质量/Da	亚细胞定位
LOC115714536	rna-XM_030643269.1	<i>CsPP2C26</i>	NC_044373.1	700	6.27	78 029.34	叶绿体
LOC115714587	rna-XM_030643333.1	<i>CsPP2C27</i>	NC_044370.1	361	5.14	39 583.36	叶绿体
LOC115714713	rna-XM_030643471.1	<i>CsPP2C28</i>	NC_044373.1	380	6.24	42 456.26	叶绿体
LOC115715787	rna-XM_030644459.1	<i>CsPP2C29</i>	NC_044374.1	400	5.35	44 671.36	叶绿体
LOC115715848	rna-XM_030644518.1	<i>CsPP2C30</i>	NC_044374.1	797	5.25	88 003.05	叶绿体
LOC115716087	rna-XM_030644798.1	<i>CsPP2C31</i>	NC_044374.1	394	5.21	43 373.76	细胞核
LOC115716196	rna-XM_030644937.1	<i>CsPP2C32</i>	NC_044374.1	357	5.48	39 519.43	叶绿体
LOC115716767	rna-XM_030645665.1	<i>CsPP2C33</i>	NC_044374.1	393	5.09	43 102.67	叶绿体
LOC115717232	rna-XM_030646204.1	<i>CsPP2C34</i>	NC_044374.1	376	8.76	41 724.28	细胞核
LOC115718843	rna-XM_030647653.1	<i>CsPP2C35</i>	NC_044375.1	425	6.40	45 675.40	细胞核
LOC115718894	rna-XM_030647702.1	<i>CsPP2C36</i>	NC_044375.1	719	5.26	80 394.64	叶绿体
LOC115718972	rna-XM_030647806.1	<i>CsPP2C37</i>	NC_044375.1	478	5.72	52 577.34	细胞核
LOC115719402	rna-XM_030648473.1	<i>CsPP2C38</i>	NC_044370.1	1 089	6.10	122 534.50	叶绿体
LOC115719439	rna-XM_030648491.1	<i>CsPP2C39</i>	NC_044375.1	413	5.58	44 974.44	细胞核
LOC115719734	rna-XM_030648908.1	<i>CsPP2C40</i>	NC_044375.1	369	6.02	41 119.36	细胞质
LOC115720616	rna-XM_030649778.1	<i>CsPP2C41</i>	NC_044375.1	382	6.67	41 153.43	叶绿体
LOC115722308	rna-XM_030651486.1	<i>CsPP2C42</i>	NC_044376.1	285	9.60	32 374.46	细胞核
LOC115722596	rna-XM_030651838.1	<i>CsPP2C43</i>	NC_044376.1	474	8.73	51 397.42	叶绿体
LOC115722610	rna-XM_030651855.1	<i>CsPP2C44</i>	NC_044376.1	513	4.80	56 391.23	细胞核
LOC115722697	rna-XM_030651971.1	<i>CsPP2C45</i>	NC_044376.1	424	4.87	46 516.58	细胞核
LOC115723145	rna-XM_030652580.1	<i>CsPP2C46</i>	NC_044376.1	576	5.67	64 195.36	细胞核
LOC115723385	rna-XM_030652857.1	<i>CsPP2C47</i>	NC_044376.1	383	6.28	42 497.09	叶绿体
LOC115723420	rna-XM_030652906.1	<i>CsPP2C48</i>	NC_044376.1	387	4.85	42 562.11	细胞质
LOC115724355	rna-XM_030653704.1	<i>CsPP2C49</i>	NC_044370.1	405	5.43	44 288.30	液泡
LOC115724774	rna-XM_030654121.1	<i>CsPP2C50</i>	NC_044377.1	384	4.83	42 452.87	细胞质
LOC115724962	rna-XM_030654348.1	<i>CsPP2C51</i>	NC_044377.1	283	7.73	30 915.75	叶绿体
LOC115725091	rna-XM_030654516.1	<i>CsPP2C52</i>	NC_044377.1	477	5.03	52 963.46	线粒体

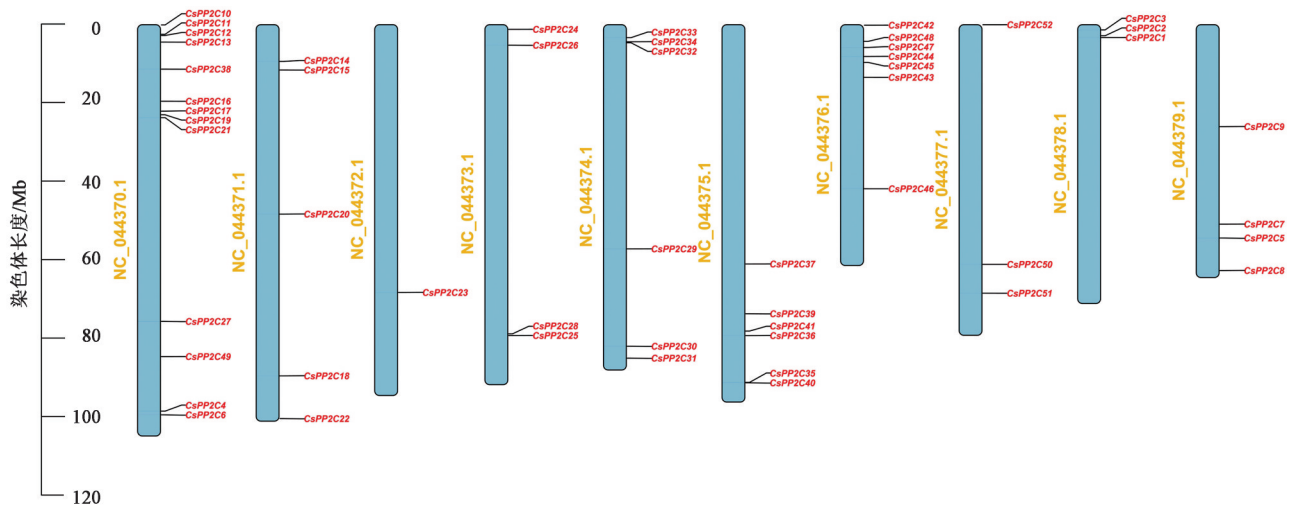
PP2C 因其结构的特殊性, N 端比 C 端保守性差, N 端延伸区的差异性决定了成员之间的功能差异性^[23,31], 为了进一步研究 CsPP2C 基因家族成员进化特性, 将 AtPP2C 与 CsPP2C 共同构建系统发育树, 见图 2。根据进化树的分类结果, 可将 CsPP2C 分为 10 个亚家族。其各亚家族成员分布情况为 A 亚族 4 个、B 亚族 4 个、C 亚族 4 个、D 亚族 7 个、E 亚族 11 个、F 亚族 6 个、G 亚族 5 个、H 亚族 3 个、I 亚族 2 个、J 亚族 1 个, 剩余 5 个成员未聚类, 分为其他类。

3.2 CsPP2C 基因家族保守基序、基因结构和顺式作用元件分析 利用 MEME 网站和 TBtools 分别对 CsPP2C 的保守基序和基因结构进行了分析。保守基序分析发现, 保守基序 1、2、3 分布于 51 个工业大麻 PP2C 基因家族成员, 除 CsPP2C4 不含保守基序

1, CsPP2C46 不含保守基序 2, CsPP2C42 不含保守基序 3 外, 其他基因家族成员均含有相应的结构域。保守基序 4、6、8 存在于 48 个 CsPP2C 成员中, 保守基序 5、9、10 分布少数基因家族成员。除此之外, 同一亚族内成员的保守基序组成几乎相同, 进一步表明进化树分类的可靠性, 而不同亚族中的保守基序组成具有一定差异。见图 3、表 2。

基因结构分析发现, *CsPP2C* 普遍含有外显子和 2 个以上的内含子。其中 A、C、D 和 G 亚族基因的外显子数量不超过 4 个。十分有趣的是, 5 个未聚类 PP2C 基因或含有 10 个以上的外显子 (*CsPP2C46*、*CsPP2C3*、*CsPP2C33*), 或仅含有 3 个外显子 (*CsPP2C34*、*CsPP2C31*), 见图 4。

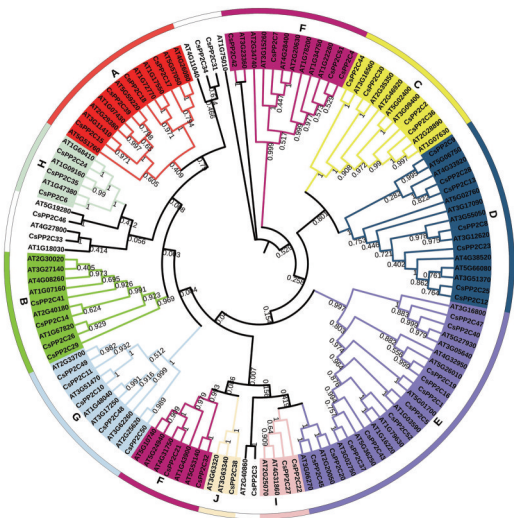
基因启动子对基因的表达和调控十分关键, 因



注:染色体的NC号以黄色字体标注;At.拟南芥;Cs.工业大麻

图1 CsPP2C基因染色体分布

Fig. 1 Chromosomal distribution of CsPP2C



A~J.分类(图3同)

图2 PP2C基因家族系统发育树

Fig. 2 Phylogenetic tree of PP2C family

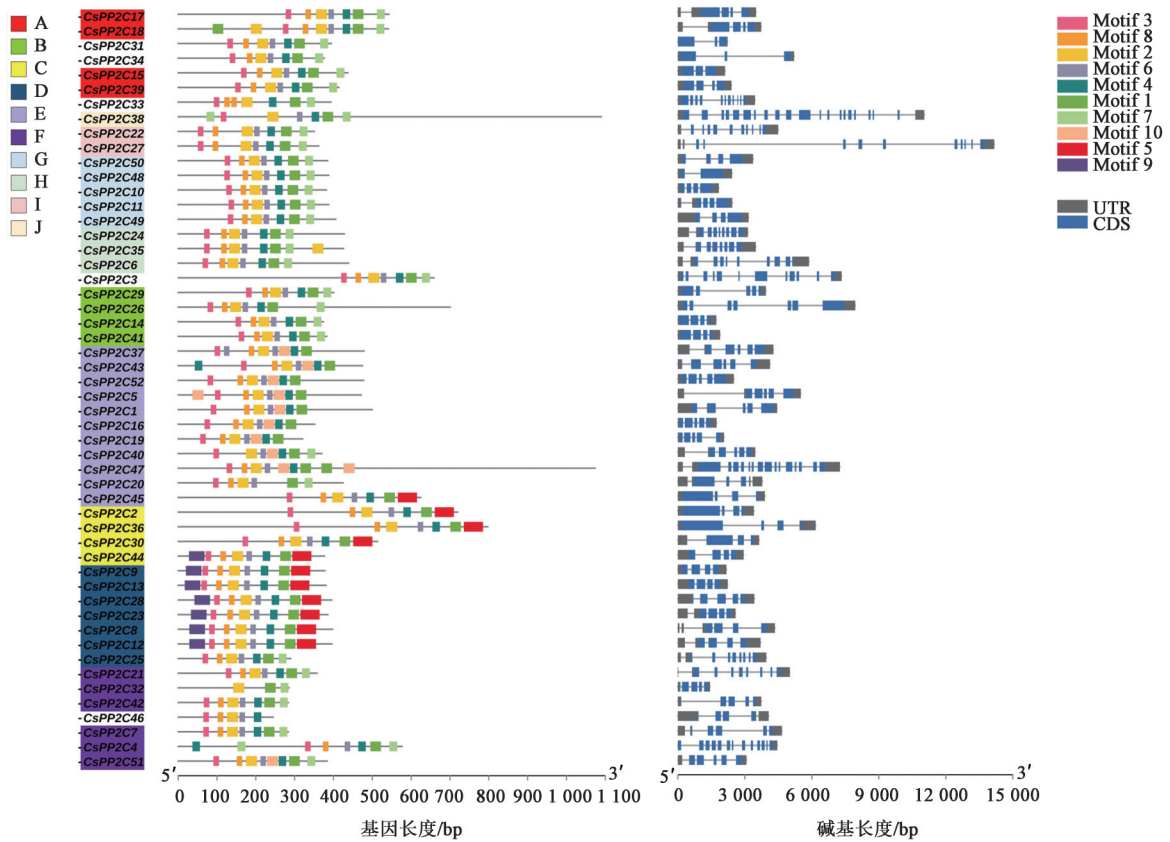
此,利用 PlantCare 对 CsPP2C 基因家族成员的启动子的顺式作用元件进行预测。共计发现 20 个顺式作用元件,其中主要顺式作用元件包括光响应元件、脱落酸响应元件、无氧诱导元件和茉莉酸甲酯响应元件,在工业大麻基因家族中的分布数量依次为 290、163、118、104 个。数量最少的是光敏色素下调表达作用元件,仅有 1 个,位于 CsPP2C24 启动子区。

3.3 工业大麻 PP2C 基因家族表达模式分析及 Real-time PCR 验证 利用工业大麻 Diku 品种的叶、花、苞片、种子的转录组测序数据和已发表的不同大麻品种的腺毛转录组数据^[32],绘制 CsPP2C 在不同品种和不同组织的表达热图,见图 5。通过聚

类分析发现, CsPP2C 在不同组织中的表达有较为明显的差异, CsPP2C33 (未聚类)在叶中高表达, CsPP2C21 (F 亚族)和 CsPP2C31 (未聚类)在种子中高表达, CsPP2C41 (B 亚族)在苞片中高表达。而 CsPP2C7、CsPP2C15、CsPP2C48、CsPP2C51 在 10 个品种中均有较高的表达量, CsPP2C1、CsPP2C2、CsPP2C4、CsPP2C6、CsPP2C9、CsPP2C10、CsPP2C16、CsPP2C17、CsPP2C19、CsPP2C26、CsPP2C31、CsPP2C38、CsPP2C40、CsPP2C42、CsPP2C47、CsPP2C49 表达量较低。相同基因在同一植物、不同品种中表达量存在差异,间接表明品种之间的表型差异与基因表达存在一定联系。

为了进一步验证转录组数据的可靠性,利用 Real-time PCR 对 CsPP2C31、CsPP2C33、CsPP2C41 在工业大麻 Diku 不同组织部位的表达情况进行了验证。Real-time PCR 结果显示, CsPP2C31、CsPP2C33、CsPP2C41 分别在种子,叶和苞片中高表达,与转录组结果一致,见表 3。

3.4 CsPP2C 基因家族共线性和可变剪切分析 拟南芥作为模式植物,其 PP2C 基因家族成员功能研究较为全面。因此,利用 TBtools 对 CsPP2C 和 AtPP2C 基因家族成员进行了共线性分析,共计发现 7 对同源基因,分别 CsPP2C12/AT4G38520、CsPP2C32/AT5G53140、CsPP2C40/AT3G05640、CsPP2C48/AT3G17250、CsPP2C42/AT3G23360、CsPP2C46/AT5G19280、CsPP2C1/AT5G01700。这与系统进化树的分类结果一致。见图 6。



注: Motif. 保守基序; UTR. 非编码区; CDS. 编码区

图3 CsPP2C保守基序、结构和顺式作用元件

Fig. 3 Conserved motifs, gene structure and *cis*-elements analyses of CsPP2C

表2 CsPP2C基因家族保守基序

Table 2 Motifs of PP2C gene family in *Cannabis sativa*

保守基序	长度/bp	氨基酸保守序列
Motif 1	28	LTPDDEFILASDGLWDVLSNZEAVDIV
Motif 2	29	SGSTALVALIQEKLVVANVGDSTRVAVLGR
Motif 3	15	SFFGVFDGHGGPGAA
Motif 4	20	LAVSRAIGDRYLKQYVVSEP
Motif 5	50	PRGGPAKRLVKAALQEAAKKREMRYS DLKKIDRGVRRHFHDDITVIVIFL
Motif 6	15	AVQLSVDHKPSREDE
Motif 7	21	LVEEALARGSKDNITVIVVDL
Motif 8	15	EAJKKAFLKTDEEFL
Motif 9	41	GRQDGLLWYKDLGQHASGDFSMVAVQ ANSLLEDQSQVESGP
Motif 10	29	RIKRCKGRVFALKDEPEVARVWLPNEDSP

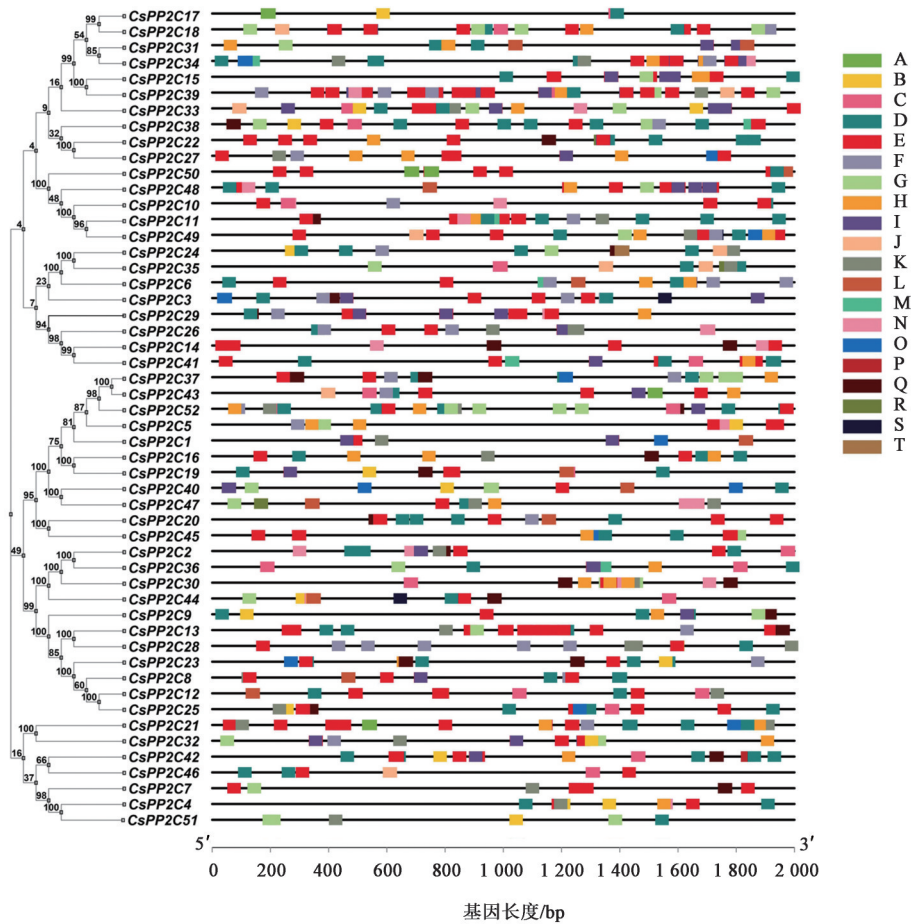
可变剪切是前体 mRNA 通过不同的剪切方式形成多种成熟 mRNA 的过程,是调控基因表达和产生蛋白质多样性的重要机制。可变剪切对植物的生长发育、信号转导、生物胁迫和非生物胁迫有重要调控作用。为此,对 CsPP2C 基因家族中 CsPP2C31、CsPP2C33、CsPP2C41 3 个基因进行可

变剪切分析。分析结果表明 CsPP2C31、CsPP2C33 存在 1 个可变剪切本, CsPP2C41 存在 4 个可变剪切本,表明该基因在进化的过程中发生了可变剪切事件,见图 6。不同的可变剪切本在不同的组织和部位可能受到不同的调控,与热图表达结果一致。

3 讨论

PP2C 蛋白在进化上十分保守,细菌、真菌、植物和动物体内均存在大量的 PP2C^[33]。植物中的已知的 PP2C 主要参与胁迫信号转导,但大多数的 PP2C 功能仍然未知。本文利用生物信息学技术,通过 Blast 比对序列、构建系统发育树、基因表达分析等方法,从工业大麻全基因组中获取了 52 个 CsPP2C 成员。并对其亚细胞定位、理化性质、基因结构和表达情况等进行了分析与预测。其中,90%CsPP2C 分布于细胞核、细胞质、叶绿体,为后续基因互作研究提供一定的参考依据。CsPP2C 基因不均匀分布在工业大麻的 10 条染色体上,大部分染色体存在基因簇,表明该基因在进化的过程中发生了复制事件,同一基因簇的成员可能含有相类似的功能。

进化树分析将 CsPP2C 分为 10 个亚家族和 5 个



注: A~T. 从上到下依次为黄酮类合成基因调控元件、分生组织表达调控元件、水杨酸响应元件、厌氧呼吸响应元件、光响应元件、低温响应元件、赤霉素响应元件、茉莉酸甲酯响应元件、脱落酸响应元件、生物钟响应元件、干旱诱导响应元件、玉米醇溶蛋白响应元件、细胞周期调控、生长素响应元件、胚乳表达、损伤诱导、防御和胁迫响应、种子特异性调控、缺氧特异性诱导、光敏色素表达负调控

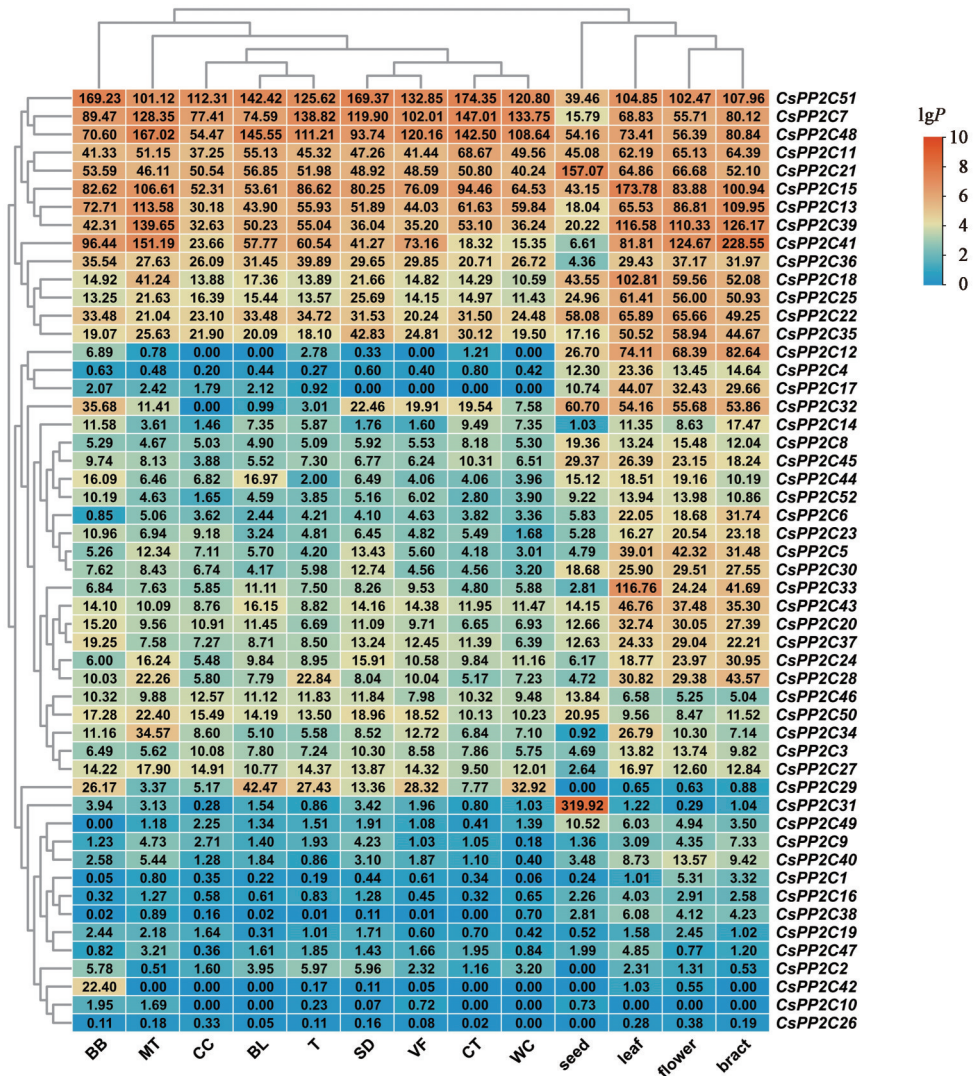
图4 CsPP2C 顺式作用元件

Fig. 4 Conserved cis-elements analyses of CsPP2C

未分类成员,与拟南芥分类结果一致,表明该基因家族在进化过程中较为保守。拟南芥中,PP2C A亚族主要参与ABA信号响应,而ABA广泛调控植物的蒸腾作用、营养生殖和种子萌发^[17,34]。CsPP2C A亚族的CsPP2C15启动子区域含有ABA响应元件,定位于细胞质。这与处于同一进化分支上的拟南芥AT3G11410非常相似,AT3G11410在拟南芥种子萌发时调控ABA信号通路^[35]。PP2C B亚族基因成员与MAPK基因家族成员有显著调控关系^[36]。由于植物不同胁迫响应共享MAPK信号通路的调控,因此,推测PP2C可能是MAPK信号转导的特异性响应的分子基础。CsPP2C B亚族含有4个成员,分别是CsPP2C41、CsPP2C14、CsPP2C26和CsPP2C19。其中,CsPP2C41定位在叶绿体,在进化中与At2G30020(*AtAP2C1*)更相近。At2G30020在拟南芥中负调控胁迫引起的MAPK信号转导。植物受伤后,At2G30020突变体中的茉莉酸含量升高^[37],过

表达株系中,损伤引起的MAPK信号降低,而CsPP2C41的启动子中具有茉莉酸甲酯响应元件,提示CsPP2C41与At2G30020可能有相似的功能。

保守基序与基因功能密切相关,保守基序9为CsPP2C D亚族所特有,AtPP2C基因家族D亚族具有明显的核定位和质膜定位信号^[23],与细胞发育相关。拟南芥PP2C基因家族的MIDA蛋白调节幼苗脱黄化,其中,MIDA11参与下胚轴伸长,MIDA9和MIDA10调节顶钩的发育,MIDA1调控子叶分离^[38],其他D亚族成员对植物幼苗的萌发及生长均有不同程度的调控作用^[39],推测CsPP2C基因家族D亚族成员可能具有相应的功能。工业大麻与拟南芥共线性分析共获取了7对同源基因,其中,蛋白质定位在细胞质E亚族基因CsPP2C40与AT3G05640(*AtPP2CF1*)具有共线性。拟南芥中外源ABA处理可显著提高AT3G05640基因在维管组织和保卫细胞的表达^[40]。另外,AT3G05640基因过表达株系表



注: CsPP2C 基因在 BB(Black Berry Kush)、MT(Mama Thai)、CC(Cherry Chem)、BL(Black Lime)、T(Terple)、S(Sour Diesel)、VF(Valley Fire)、CT(Canna Tsu)、WC(White Cookies)品种腺毛的表达;种子(seed)、花(flower)、叶(leaf)和苞片(bract)为 CsPP2Cs 在 Diku 品种不同组织的表达

图5 CsPP2C 基因在不同品种和组织的表达热图

Fig. 5 Heat map of CsPP2C expression in different varieties and tissue.

表3 CsPP2C 家族 Real-time PCR 检测 ($\bar{x} \pm s, n=3$)

Table 3 Real-time PCR of CsPP2C family ($\bar{x} \pm s, n=3$)

基因名称	花	苞片	叶	种子
CsPP2C31	1.06±0.50	10.69±0.63	13.96±2.34	82.20±8.00
CsPP2C33	1.65±1.30	13.49±0.65	74.43±3.75	1.01±0.21
CsPP2C41	16.54±2.14	55.89±3.07	20.69±4.43	1.01±0.20

注:内参基因为 EFla

现出对 ABA 处理的低敏感表型^[40]。CsPP2C48 基因共线性基因 AT3G17250 在拟南芥胚乳中表达,其负调控蛋白激酶途径、脂类信号转导途径和茉莉酸生物合成途径^[41]。因此,CsPP2C40 和 CsPP2C48 基因在大麻中可能具有类似的功能。

通过 Real-time PCR 对部分差异性表达的基因

进行分析发现,CsPP2C33 基因在叶中高表达。CsPP2C33 与 AT4G27800 在进化中处于同一分支,AT4G27800.1 基因为拟南芥 AT4G27800 基因的 3 种剪切形式中的一种,在叶片中表达,与植物的光合作用密切相关^[42];种子高表达基因 CsPP2C31 与 AT4G11040 基因同源,AT4G11040 在种子吸胀期抑制去磷酸化,影响种子的休眠和萌发^[43]。因此,CsPP2C33 与 CsPP2C31 基因在大麻中极有可能具有类似的功能。

本研究通过生物信息学技术和方法对 CsPP2C 基因家族成员进行了系统且全面的分析,为工业大麻 PP2C 基因家族的功能研究和大麻的分子育种品种选育提供了参考依据。

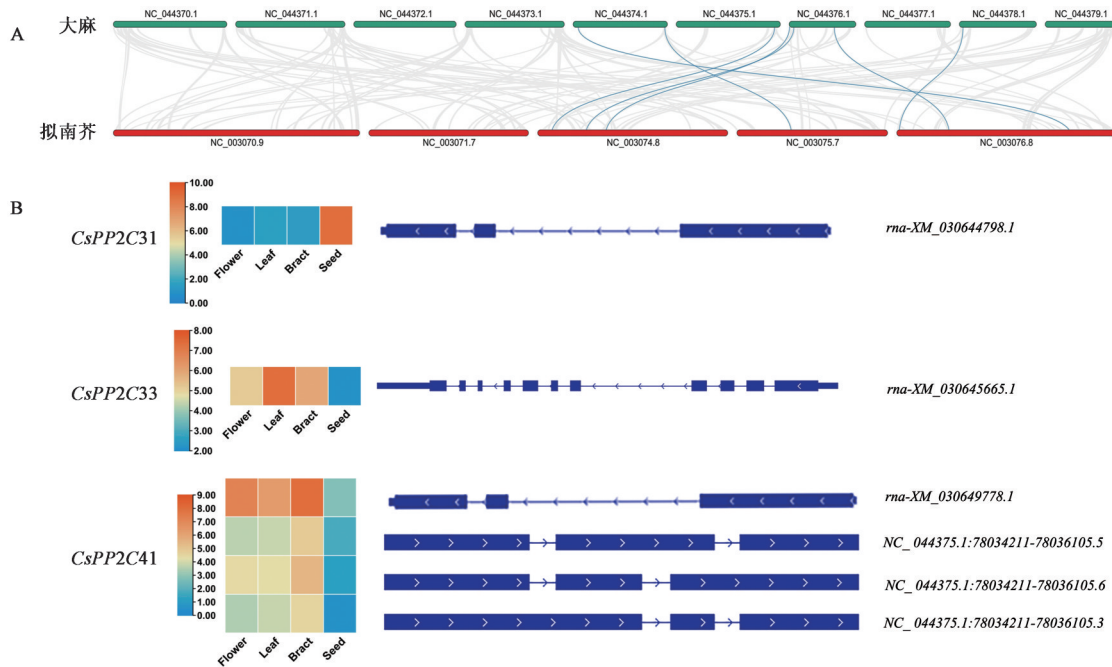


图6 大麻与拟南芥 *PP2C* 基因共线性分析(A)和 *CsPP2C* 基因可变剪切分析(B)

Fig. 6 Collinearity between *CsPP2C* and *AtPP2C* and Alternative splicing analysis of *CsPP2C*

[利益冲突] 本文不存在任何利益冲突。

[参考文献]

- [1] 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.
- [2] AMIN M R, ALI D W. Pharmacology of medical cannabis[J]. Adv Exp Med Biol, 2019, 1162: 151-165.
- [3] 贺海波,石孟琼. 火麻仁的化学成分和药理活性研究进展[J]. 中国民族民间医药, 2010, 19(15): 56-57.
- [4] 韦凤,涂冬萍,王柳萍. 火麻仁食用开发和药理作用研究进展[J]. 中国老年学杂志, 2015, 35(12): 3486-3488.
- [5] 弓佩含,杨洋,刘玉婷,等. 大麻化学成分及药理作用的研究进展[J]. 中国实验方剂学杂志, 2017, 23(13): 212-219.
- [6] DARIŠ B, TANCER VERBOTEN M, KNEZ Ž, et al. Cannabinoids in cancer treatment: Therapeutic potential and legislation[J]. Bosn J Basic Med Sci, 2019, 19(1): 14-23.
- [7] HILL K P, PALASTRO M D, JOHNSON B, et al. Cannabis and pain: A clinical review [J]. Cannabis Cannabinoid Res, 2017, 2(1): 96-104.
- [8] 周永凯,张建春,张华. 大麻纤维的抗菌性及抗菌机制[J]. 纺织学报, 2007, 28(6): 12-15.
- [9] 陈建华,臧巩固,赵立宁,等. 大麻化学成分研究进展与开发我国大麻资源的探讨[J]. 中国麻业, 2003, 25(6): 266-271.
- [10] 孙小寅,管映亭,温桂清,等. 大麻纤维的性能及其应用研究[J]. 纺织学报, 2001, 22(4): 34-36.
- [11] JOHNSON R. Defining hemp: A fact sheet [J]. Congressional Res Service, 2019, 44742.
- [12] REN G, ZHANG X, LI Y, et al. Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*[J]. Sci Adv, 2021, 7(29): eabg2286.
- [13] COHEN P. The structure and regulation of protein phosphatases[J]. Annu Rev Biochem, 1989, 58(1): 453-508.
- [14] HUBBARD M J, COHEN P. On target with a new mechanism for the regulation of protein phosphorylation [J]. Trends Biochem Sci, 1993, 18(5): 172-177.
- [15] BORK P, BROWN N P, HEGYI H, et al. The protein phosphatase 2C (PP2C) superfamily: Detection of bacterial homologues [J]. Protein Sci, 1996, 5(7): 1421-1425.
- [16] FUCHS S, GRILL E, MESKIENE I, et al. Type 2C protein phosphatases in plants[J]. FEBS J, 2013, 280(2): 681-693.
- [17] UMEZAWA T, SUGIYAMA N, MIZOGUCHI M, et al. Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in *Arabidopsis* [J]. Proc Natl Acad Sci U S A, 2009, 106(41): 17588-17593.
- [18] ZHANG F, FU X, LV Z, et al. Type 2C phosphatase 1 of *Artemisia annua* L. is a negative regulator of ABA signaling[J]. Bio Med Res Int, 2014, 2014: 521794.

- [19] RODRIGUEZ M C, PETERSEN M, MUNDY J. Mitogen-activated protein kinase signaling in plants [J]. *Annu Rev Plant Biol*, 2010, 61(1): 621-649.
- [20] UMBRASAITE J, SCHWEIGHOFER A, KAZANAVICIUTE V, et al. MAPK phosphatase AP2C3 induces ectopic proliferation of epidermal cells leading to stomata development in *Arabidopsis* [J]. *PLoS One*, 2010, 5(12): e15357.
- [21] MESKIENE I, BAUDOUIN E, SCHWEIGHOFER A, et al. Stress-induced protein phosphatase 2C is a negative regulator of a mitogen-activated protein kinase [J]. *J Biol Chem*, 2003, 278(21): 18945-18952.
- [22] LIU X, SINGH S K, PATRA B, et al. Protein phosphatase NtPP2C2b and MAP kinase NtMPK4 act in concert to modulate nicotine biosynthesis [J]. *J Exp Bot*, 2021, 72(5): 1661-1676.
- [23] SCHWEIGHOFER A, HIRT H, MESKIENE I. Plant PP2C phosphatases: Emerging functions in stress signaling [J]. *Trends Plant Sci*, 2004, 9(5): 236-243.
- [24] XUE T, WANG D, ZHANG S, et al. Genome-wide and expression analysis of protein phosphatase 2C in rice and *Arabidopsis* [J]. *BMC Genomics*, 2008, 9: 550.
- [25] 闵东红, 薛飞洋, 马亚男, 等. 谷子 PP2C 基因家族的特征 [J]. *作物学报*, 2013, 39(12): 2135-2144.
- [26] 胡秋涛, 侯丹, 赵钟毓, 等. 毛竹 PP2C 基因家族鉴定与表达分析 [J]. *农业生物技术学报*, 2020, 28(10): 1776-1787.
- [27] 王丹丹. 白菜、甘蓝、拟南芥 PP2C 型蛋白磷酸酶基因的比较分析与功能研究 [D]. 杨凌: 西北农林科技大学, 2017.
- [28] YANG Q, LIU K, NIU X, et al. Genome-wide identification of PP2C genes and their expression profiling in response to drought and cold stresses in *medicago truncatula* [J]. *Sci Rep*, 2018, 8(1): 12841.
- [29] 徐志超, 浦香东, 宋经元. 基于丹参基因组的蛋白磷酸酶 2C 家族的系统分析 [J]. *中国现代中药*, 2018, 20(6): 652-657.
- [30] 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.
- [31] LU X, XIONG Q, CHENG T, et al. A PP2C-1 Allele underlying a quantitative trait locus enhances soybean 100-seed weight [J]. *Mol Plant*, 2017, 10(5): 670-684.
- [32] ZAGER J J, LANGE I, SRIVIDYA N, et al. Gene networks underlying cannabinoid and terpenoid accumulation in cannabis [J]. *Plant Physiol*, 2019, 180(4): 1877-1897.
- [33] RODRIGUEZ P L. Protein phosphatase 2C (PP2C) function in higher plants [J]. *Plant Mol Biol*, 1998, 38(6): 919-927.
- [34] FINKELSTEIN R R, GAMPALA S S, ROCK C D. Abscisic acid signaling in seeds and seedlings [J]. *Plant Cell*, 2002, 14 Suppl(Suppl): S15-S45.
- [35] YOSHIDA T, NISHIMURA N, KITAHATA N, et al. ABA-hypersensitive germination3 encodes a protein phosphatase 2C (AtPP2CA) that strongly regulates abscisic acid signaling during germination among *arabidopsis* protein phosphatase 2Cs [J]. *Plant Physiol*, 2006, 140(1): 115-126.
- [36] BROCK A K, WILLMANN R, KOLB D, et al. The *Arabidopsis* mitogen-activated protein kinase phosphatase PP2C5 affects seed germination, stomatal aperture, and abscisic acid-inducible gene expression [J]. *Plant Physiol*, 2010, 153(3): 1098-1111.
- [37] SCHWEIGHOFER A, KAZANAVICIUTE V, SCHEIKL E, et al. The PP2C-type phosphatase AP2C1, which negatively regulates MPK4 and MPK6, modulates innate immunity, jasmonic acid, and ethylene levels in *Arabidopsis* [J]. *Plant Cell*, 2007, 19(7): 2213-2224.
- [38] SENTANDREU M, MARTÍN G, GONZÁLEZ-SCHAIN N, et al. Functional profiling identifies genes involved in organ-specific branches of the PIF3 regulatory network in *Arabidopsis* [J]. *Plant Cell*, 2011, 23(11): 3974-3991.
- [39] ROVIRA A, SENTANDREU M, NAGATANI A, et al. The sequential action of MIDA9/PP2C. D1, PP2C. D2, and PP2C. D5 is necessary to form and maintain the hook after germination in the dark [J]. *Front Plant Sci*, 2021, 12: 636098.
- [40] SUGIMOTO H, KONDO S, TANAKA T, et al. Overexpression of a novel *Arabidopsis* PP2C isoform, AtPP2CF1, enhances plant biomass production by increasing inflorescence stem growth [J]. *J Exp Bot*, 2014, 65(18): 5385-5400.
- [41] HSIEH T F, SHIN J, UZAWA R, et al. Regulation of imprinted gene expression in *Arabidopsis* endosperm [J]. *Proc Natl Acad Sci USA*, 2011, 108(5): 1755-1762.
- [42] PRIBIL M, PESARESI P, HERTLE A, et al. Role of plastid protein phosphatase TAP38 in LHC II dephosphorylation and thylakoid electron flow [J]. *PLoS Biol*, 2010, 8(1): e1000288.
- [43] XIANG Y, SONG B, NÉE G, et al. Sequence polymorphisms at the REDUCED DORMANCY5 pseudophosphatase underlie natural variation in *Arabidopsis* dormancy [J]. *Plant Physiol*, 2016, 171(4): 2659-2670.

[责任编辑 顾雪竹]