

• 药材与资源 •

汉麻聚酮合酶基因家族成员鉴定与表达分析

朱雪雯¹, 米要磊¹, 孟祥霄¹, 陈伟强¹, 王思凡¹, 曹 雪¹, 陈士林¹, 徐志超², 苏 畅³, 孙 伟^{1*}, 万会花^{1*}

1. 中国中医科学院中药研究所, 中药鉴定与安全性评估重点实验室, 北京 100700

2. 东北林业大学生命科学学院, 黑龙江 哈尔滨 150040

3. 深圳市药品检验研究院, 广东 深圳 518057

摘要:目的 在从全基因组层面对汉麻 *Cannabis sativa* 聚酮合酶 (polyketide synthases, PKSs) 基因家族进行鉴定和功能分析, 为解析汉麻次生代谢产物的合成与积累奠定理论基础。方法 基于已发布的高质量汉麻基因组, 利用生物信息学工具对汉麻 PKS 基因家族进行鉴定并对其理化性质、基因结构、蛋白构象以及表达模式进行分析, 利用 AlphaFold 算法对关键 PKS 蛋白结构进行预测。结果 共鉴定出 9 个汉麻 PKSs (*CsPKS1*~*CsPKS9*), 分布在 5 条染色体上。通过对汉麻和其他物种 PKS 的系统发育分析, *CsPKSs* 主要被分为 3 组: group1 包括 *CsPKS2*~*CsPKS4* 与花药特异性 CHS 样酶 (anther specific chalcone synthase-like, ASCL) 同源; group2 含有 *CsPKS1* 和 *CsPKS5*~*CsPKS8* 为类 CHS 超家族 (chalcone synthase like, CHSL) 与苯戊酮合酶 (valerophenone synthase, VPS) 和芪类合酶 (stilbene synthase, STS) 等关系较密切; group3 仅含有 *CsPKS9*, 为查耳酮合酶 (chalcone synthase, CHS) 参与类黄酮的合成。其中, *CsPKS1* 和 *CsPKS7* 串联重复, 是汉麻中酚萜类化合物合成通路中的橄榄醇合成酶 (olivitol synthases, OLS)。此外, *CsPKS5* 和 *CsPKS8* 可能参与汉麻芪类化合物的生物合成。*CsCHS* 的功能在进化中相对保守, 而 CHS-like 的 *CsPKS* 的功能出现了分化。通过对汉麻不同组织部位和不同品种腺毛的 RNA-seq 数据分析, 汉麻 PKSs 基因的表达存在组织特异性, 尤其与酚萜类、类黄酮、芪类次生代谢产物密切相关的基因在雌花和苞片中特异性表达。结论 为深入研究汉麻 PKS 基因家族的功能提供了方向, 为进一步阐明汉麻中萜酚类化合物的生物合成机制和 *CsPKSs* 在汉麻中的生理功能奠定基础, 并对开发新聚酮化合物药用资源提供理论基础。

关键词: 汉麻; 基因家族; 聚酮合酶; 生信分析; 基因表达; 次生代谢产物

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)03-0886-12

DOI: 10.7501/j.issn.0253-2670.2023.03.023

Identification and expression analysis of Polyketide synthase gene family in *Cannabis sativa*

ZHU Xue-wen¹, MI Yao-lei¹, MENG Xiang-xiao¹, CHEN Wei-qiang¹, WANG Si-fan¹, CAO Xue¹, CHEN Shi-lin¹, XU Zhi-chao², SU Chang³, SUN Wei¹, WAN Hui-hua¹

1. Key Laboratory of Beijing for Identification and Safe Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

2. School of Life Sciences, Northeast Forestry University, Harbin 150040, China

3. Shenzhen Institute of Drug Control, Shenzhen 518057, China

Abstract: Objective To conduct identification and functional analysis of the polyketide synthase (PKS) gene family of Hanma (*Cannabis sativa* L.) from the genome-wide level as a theoretical basis for resolving the synthesis and accumulation of secondary metabolites in *Cannabis sativa* L. **Methods** Based on the published high-quality *C. sativa* genome, the *C. sativa* PKS (*CsPKS*) gene family were identified and its physicochemical properties, gene structure, protein conformation and expression pattern were

收稿日期: 2022-09-09

基金项目: 中国中医科学院科技创新工程 (C12021A03710)

作者简介: 朱雪雯, 硕士研究生。E-mail: 3443429211@qq.com

*通信作者: 万会花, 助理研究员, 研究方向为药用成分合成途径挖掘。E-mail: hhwan@icmm.ac.cn

孙 伟, 研究员, 研究方向为中药材鉴定及药用成分合成途径及调控机制挖掘。E-mail: wsun@icmm.ac.cn

analyzed through bioinformatics analysis tools, in addition, this study also used the Alphafold algorithm to predict the key PKS protein structure. **Results** A total of nine CsPKS members (CsPKS1-CsPKS9) were identified in this study and they distributed on five chromosomes. Through the phylogenetic analysis of CsPKSs in *C. sativa* along with other species, CsPKSs are mainly divided into three groups. Group1 contains CsPKS2-CsPKS4 are homologous to anther specific chalcone synthase (CHS)-like (ASCL) enzymes; Group2 contains CsPKS1 and CsPKS5-CsPKS8, which belongs to the CHS-like superfamily (CHSL) and is closely related to valerophenone synthase (VPS) and stilbene synthase (STS); Group3 contains only CsPKS9, a CHS involved in the synthesis of flavonoids. Among them, *CsPKS1* and *CsPKS7* are tandem repeat genes, which are olivetol synthases (OLS) in the synthesis pathway of phenolic terpenoids in hemp, *CsPKS9* is the CHS of the hemp flavonoid synthesis pathway and is involved in the synthesis of flavonoids. In addition, *CsPKS5* and *CsPKS8* may be involved in the biosynthesis of stilbene compounds. The functions of CsCHS are relatively conserved in evolution, while the functions of CHS-like have diverged. According to RNA-seq data, *CsPKSs* were differentially expressed in five tissues in cultivar Diku, especially the genes closely related to phenolic terpenoids, flavonoids, and stilbene secondary metabolites in specific expression in female flowers and bracts. **Conclusion** This study will provide a direction for further functional study of the CsPKS gene family, lays a foundation for further elucidating the biosynthetic mechanism of terpene phenolic compounds in *C. sativa* and the physiological function of CsPKSs, and provides theoretical basis for the development of new polyketides for medicinal resource.

Key words: *Cannabis sativa* L.; gene family; polyketide synthases; bioinformatics; gene expression; secondary metabolites

聚酮类化合物 (polyketide, PK) 是一类由细菌、真菌、放线菌及植物产生的具有重要生物活性的天然次生代谢产物, 它们在植物与环境相互作用以及植物生长发育方面发挥着重要的生物学意义^[1]。聚酮合酶 (polyketide synthases, PKSs) 是催化聚酮化合物合成过程中的关键酶, 该类酶可催化起始底物酰基辅酶 A (CoA) 与丙二酰-CoA 缩合形成一系列的聚酮类化合物骨架。PKSs 根据其蛋白结构和催化机制的不同分为 I、II 和 III 型^[2]。I 型和 II 型 PKS 主要存在于微生物中。I 型 PKS 是以模块形式存在的复合酶, 每个模块含有多个催化功能域, 每个模块负责完成一次链延伸反应, 这些模块包含不同的活性位点, 以至于每个酶催化的反应均具有不同的活性^[3]; II 型 PKS 是一组多酶复合物, 主要催化多环芳族聚酮化合物的生物合成^[2]。III 型 PKS 广泛存在于植物中, 是一类同源二聚体蛋白, 主要由查耳酮合酶 (chalcone synthase, CHS) 和类 CHS 超家族组成^[4-5]。目前基于基因组水平的 *PKS* 基因家族已在多种被子植物中 (例如水稻、玉米、香蕉、大豆和陆地棉等) 被鉴定^[4, 6-11], 且它们的研究主要集中在苯丙烷两大代谢分支类黄酮和木脂素的生物合成中。例如, 砀山酥梨 *PbPKS* 参与梨果实石细胞和木质素的合成^[12], 从而决定梨果实品质; 而 CHS 作为花青素生物合成的关键酶已经在多种植物中被鉴定。植物中已发现的 III 型 PKS 根据其催化产物结构的区别可分为 CHS 以及类 CHS 超家族 (chalcone synthase-like, CHSL) 包括 2-吡喃酮合酶

(2-pyrone synthase, 2-PS)、芪类合酶 (stilbene synthase, STS)、联苄基合酶 (bibenzyl synthase, BBS)、吡啶酮合酶 (acridone synthase, ACS)、二苯甲酮合酶 (benzophenone synthase, BPS)、苯戊酮合酶 (valerophenone synthase, VPS)、苯亚甲基丙酮合酶 (benzalacetone synthase, BAS) 以及橄榄醇合成酶 (olivetol synthases, OLS) 等^[6]。

汉麻 *Cannabis sativa* L. 又名火麻、胡麻和线麻, 是大麻科 (Cannabinaceae) 大麻属 *Cannabis* L. 一年生草本植物。汉麻具有重要的经济和药用价值, 其干燥成熟果实即火麻仁可润肠通便, 滋阴养血^[13], 其花和苞片中富含的萜酚类化合物, 具有镇痛、抗肿瘤、抗炎等作用。其中大麻二酚 (cannabidiol, CBD) 和精神活性物质四氢大麻酚 (tetrahydrocannabinol, THC) 的结构、生物合成和生物活性已经得到了深入的研究^[14-16]。CBD 作为非精神活性物质, 其巨大的药用潜力正逐步被挖掘^[17]。汉麻中萜酚类化合物的生物合成途径主要来自脂肪酸途径和萜类途径。OLS 作为新型 PKS, 是汉麻萜酚类化合物前体物质合成的关键酶^[18]。OLS 与橄榄酸环化酶 (olivetolic acid cyclase, OAC) 将丙二酰辅酶 A (malonyl-CoA) 与己酰辅酶 A (hexanoyl-CoA) 缩合生成 2,4-二羟基-6-戊基-苯甲酸 (olivetolic acid, OA)。OA 和香叶基二磷酸 (geranyl diphosphate, GPP) 在大麻醇酸合成酶 (cannabigerolic acid synthase, CBGAS) 的催化下合成汉麻酚萜类化合物的主要前体物质大麻醇酸 (cannabigerolic acid, CBGA)。此外, 汉麻

中的 PKS 还参与多种次生代谢产物如类黄酮和萜类等的生物合成^[19]。但是目前在全基因组层面对汉麻 PKS 基因家族的分析还未有报道,因此对汉麻中 PKS 基因家族的物理化学性质,基因和蛋白结构特征以及基因的表达模式等信息的挖掘,对汉麻中萜酚类化合物以及其他次生代谢产物的合成有着重要的意义。

本研究利用生物信息学工具,对汉麻基因组中聚酮合酶基因家族进一步挖掘与分析。根据 CsPKS 基因家族成员的同源性和基因结构特征,探究 CsPKS 基因家族在进化过程中的复制事件,结合基因的表达模式探讨功能上是否存在部分或者完全冗余;同时研究 CsPKSs 基因在染色体上的分布情况,从而说明 CsPKS 基因家族的规模在进化过程中经历了收缩或者扩张;通过分析启动子区域窥探基因的表达模式,为揭示 CsPKSs 功能提供了方向,为进一步研究汉麻中萜酚类化合物以及类黄酮的生物合成机制和 CsPKSs 在汉麻中的生理功能奠定基础,并对开发新聚酮化合物药用资源提供理论基础。

1 材料与方法

1.1 CsPKS 家族成员的获取

从 NCBI (<https://www.ncbi.nlm.nih.gov>) 网站上获取雌性汉麻 CBDRx 的全基因组及注释文件 (accession number: GCA_900626175.1), 应用 TBtools^[20] 软件提取汉麻蛋白序列, 于 Uniport (<https://www.uniprot.org/>) 网站进行预测, 预测得到的所有 PKS 家族成员, 进一步去除重复转录本, 利用 TBtools 软件获得 CsPKS 家族的 gene ID 所对应的蛋白序列。同时从 Uniport 获得拟南芥和水稻的 PKS 家族成员, 利用 BlastTP 工具比对汉麻 CBDRx 的蛋白序列文件, 以 E 值 $<1 \times 10^{-5}$ 获得汉麻 CsPKS 序列, 去冗余后与上述方法进行比较, 获得最终的汉麻 PKS 家族成员。使用生物信息学工具 ExPASy-ProSite (<http://web.expasy.org/protparam/>) 对 CsPKSs 蛋白的相对分子质量和等电点进行预测。利用 softberry 网站 (<http://www.softberry.com/cgi-bin/programs/proloc/protcomppl.pl>) 进行亚细胞定位预测。

1.2 CsPKSs 蛋白序列的保守型和系统发育树分析

利用 DNAMAN 将 CsPKSs 与其他物种的 PKS 氨基酸序列进行保守结构域的同源性比对。使用 MEGA6.0^[21] 通过邻接 (neighbour-joining, NJ) 方法构建系统发育树, 并进行重复 1000 次的 Bootstrap 检验。利用 MEME (<http://meme-suite.org/tools/>

meme) 网站预测保守基序 Motif (参数设为 10)。用 TBtools 软件对 CsPKSs 基因进行外显子和内含子、保守结构域和系统进化树三者可视化处理。

1.3 CsPKSs 基因的染色体定位及共线性分析

使用 TBtools 软件解析汉麻的基因组注释文件, 确定每个 PKS 基因的位置信息并绘制其所对应的染色体物理位置图。从 NCBI 网站下载拟南芥、汉麻、水稻和番茄的全基因组数据, 使用 TBtools 软件对其共线性关系进行可视化处理, 并找到 PKS 基因的同源基因对。

1.4 CsPKSs 基因顺式作用元件预测

使用 PlantCare (<http://www.plantcare.co.uk/>) 网站对 CsPKSs 基因顺式作用元件进行预测, 并使用 TBtools 软件对其可视化构图。

1.5 CsPKSs 蛋白二级和三级结构预测

通过 SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/secpred_sopma.pl) 在线预测蛋白质二级结构。利用 AlphaFold^[22] 构建 CsPKSs 蛋白三维结构模型。首先分析靶蛋白序列并进行多序列比对, 随着深度神经网络学习预测原子对的距离和扭转分布, 通过梯度下降的蛋白特异性电位来构建靶蛋白结构^[23]。使用的工具包括: HHblits^[24] (3 次迭代, $E=1 \times 10^{-3}$); PSI-BLAST v.2.6.0 nr 数据集^[25] (3 次迭代, $E=1 \times 10^{-3}$); BioPython v.1.65^[26]; 用于结构可视化的 PyMol 2.2.0^[27]。

1.6 CsPKSs 基因表达模式分析

根据已公布的 9 个不同栽培品种 (Black Berry Kush (BB)、Sour Diesel (SD)、Black Lime (BL)、Canna Tsu (CT)、Cherry Chem (CC)、Terple-3 (T3)、Mama Thai (MT)、Valley Fire (VF)、White Cookies (WC) 雌株腺毛转录组数据 (PRJNA498707) 以及商业品种 Dinamed kush CBD autoflowering (Diku) 的花、叶、苞片、茎、种子和根的转录组数据 (Accession No.: bract: SMAN16122880 ~ SAMN16122882; stem: SAMN16122883 ~ SAMN16122885; flower: SAMN16122886 ~ SAMN16122888; leaf: SAMN16122889 ~ SAMN16122891) 计算 CsPKSs 基因的表达量 (FPKM), 并运用 TBtools 软件绘制热图, 进行基因聚类 and 差异表达分析。具体比对以及计算方法如下: 使用 star 软件将 RNA-seq 数据比对至参考基因组, 然后使用 rsem 软件包中的 rsem-calculate-expression 工具对比对结果进行定量分析, 获得表达量 FPKM 值。

1.7 CsPKS 基因表达

根据进化关系以及 RNA-seq 数据对本研究中筛选的 *CsPKS1*、*CsPKS4-5* 和 *CsPKS9* 基因进行实时荧光定量 PCR (quantitative real-time PCR, qRT-PCR) 验证。汉麻 Diku 品种的花、苞片、种子、根、叶和茎的 RNA 提取使用 RNeasy Pure Plant Kit 试剂盒 (北京天根有限公司), 采用 TransScript® One-Step gDNA Removal and cDNA Synthesis

SuperMix 试剂盒 (北京全式金公司) 逆转录合成 cDNA, 使用 StarLighter SYBR Green qPCR Mix 试剂盒 (北京启衡星公司) 在 Rotor-Gene Q (QIAGEN 公司, 德国) 上进行 qRT-PCR 检测, 反应程序: 95 °C 酶激活 5 min, 95 °C 变性 30 s, 58 °C 退火 20 s, 72 °C 延伸 15 s, 循环次数 40, 用 $2^{-\Delta\Delta C_t}$ 法计算基因相对表达量^[28]。相关引物见表 1, 内参基因选择 *EF1 α* ^[29]。本实验进行了 3 次生物学重复。

表 1 qRT-PCR 引物

Table 1 Primers for qRT-PCR

基因	正向 (5'-3')	反向 (5'-3')	长度/bp
<i>EF1α</i>	ACCAAGATTGACAGGCGTTC	CCTTCTTCTCCACAGCCTTG	100
<i>CsPKS1</i>	CCCTCAGTGAAGCGTGTGAT	TATGTCACAACAGACGGCGA	126
<i>CsPKS4</i>	TCTCGAAGGAAAACCCACCG	CGCTGGATGAGACGTAGACC	149
<i>CsPKS5</i>	GCTGTTGAGGCTAAGCTCCA	CTGTGGTGGACTTCCCTTCC	151
<i>CsPKS9</i>	GCCAGGCCCTATTTGGTGAT	ATAGTTTGGGCCCGCAGAGAC	106

2 结果与分析

2.1 CsPKS 家族成员的获取

从 NCBI 网站获取汉麻的全基因组及注释文件 (accession number: GCA_900626175.1), 利用 TBtools 软件提取汉麻蛋白序列, 将 PlantTFDB 网站获得的拟南芥和水稻 PKSs 蛋白序列与汉麻蛋白

序列数据进行比对, 去除重复转录本后, 共筛选得到 *CsPKS* 基因家族成员 9 个, 依次命名为 *CsPKS1*~*CsPKS9*, 其编码氨基酸长度为 356~412 aa; 蛋白相对分子质量的大小为 39 285.38~45 282.99; 等电点介于 5.52~8.07。亚细胞定位预测 9 个基因全部定位于细胞质 (表 2)。

表 2 *CsPKS* 基因家族成员基本信息及特征

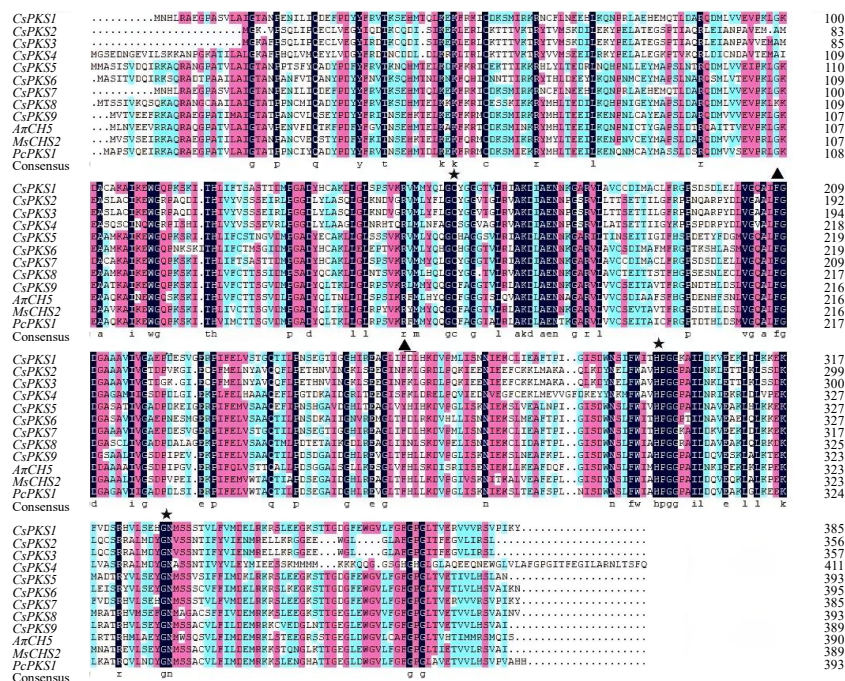
Table 2 Information and characteristics of *CsPKSs*

基因号 Locus	转录本登录号	基因名称	长度/aa	等电点	相对分子质量	亚细胞定位
LOC115700696	rna-XM_030628326.1	<i>CsPKS1</i>	385	6.04	42 576.11	细胞质
LOC115708801	rna-XM_030636816.1	<i>CsPKS2</i>	356	5.52	39 285.38	细胞质
LOC115715682	rna-XM_030644341.1	<i>CsPKS3</i>	357	5.67	39 392.50	细胞质
LOC115704317	rna-XM_030631533.1	<i>CsPKS4</i>	412	5.53	45 282.99	细胞质
LOC115713254	rna-XM_030641739.1	<i>CsPKS5</i>	393	6.68	43 367.82	细胞质
LOC115699292	rna-XM_030626628.1	<i>CsPKS6</i>	395	8.07	43 696.44	细胞质
LOC115699293	rna-XM_030626629.1	<i>CsPKS7</i>	385	6.04	42 576.11	细胞质
LOC115724130	rna-XM_030653598.1	<i>CsPKS8</i>	393	7.93	42 932.80	细胞质
LOC115724170	rna-XM_030653640.1	<i>CsPKS9</i>	389	6.04	42 720.37	细胞质

2.2 CsPKS 序列的保守型和系统发育树分析

利用 DNAMAN 软件将 9 条 *CsPKS* 氨基酸序列与 GenBank 数据库中其他植物 PKS 氨基酸序列进行比对分析 (图 1)。*CsPKS* 氨基酸序列与金鱼草 *Antirrhinum majus* L. (BAE80511)、紫花苜蓿 *Medicago sativa* L. (P30074)、虎杖 *Polygonum cuspidatum* L. (ABK92282) 中的氨基酸序列具有较高的相似度。*CsPKS* 具有植物 III 型 PKS 超家族三

联体活性中心位点 Cys164、His303 和 Asn336 (图 1)^[30], 其中 Cys164 是聚酮链形成过程中的亲核活性位点, 通过共价键与起始底物分子和反应中间体结合, 而 His303 和 Asn336 则起着稳定或活化起始底物和延伸底物的作用。同时 *CsPKSs* 氨基酸序列具有 CHS 典型的活性位点^[31], 如 Arg38、Phe49、Thr50、Thr132、Gly211、Gly216 等。起“Gatekeeper”作用的 2 个氨基酸残基 Phe215 和 Phe265 在大多数



★-PKS 中的活性位点残基 (Cys164、His303 和 Asn336) ▲-“Gatekeeper” (Phe215 和 Phe265) MsCHS2-紫花苜蓿 CHS2 (P30074) AmCHS-金鱼草 CHS (BAE80511) PcPKS1-虎杖 PKS1 (ABK92282)
★-active site residues in PKS (Cys164、His303 和 Asn336) ▲-“Gatekeeper” (Phe215 和 Phe265) MsCHS2-Medicago sativa CHS2 (P30074) AmCHS-Antirrhinum majus CHS (BAE80511) PcPKS1-Polygonum cuspidatum PKS1 (ABK92282)

图 1 PKS 氨基酸序列比对

Fig. 1 Alignment of PKSs

CsPKSs 中高度保守, 但 CsPKS5 和 CsPKS8 中 Phe256 分别被 Tyr 和 Ile 取代, 与其他 PKS 存在较为明显的差异。

为了阐明 CsPKS 与其他植物物种 PKS 之间的进化关系, 对 CsPKS 与拟南芥、水稻、苜蓿等植物的 PKS、CHS、CHSL、STS 和 VPS 等进行了系统发育树分析。系统发育重建后发现, CsPKS 蛋白共划分为 3 个组, 分别命名为 group 1、group 2 和 group 3 (图 2)。针对 9 个 CsPKS, 发现第 1 组共包含 CsPKS2、CsPKS3 和 CsPKS4 3 个成员, 其中 CsPKS2 和 CsPKS3 聚为一支后与 NsCHSL 聚在一起, CsPKS4 与 AtPBSB 聚为一支。第一组 PKS 多与模式植物如拟南芥、水稻和烟草等关系密切; 第 2 组包含 5 个 CsPKS, 分别是 CsPKS1、CsPKS 5~8, 该组 5 个成员与 VvSTS1、RpBAS 和 HIVPS 关系较为密切, 其中 CsPKS5 和 CsPKS8 聚为一支; CsPKS1、7 和 6 聚为一支后与啤酒花的 VPS 聚为一支; 第 3 组仅包含 CsPKS9, 其与啤酒花 CHS 同源。

CsPKSs 基因结构分析显示 (图 3) 所有的 CsPKSs 基因中仅有 1 个内含子这与其他植物 III 型

PKS 基因的结构较为保守相符。裸子植物和被子植物中, 除已报道的金鱼草外, 其余已报道 III 型 PKS 基因均含有 1 个内含子且该内含子位置保守^[31-33]。

CsPKS2 和 CsPKS3 在 3'端无非翻译区 (untranslated region, UTR), CsPKS4 和 CsPKS6 内含子较长, 分别为 1187、2075 bp。通过保守基序分析发现, Motif1、Motif2、Motif4、Motif5、Motif7、Motif8、Motif9 存在于所有 CsPKSs 中, Motif3 和 Motif6 存在于 CsPKS1、CsPKS5、CsPKS6、CsPKS7、CsPKS8、CsPKS9 中, Motif10 存在于 CsPKS2、CsPKS3、CsPKS4 中。

2.3 CsPKS 基因的染色体定位及共线性分析

通过对 CsPKS 家族成员染色体进行定位分析 (图 4), 发现该家族 9 个成员分布在 Chr1、Chr3、Chr5、Chr7、Chr10 号染色体上。其中, CsPKS1、CsPKS6 和 CsPKS7 成簇位于 10 号染色体运用比较基因组学的方法和共线性原理挖掘汉麻与双子叶模式植物拟南芥、番茄以及单子叶模式植物水稻之间的共线性关系。图 5 中高亮区域显示 PKS 家族在 3 个物种间的同源基因对, 汉麻和拟南芥中, 一共发现 2 对同源基因 (CsPKS4 和 AtPKSB, CsPKS1 或 7

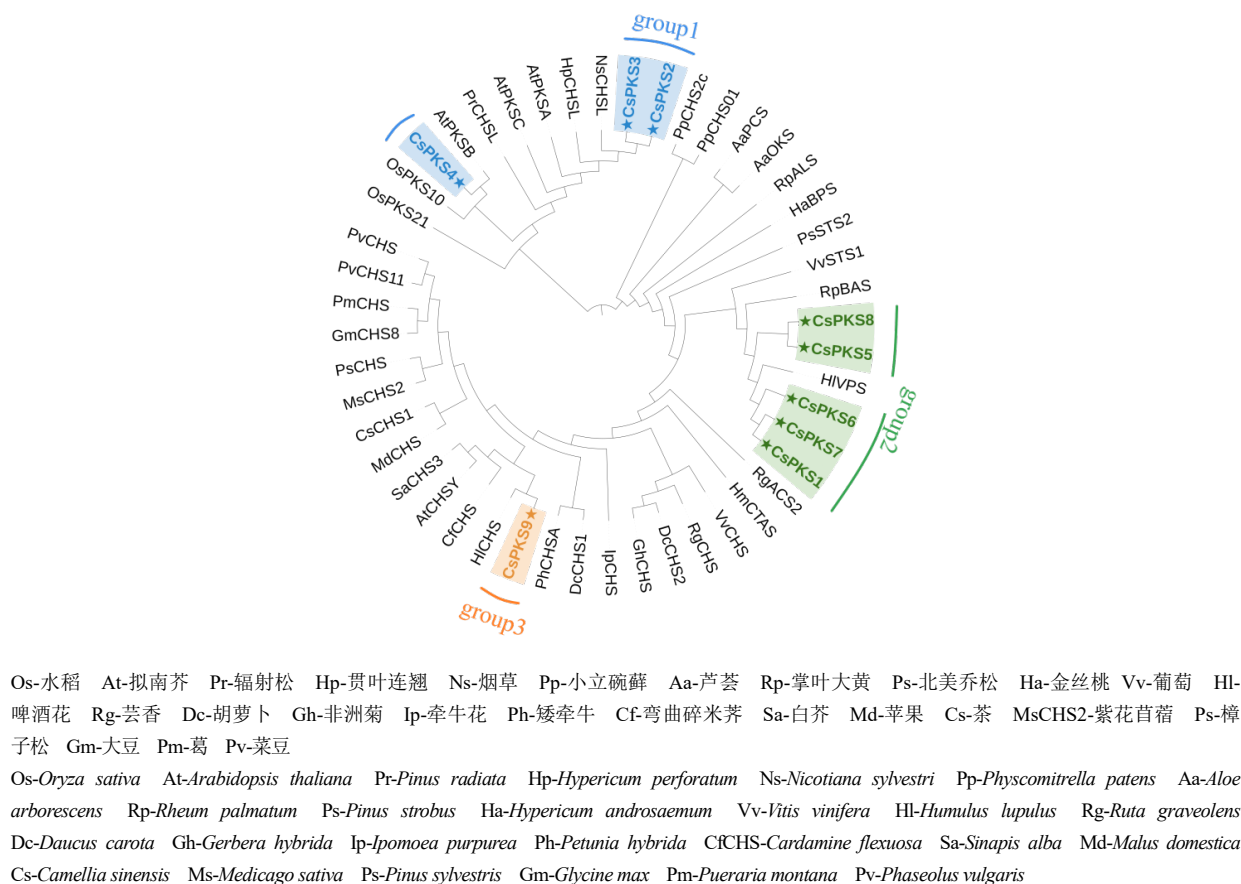


图 2 CsPKS 家族基因系统发育树及其分类

Fig. 2 Phylogenetic analysis of PKSs from *C. sativa* and other plant species

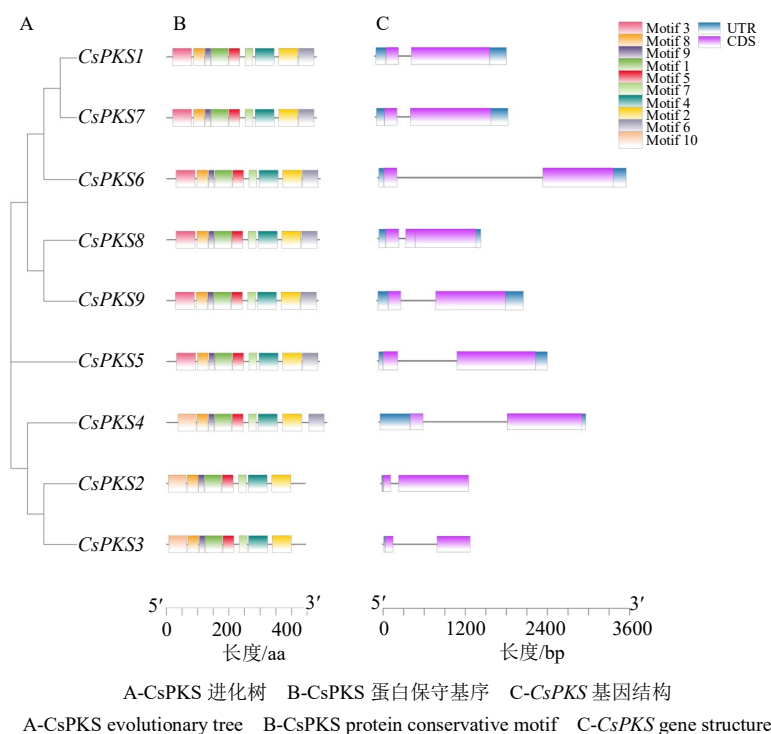


图 3 CsPKS 的基因结构及蛋白质保守基序分析

Fig. 3 Gene structure and conserved motif analysis of *CsPKS*

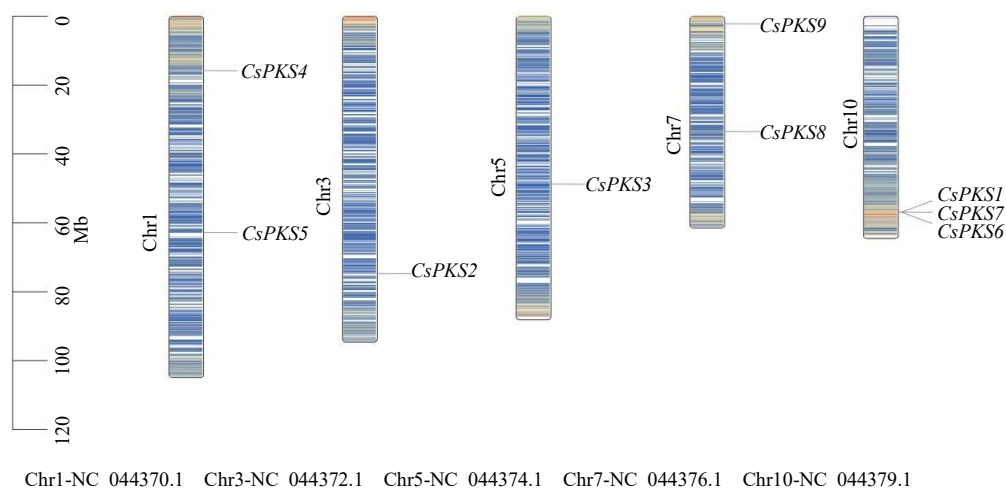
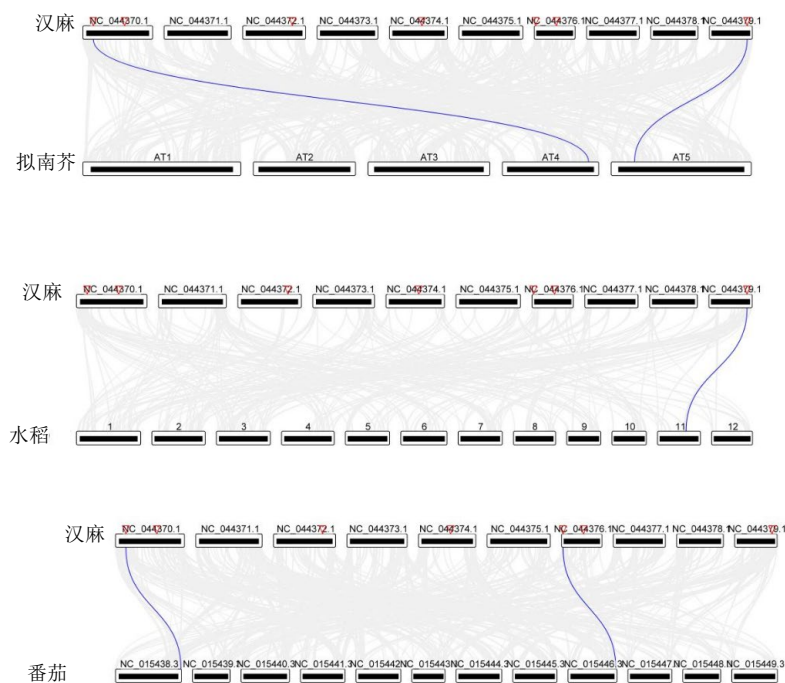
图 4 *CsPKS* 基因染色体定位Fig. 4 Chromosomal locations of *CsPKS* genes in *C. sativa*

图 5 汉麻与拟南芥、水稻和番茄的共线性分析

Fig. 5 Collinear relationship of *C. sativa* with respect to *Arabidopsis thaliana*, *Oryza sativa* and *Solanum lycopersicum*

和 *AtCHSY*)，其中 *CsPKS1* 和 *CsPKS7* 为一对串联重复基因，且氨基酸序列完全一致；在汉麻和水稻中，一共发现了 1 对同源基因 (*CsPKS1* 或 *CsPKS7* 和 *OsCHS1*)；汉麻和番茄中，一共发现 2 对同源基因 (*CsPKS4* 和 *SIPKSB*，*CsPKS9* 和 *NM_001247104.2*)。

2.4 *CsPKS* 基因顺式作用元件预测

提取 *CsPKS* 基因的翻译起始密码子 (ATG) 上

游的 2 kb 作为启动子序列，对 *CsPKS* 基因家族成员顺式作用元件进行预测和可视化分析 (图 6)，结果发现，核启动元件、光响应元件存在于汉麻所有 *PKS* 基因的启动子区域。*CsPKS1*、*CsPKS6*、*CsPKS7* 中顺式作用元件的数目和种类均较多，*CsPKS3* 所含顺式作用元件数目最少，且干旱诱导元件仅存在于 *CsPKS3* 中。*CsPKS* 基因启动子区的顺式作用元件种类及分布具有多样性。

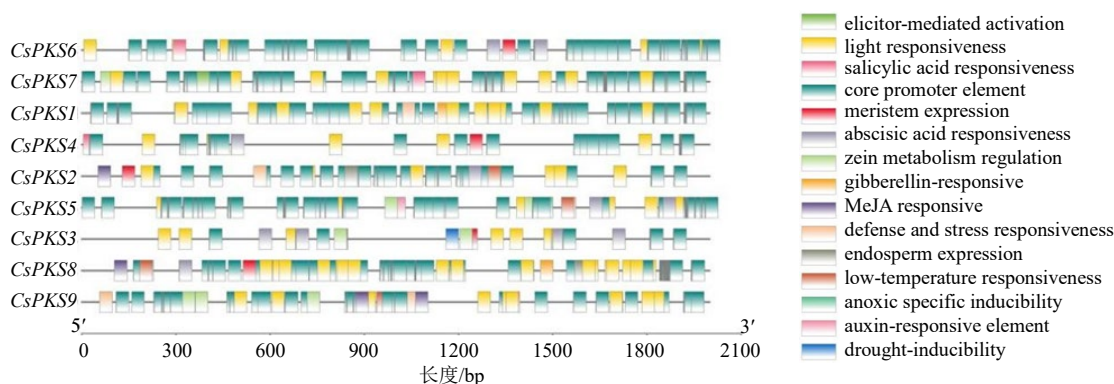


图 6 *CsPKS* 家族基因的启动子区顺式作用元件

Fig. 6 *Cis*-acting elements within the promoters of *CsPKS* genes

2.5 *CsPKSs* 蛋白二级和三级结构分析

蛋白质二级结构预测结果如表 3 所示, *CsPKSs* 的 α -螺旋介于 38.11%~46.56%; 延伸链和 β -折叠在整个二级结构中占比较少, 分别为 14.61%~18.45%和 5.85%~8.50%; 随机卷曲介于 32.88%~62.40%。使用 Alphafold 对目前已鉴定功能的 OLS (*CsPKS1*/*CsPKS7*) 和 *CsPKS9* 蛋白进行三维建模(图 7)并通过 Pymol 软件实现蛋白构象的可视化。AlphaFold 的核心组件是一个卷积神经网络, 它通过对 Protein Data Bank (PDB) 数据库中的蛋白结构深度学习来预测蛋白质残基的 C_{β} 原子对 ij 之间的距离 (d_{ij})^[23]。OLS 和 *CsPKS9* 蛋白主要由 α -螺旋和无规则卷曲形成活性口袋组成, Asn330、His257、Cys157、Phe208 和 Phe259 为 OLS 活性位点; Asn330、His257 和 Cys157 是 OLS 活性中心位点, Phe208 和 Phe259 为 OLS 起“Gatekeeper”作用的两个氨基酸残基; Asn336、His303、Cys167、Phe215 和 Phe265 为

CsPKS9 活性位点: Asn336、His303 和 Cys167 是 *CsPKS9* 活性中心位点, Phe215 和 Phe265 为 *CsPKS9* 起“Gatekeeper”作用的 2 个氨基酸残基。

2.6 *CsPKS* 基因表达模式分析

为探究 *CsPKS* 基因的表达模式, 对汉麻品种 Diku 的花、苞片、种子、根、叶、茎 6 个不同组织和 9 个不同汉麻品种的腺毛转录组数据进行基因表达差异分析, 并绘制热图。研究发现 *CsPKS9* 在花中表达量最高, 苞片和叶中表达量均较高, 茎中表达较低, 种子和根中几乎不表达; *CsPKS1*、*CsPKS7* 在茎中表达量较低, 种子和根中几乎不表达, 花和苞片中表达较高, 且在苞片表达量高于花; *CsPKS5* 在花和苞片中表达量较低, 在其他组织中不表达; 其余基因在 6 个组织中都几乎不表达。早期的研究显示 *CsPKS1* 和 *CsPKS7* 为汉麻中萜酚类化合物合成途径的关键酶基因^[17]。通过热图聚类可以看出, 在汉麻 9 个不同品种的腺毛中(图 8), *CsOLS*

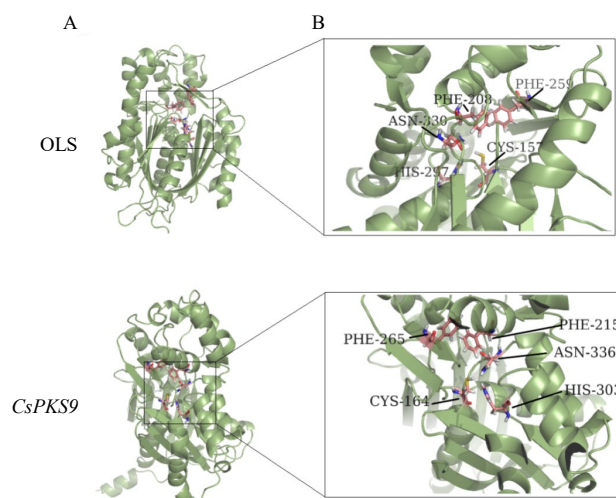
表 3 *CsPKS* 蛋白二级结构

Table 3 Predicted protein secondary structures of *CsPKS*

名称	α -螺旋/%	延伸链/%	β -折叠/%	无规则卷曲/%	二级结构原件
<i>CsPKS1</i>	42.08	15.06	7.53	35.32	
<i>CsPKS2</i>	46.35	14.61	7.30	31.74	
<i>CsPKS3</i>	45.10	15.97	6.44	32.49	
<i>CsPKS4</i>	38.11	18.45	8.50	34.95	
<i>CsPKS5</i>	46.56	15.27	7.38	30.79	
<i>CPKS6</i>	42.53	15.19	6.84	35.44	
<i>CsPKS7</i>	42.08	15.06	7.53	35.32	
<i>CsPKS8</i>	44.02	15.78	5.85	34.35	
<i>CsPKS9</i>	44.47	14.91	7.71	32.90	

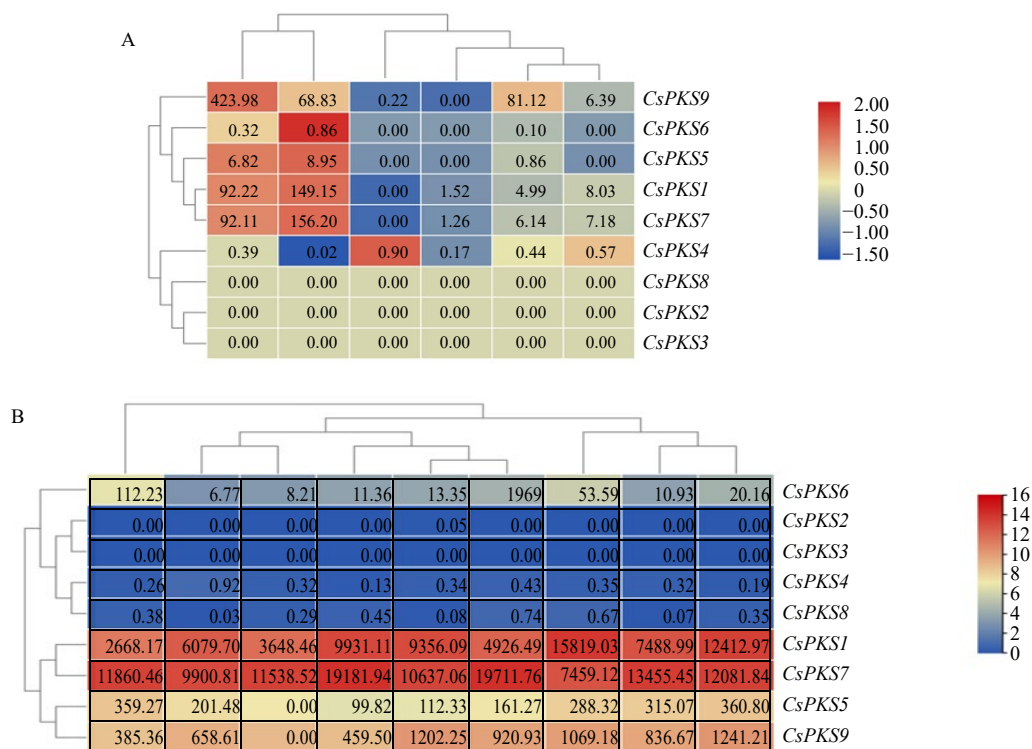
蓝色- α -螺旋 绿色- β -折叠 黄色-无规则卷曲 红色-延伸链

blue-alpha helix green-beta turn yellow-random coil red-extended strand



A-OLS 和 CsPKS9 蛋白整体的 3D 构象, 以及关键氨基酸残基所在的空间位置 B-关键氨基酸残基之间在相应蛋白中的空间分布状态
A-OLS and CsPKS9 protein as a whole, and the spatial position of the key amino acid residues B-the spatial distribution of the key amino acid residues in the corresponding protein

图 7 CsOLS (CsPKS1 或 CsPKS7) 和 CsPKS9 蛋白三维结构
Fig. 7 Three-dimensional structure of CsOLS and CsPKS9



A-CsPKS 基因在 Diku 品种不同组织器官的表达模式, 数据经过 $\log_2(\text{FPKM}+1)$ 对数转化处理和 row scale 均一化处理 (即颜色随着该基因的组织分布差异而变化) B-CsPKS 在 9 个汉麻品种腺毛的表达模式, 数据经过 $\log_2(\text{FPKM}+1)$ 对数转化处理
A-CsPKS gene in different tissues and organs of Diku variety were processed by $\log_2(\text{FPKM}+1)$ logarithmic transformation and row scale homogenization (that is, the color changes with the difference in the tissue distribution of the gene) B-CsPKS in the glandular hair of 9 hemp varieties were processed by $\log_2(\text{FPKM}+1)$ logarithmic transformation. The number is its original FPKM value

图 8 CsPKS 基因的表达模式分析
Fig. 8 Expression pattern of CsPKS genes in different tissues of Diku and trichomes of different varieties

(*CsPKS1* 和 *CsPKS7*) 在所有品种的腺毛中都高度表达, *CsPKS5* 和 *CsPKS9* 在所有腺毛中也均有较高的表达, 其次是 *CsPKS6*。 *CsPKS2*、*CsPKS3*、*CsPKS4* 和 *CsPKS8* 在所有品种中表达量很低或者不表达。

2.7 qRT-PCR 验证 *CsPKS* 基因表达

为了进一步验证转录组数据的表达分析结果, 本

研究利用 qRT-PCR 验证 *OLS* (*CsPKS1/7*)、*CsPKS4-5*、*CsPKS9* 在汉麻 Diku 品种不同组织部位的表达情况(图 9)。结果显示 *OLS*、*CsPKS5*、*CsPKS9* 在汉麻花和苞片中的表达量比在其他组织部位的表达量高 ($P < 0.05$), 与转录组结果一致。由于 *CsPKS4* 基因整体表达量极低, 其 qRT-PCR 结果与转录组趋势略有不同。

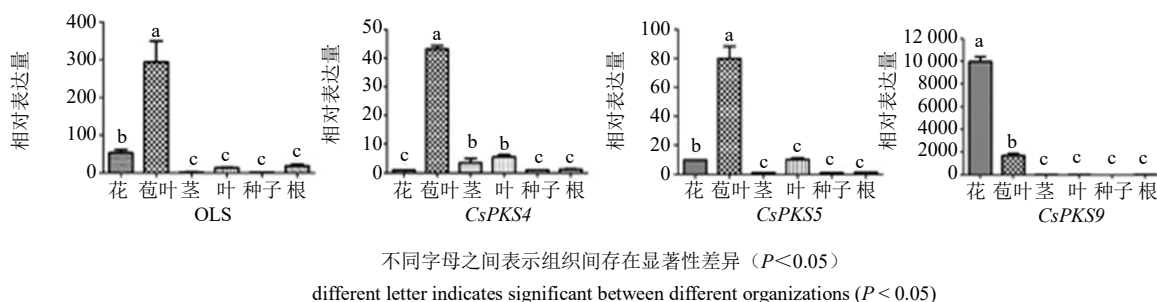


图 9 *CsPKS* 在不同组织中的相对表达水平

Fig. 9 Relative gene expression of the *CsPKS* in different tissues

3 讨论

PKS 是聚酮化合物合成过程中的关键酶, 在植物生长发育、抗逆以及植物与环境之间的信息交互等方面起着重要的作用^[1]。基于基因组水平的 PKS 基因家族成员已在多种植物中被鉴定^[7], 目前水稻和蔷薇科梨^[12]中含有数量最多的 PKS 成员, 达 27 个, 而在其他被子植物中 PKS 的数量基本在 4-20 个之间(香蕉 20 个^[9]、玉米 14 个^[8]、大豆 14 个^[10]、悬钩子 10 个^[11]和兰花 5 个^[34]等)。通过基因组水平的研究发现, PKS 在植物基因组中的数量差异可能与物种是否存在古老的全基因组复制或者近期的串联重复复制有关^[12]。本研究在汉麻基因组中共鉴定得到 9 个 *CsPKS* 基因, 其中 *CsPKS1*、*CsPKS7* 和 *CsPKS6* 在基因组上成簇存在, *CsPKS1* 和 *CsPKS7* 核酸序列相似度为 99.40%, 氨基酸序列相似度为 100%; *CsPKS6* 与 *CsPKS7* 的核酸序列相似度为 69.1%, 并且它们具有相似的保守基序和保守结构域分布。Gao 等^[35]通过对 1 个野生品种的汉麻深度测序, 发现其发生了 3 次近代全基因组复制事件和一次大规模的基因复制事件, 因此推测全基因组复制事件可能是产生 *CsPKS1* 和 *CsPKS7* 基因对的直接原因。

CsPKS 在汉麻类黄酮和酚萜类化合物的生物合成中发挥着重要的作用。*CsPKS* 在酚萜类化合物合成途径中以己酰辅酶 A (hexanoyl-CoA) 和丙二酰辅酶 A (malonyl-CoA) 作为底物生成中间产物丁烯酮辅酶 A (tetraketide-CoA); *CsPKS* 在类黄酮合

成途径中以香豆酰辅酶 A (p-coumaroyl-CoA) 和丙二酰辅酶 A 为底物生成柚配基查耳酮。*CsPKS1/7* 与 Taura 等^[18]克隆鉴定的汉麻酚萜类化合物合成途径基因 *OLS* 为同一基因, 因此 *CsPKS1/7* 是参与汉麻酚萜类化合物生物合成的关键酶基因。另外, *CsPKS9* 与 Raharjo 等^[36]在汉麻中克隆的参与汉麻类黄酮化合物合成的 *CHS* 基因序列一致, 这也表明了 *CsPKS9* 是汉麻类黄酮合成的关键酶基因。*CsPKS1/7* 和 *CsPKS6* 在系统发育树中与 HIVPS 聚类^[37], 以异戊酰辅酶 A (isovaleryl-CoA) 和异丁基辅酶 A (isobutyryl-CoA) 作为偏好性底物。在汉麻与拟南芥以及水稻的共线性分析中发现 *CsPKS1/7* 与 AtCHSY 和 OsCHS1 同源, 而 AtCHSY 和 OsCHS1 已报道的催化产物是柚配基查耳酮, 这也说明了 PKS 在进化过程中功能产生了分化。在系统发育树中 *CsPKS9* 与其它多种植物的 CHS 聚类, 且在共线性分析中 *CsPKS9* 和番茄 SlCHS 同源, 它们都参与了柚配基查耳酮的生成, 这也表明 CHS 作为类黄酮合成的关键酶, 其功能是相对保守的。值得一提的是, *CsPKS9* 不仅可以香豆酰辅酶 A (p-coumaroyl-CoA) 作为底物, 也可以异戊酰辅酶 A 或异丁基辅酶 A 作为底物分别生成间苯二酮和间苯异丁苯甲酮^[38], 这也表明 *CsPKS9* 具有 VPS 活性。

CsPKS2、*CsPKS3* 和 *CsPKS4* 与 NsCHSLK、AtPKSA 和 AtPKSB 聚为一支, 研究发现烟草 NsCHSLK 在花药发育中, 特别是对花粉粒外壁合成中起促进作用, 它的 mRNA 的表达仅限于小孢子

和绒毛细胞且 *NsCHSLK* 启动子表现出花药特异性^[39]；在拟南芥花药的孢子发育过程中，*AtPKSA* 和 *AtPKSB* 在绒毛膜细胞中存在组织特异性和时空特异性表达，其参与了孢粉蛋白的合成，被分类为花药特异性 *CHS* like (anther specific chalcone synthase-like, ASCL) 蛋白^[36,40]。由于本研究中涉及的汉麻为雌性植株，不存在花药的发育过程，所以转录组中 *CsPKS2*、*CsPKS3* 和 *CsPKS4* 基因的表达量很低甚至不表达。因此，*CsPKS2*、*CsPKS3* 和 *CsPKS4* 基因可能在雄性大麻花药发育过程中发挥作用。另外，*CsPKS5* 和 *CsPKS8* 与 *STS* 聚为一支，因此这 2 个 *PKS* 可能参与汉麻萜类化合物的生物合成。*CsPKS* 启动子的激素诱导和抗逆胁迫类顺式作用元件数量较多，表明 *PKS* 基因家族在汉麻生长发育过程中，尤其是在响应胁迫中可能具有重要的调控作用^[41]。

植物中聚酮类化合物的积累以及关键 III 型 *PKS* 基因的表达往往具有一定的组织特异性。例如，*GhCHS* 主要在陆地棉 *Gossypium hirsutum* Linn. 的纤维中表达，与纤维素的生物合成密切相关^[42]；凹缘金虎尾 *Malpighia emarginata* L. 中的 *MeCHS* 的表达在果实成熟过程中显著上调^[43]；*HIVPS* 基因在啤酒花蛇麻腺中特异性表达，直接影响苦味酸的生物合成^[37]。汉麻雌花和苞片是萜类化合物等次生代谢产物的主要合成场所^[36]，而与这些次生代谢产物密切相关的基因比如 *CsPKS1*、*CsPKS5-7* 和 *CsPKS9* 在花和苞片以及不同品种的汉麻腺毛中都高度表达。

PKS 是聚酮类化合物的合成关键酶和限速酶。本研究通过生物信息学，对 *CsPKS* 在全基因组水平上进行了鉴定与分析，对该家族成员的物理特性、基因结构、保守基序等进行分析，并探究了汉麻与其他物种间 *PKS* 的进化关系，确定了各基因在汉麻的组织表达特异性，为探明 *CsPKS* 基因功能提供理论支撑，并为寻找潜在的汉麻中聚酮类合成的关键酶提供理论依据。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress [J]. *Curr Opin Plant Biol*, 2002, 5(3): 218-223.
- [2] Flores-Sanchez I J, Verpoorte R. Plant polyketide synthases: A fascinating group of enzymes [J]. *Plant Physiol Biochem*, 2009, 47(3): 167-174.

- [3] Moore B S, Hopke J N. Discovery of a new bacterial polyketide biosynthetic pathway [J]. *Chembiochem*, 2001, 2(1): 35-38.
- [4] Hu L F, He H H, Zhu C L, et al. Genome-wide identification and phylogenetic analysis of the *Chalcone synthase* gene family in rice [J]. *J Plant Res*, 2017, 130(1): 95-105.
- [5] Eom S H, et al. Genome-wide identification and transcriptional expression analysis of *Chalcone synthase* in flax (*Linum usitatissimum* L.) [J]. *Gene Rep*, 2016, 5: 51-56.
- [6] Lussier F X, Colatriano D, Wiltshire Z, et al. Engineering microbes for plant polyketide biosynthesis [J]. *Comput Struct Biotechnol J*, 2012, 3: e201210020.
- [7] Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation [J]. *Nucleic Acids Res*, 2016, 44(D1): D457-D462.
- [8] Han Y H, Ding T, Su B, et al. Genome-wide identification, characterization and expression analysis of the *Chalcone synthase* family in maize [J]. *Int J Mol Sci*, 2016, 17(2): 161.
- [9] Pothiraj R, Ravikumar M J, Suthanthiram B, et al. Genome-scale analyses of polyketide synthases in banana: Phylogenetics and expression profiling forecast their candidacy in specialized metabolism [J]. *Gene*, 2021, 778: 145472.
- [10] Anguraj Vadivel A K, Krysiak K, Tian G, et al. Genome-wide identification and localization of *Chalcone synthase* family in soybean (*Glycine max*[L]Merr) [J]. *BMC Plant Biol*, 2018, 18(1): 325.
- [11] Kumar A, Ellis B E. A family of polyketide synthase genes expressed in ripening *Rubus* fruits [J]. *Phytochemistry*, 2003, 62(3): 513-526.
- [12] 李国辉, 李亿红, 钱冉, 等. 梨 III 型聚酮合成酶家族的比较基因组学研究及表达模式分析 [J]. 植物生理学报, 2018, 54(6): 1005-1017.
- [13] 代丹, 邸莎, 吴浩然, 等. 大麻仁的临床应用及其用量探究 [J]. 吉林中医药, 2020, 40(2): 242-244.
- [14] Torrens A, Vozella V, Huff H, et al. Comparative pharmacokinetics of Δ^9 -tetrahydrocannabinol in adolescent and adult male mice [J]. *J Pharmacol Exp Ther*, 2020, 374(1): 151-160.
- [15] Reddy D S, Golub V M. The pharmacological basis of *Cannabis* therapy for epilepsy [J]. *J Pharmacol Exp Ther*, 2016, 357(1): 45-55.
- [16] 王欣, 于京. 大麻二酚抗肿瘤作用及其机制研究进展 [J]. 国际药学研究杂志, 2020, 47(12): 1057-1061.
- [17] Izzo A A, Camilleri M. Cannabinoids in intestinal inflammation and cancer [J]. *Pharmacol Res*, 2009, 60(2): 105-115.

- 117-125.
- [18] Taura F, Tanaka S, Taguchi C, *et al.* Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway [J]. *FEBS Lett*, 2009, 583(12): 2061-2066.
- [19] Flores-Sanchez I J, Verpoorte R. Secondary metabolism in *Cannabis* [J]. *Phytochem Rev*, 2008, 7(3): 615-639.
- [20] Chen C J, 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.
- [21] Kumar S, Tamura K, Nei M. MEGA: Molecular evolutionary genetics analysis software for microcomputers [J]. *Comput Appl Biosci*, 1994, 10(2): 189-191.
- [22] Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold [J]. *Nature*, 2021, 596(7873): 583-589.
- [23] Senior A W, Evans R, Jumper J, *et al.* Improved protein structure prediction using potentials from deep learning [J]. *Nature*, 2020, 577(7792): 706-710.
- [24] Remmert M, Biegert A, Hauser A, *et al.* HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment [J]. *Nat Methods*, 2011, 9(2): 173-175.
- [25] Altschul S F, Madden T L, Schäffer A A, *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs [J]. *Nucleic Acids Res*, 1997, 25(17): 3389-3402.
- [26] Cock P J, Antao T, Chang J T, *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics [J]. *Bioinformatics*, 2009, 25(11): 1422-1423.
- [27] Seeliger D, de Groot B L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina [J]. *J Comput Aided Mol Des*, 2010, 24(5): 417-422.
- [28] Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} Method [J]. *Methods*, 2001, 25(4): 402-408.
- [29] Guo R, Guo H Y, Zhang Q Y, *et al.* Evaluation of reference genes for RT-qPCR analysis in wild and cultivated *Cannabis* [J]. *Biosci Biotechnol Biochem*, 2018, 82(11): 1902-1910.
- [30] Jez J M, Austin M B, Ferrer J, *et al.* Structural control of polyketide formation in plant-specific polyketide synthases [J]. *Chem Biol*, 2000, 7(12): 919-930.
- [31] Jez J M, Bowman M E, Dixon R A, *et al.* Structure and mechanism of the evolutionarily unique plant enzyme *Chalcone isomerase* [J]. *Nat Struct Biol*, 2000, 7(9): 786-791.
- [32] Schröder J. A family of plant-specific polyketide synthases: Facts and predictions [J]. *Trends Plant Sci*, 1997, 2(10): 373-378.
- [33] Durbin M L, McCaig B, Clegg M T. Molecular evolution of the *Chalcone synthase* multigene family in the morning glory genome [J]. *Plant Mol Biol*, 2000, 42(1): 79-92.
- [34] Kaur A, Ghai D, Yadav V G, *et al.* Polyketide synthases (PKSs) of secondary metabolism: *in silico* identification and characterization in orchids [J]. *J Biomol Struct Dyn*, 2022, 16: 1-13.
- [35] Gao S, Wang B S, Xie S S, *et al.* A high-quality reference genome of wild *Cannabis sativa* [J]. *Hortic Res*, 2020, 7(1): 73.
- [36] Raharjo T J, Widjaja I, Roytrakul S, *et al.* Comparative proteomics of *Cannabis sativa* plant tissues [J]. *J Biomol Tech*, 2004, 15(2): 97-106.
- [37] Okada Y, Ito K. Cloning and analysis of valerophenone synthase gene expressed specifically in lupulin gland of hop (*Humulus lupulus* L.) [J]. *Biosci Biotechnol Biochem*, 2001, 65(1): 150-155.
- [38] Raharjo T J, Chang W T, Verberne M C, *et al.* Cloning and over-expression of a cDNA encoding a polyketide synthase from *Cannabis sativa* [J]. *Plant Physiol Biochem*, 2004, 42(4): 291-297.
- [39] Atanassov I, Russinova E, Antonov L, *et al.* Expression of an anther-specific *Chalcone synthase*-like gene is correlated with uninucleate microspore development in *Nicotiana glauca* [J]. *Plant Mol Biol*, 1998, 38(6): 1169-1178.
- [40] Kim S S, Grienemberger E, Lallemand B, *et al.* lap6/polyketide synthase a and lap5/polyketide synthase b encode hydroxyalkyl α -pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana* [J]. *Plant Cell*, 2010, 22(12): 4045-4066.
- [41] Dare A P, Hellens R P. RNA interference silencing of CHS greatly alters the growth pattern of apple (*Malus domestica*) [J]. *Plant Signal Behav*, 2013, 8(8): e25033.
- [42] Su X Q, Sun X, Cheng X, *et al.* Comparative genomic analysis of the PKS genes in five species and expression analysis in upland cotton [J]. *PeerJ*, 2017, 5: e3974.
- [43] Xu M, Shen C, Zhu Q, *et al.* Comparative metabolomic and transcriptomic analyses revealed the differential accumulation of secondary metabolites during the ripening process of acerola cherry (*Malpighia emarginata*) fruit [J]. *J Sci Food Agr*, 2022, 102(4): 1488-1497.

[责任编辑 时圣明]