

• 药材与资源 •

荆芥 1-脱氧-D-木酮糖-5-磷酸合成酶基因克隆及生物信息学分析

林谷音^{1,2}, 周佩娜¹, 尹梦娇^{1,2}, 刘莉成^{1,2}, 戴仕林^{1,2}, 刘潺潺^{1,2*}, 吴启南^{1,2,3*}

1. 南京中医药大学药学院, 江苏 南京 210023

2. 江苏省中药资源产业化过程协同创新中心, 江苏 南京 210023

3. 中药资源产业化与方剂创新药物国家地方联合工程研究中心, 江苏 南京 210023

摘 要: 目的 克隆荆芥 1-脱氧-D-木酮糖-5-磷酸合成酶 (1-deoxy-D-xylulose 5-phosphate synthase, *StDXS*) 基因, 并进行生物信息学分析。方法 根据荆芥转录组数据获得的 *StDXS* 基因序列设计特异性引物, 通过 RT-PCR 技术获得 *StDXS* 基因 cDNA 的全长序列, 并进行生物信息学分析。结果 荆芥 *StDXS* 基因全长 2177 bp, 包含一个长度为 2157 bp 的开放阅读框, 编码 718 个氨基酸, 其理论相对分子质量为 77 240, 等电点为 6.32, 定位于叶绿体, 不存在跨膜区及信号肽, 为非分泌蛋白。同源系统进化树分析表明, 该序列与同科植物鼠尾草的 *DXS* 基因进化关系较近, 均属于 *DXS1* 亚家族。密码子偏性分析表明, 该基因偏好使用以 A/T 结尾的密码子, 具有 28 个偏性密码子, 烟草为该基因最适合的外源表达宿主。结论 成功克隆荆芥 *StDXS* 基因并进行生物信息学分析, 为从分子水平调控荆芥的生长发育和改善药材产量和品质提供理论基础。

关键词: 荆芥; 1-脱氧-D-木酮糖-5-磷酸合成酶; 基因克隆; 生物信息学分析; RT-PCR

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2021)02-0527-11

DOI: 10.7501/j.issn.0253-2670.2021.02.026

Cloning and bioinformatics analysis of 1-deoxy-D-xylulose 5-phosphate synthase *DXS* gene from *Schizonepeta tenuifolia*LIN Gu-yin^{1,2}, ZHOU Pei-na¹, YIN Meng-jiao^{1,2}, LIU Li-cheng^{1,2}, DAI Shi-lin^{1,2}, LIU Chan-chan^{1,2}, WU Qi-nan^{1,2,3}

1. College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China

2. Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Nanjing 210023, China

3. National and Local Collaborative Engineering Center of Chinese Medicinal Resources Industrialization and Formulae Innovative Medicine, Nanjing 210023, China

Abstract: Objective To clone 1-deoxy-D-xylulose 5-phosphate synthase (*StDXS*) gene from *Schizonepeta tenuifolia* and perform the bioinformatics analysis. **Methods** The full-length cDNA sequence of *StDXS* gene was cloned by using RT-PCR based on transcriptome sequencing data of *S. tenuifolia* generated in the previous study and further analyzed by bioinformatic methods. **Results** The cloned *StDXS* gene was 2 177 bp, containing a 2 157 bp open reading frame (ORF) which encoded 718 amino acids. The theoretical molecular weight was 77 240 and its isoelectric point was 6.32. *StDXS* protein might be located in the chloroplast, and it had no signal peptide and membrane spanning domain, it was non-secretory protein. Phylogenetic analysis indicated that sequence of amino acids was closely related to the evolution of *Salvia officinalis* of the same family, and they both belonged to *DXS1* family. Codon bias analysis showed that *StDXS* gene preferred to use A/T ending codon, with 28 skewed codons. *Nicotiana tabacum* was the most suitable host for exogenous expression of *StDXS* gene. **Conclusion** The *StDXS* gene was cloned from *S. tenuifolia* successfully and its bioinformatics analysis was performed, which provided a theoretical basis for regulating the growth and development of *S. tenuifolia* and improving the yield and quality of *S. tenuifolia*.

Key words: *Schizonepeta tenuifolia* Briq.; 1-deoxy-D-xylulose 5-phosphate synthase; gene cloning; bioinformatics analysis; RT-PCR

收稿日期: 2020-03-06

基金项目: 国家自然科学基金面上项目 (81973435); 国家自然科学基金面上项目 (81473313); 国家自然科学基金青年科学基金项目 (81903756)

作者简介: 林谷音 (1996—), 女, 硕士研究生, 研究方向为中药学。E-mail: lgy_grace@sina.com

*通信作者: 刘潺潺, 女, 讲师, 主要从事中药资源生产与品质评价研究。E-mail: liuchanchan0317@outlook.com

吴启南, 男, 教授, 主要从事中药资源生产与品质评价研究。Tel: (025)85811521 E-mail: qnwyjs@163.com

中药荆芥为唇形科一年生草本植物荆芥 *Schizonepeta tenuifolia* Briq. 的干燥地上部分, 具有解表散风、透疹、消疮的功效。其茎、叶、花穗多个部位富含挥发油, 经提取可得到荆芥精油, 被广泛用于医药、化工、食品和香料等领域。荆芥精油主要含有单萜类成分, 也含有少量倍半萜和醛、酮、醌类等化合物, 其中薄荷酮 (menthone)、胡薄荷酮 (pulegone) 相对含量较高, 占总挥发油含量的 70% 以上^[1], 具有镇痛、抗炎以及体内抗病毒活性^[2,3]。

单萜类成分在植物体内主要通过 2-甲基-D-赤藓醇-4-磷酸途径 (MEP Pathway) 合成, 该途径以丙酮酸 (pyruvate) 和 3-磷酸-甘油醛 (glyceraldehyde-3-phosphate) 为原料, 经过多步酶促反应生成异戊烯焦磷酸 (isopentenyl pyrophosphate, IPP) 和其异构体二甲丙烯焦磷酸 (dimethylallyl pyrophosphate, DMAPP), 再进一步生成牻牛儿基焦磷酸 (geranyl pyrophosphate, GPP), 人们熟知的樟脑、柠檬烯、薄荷酮均是 GPP 的下游产物——单萜类成分^[4-5]。1-脱氧-D-木酮糖-5-磷酸合酶 (1-deoxy-D-Xylulose-5-phosphate synthase, DXS) 是 MEP 途径的限速酶之一, 是催化丙酮酸和 3-磷酸-甘油醛合成 1-脱氧-D-木酮糖-5-磷酸 (1-deoxy-D-xylulose 5-phosphate, DXP) 的关键酶^[6-7]。目前, DXS 基因已在厚朴^[8]、黄花蒿^[9]、滇龙胆^[10]、冬凌草^[11]等多种药用植物中被分离并鉴定, 发现该基因在调节植物生长发育、次生代谢产物的合成与积累以及自身保护和防御等方面具有重要作用。

迄今为止, 有关荆芥的研究多聚焦于植物形态学、化学成分以及药理作用方面, 从分子水平层面仅报道了荆芥单萜类成分生物合成相关的 *d*-柠檬烯合酶 (*d*-limonene synthase) 基因的分子克隆和酶促实验, 并推测了该酶的立体化学结构^[12], 有关荆芥单萜类成分生物合成分子机制的报道依然鲜有报道。因此, 本研究基于课题组前期荆芥转录组测序数据, 筛选得到注释为 DXS 的序列, 对其进行特异性引物设计和 cDNA 全长克隆及生物信息学分析。以期为进一步研究其功能, 探索荆芥单萜类成分生物合成机制奠定基础, 也为提升中药荆芥的药品品质提供理论指导。

1. 材料与试剂

1.1 材料

本实验植物材料荆芥由购买于河北的荆芥种子繁殖而得, 经课题组前期培养 30 d 得到荆芥新鲜植株, 经南京中医药大学吴啟南教授鉴定为唇形科荆芥属植物荆芥 *S. tenuifolia* Briq., 其新鲜叶片清水洗净后用滤纸吸干水分, 作为提取荆芥 RNA 的材料。

1.2 试剂

多糖多酚植物总 RNA 提取试剂盒、DH5 α 大肠杆菌感受态细胞购于北京天根生化科技有限公司; DNA 凝胶回收试剂盒购于北京擎科新业生物技术有限公司; PrimeScriptTM II 1st Strand cDNA Synthesis Kit、2 $\times$ Taq Master Mix、pMD18-T Vector Cloning Kit 购于大连宝生物科技有限公司; 2 $\times$ Easy Taq Mix 购于北京全式金生物技术有限公司; 引物合成和测序技术服务由生工生物工程 (上海) 股份有限公司提供。

2 方法

2.1 荆芥叶片总 RNA 的提取及 cDNA 的合成

将荆芥叶片样品置于液氮中研磨成细粉, 按照多糖多酚植物总 RNA 提取试剂盒说明书提取总 RNA, 按照 PrimeScriptTM II 1st Strand cDNA Synthesis Kit 反转录体系进行 cDNA 的合成。

2.2 荆芥 *SdDXS* 基因全长的克隆

根据课题组前期荆芥转录组数据获得 DXS 基因序列, 使用 Primer Premier 5.0 软件设计特异性引物, 上游引物 DXS-F 的序列为 5'-CACACAC-ACACAAGATGGCT-3', 下游引物 DXS-R 的序列为 5'-CCATCTAGGACATCATCTC-3'。扩增体系总体积 25 μ L, 包括 2 $\times$ Taq Master Mix 12.5 μ L, 叶 cDNA 模板 2.5 μ L, 上游和下游引物各 1.0 μ L, ddH₂O 9.0 μ L。PCR 反应条件为 94 $^{\circ}$ C 预变性 1.5 min; 94 $^{\circ}$ C 变性 30 s; 50 $^{\circ}$ C 退火 45 s; 72 $^{\circ}$ C 延伸 5 min, 共 35 个循环; 最后 72 $^{\circ}$ C 保持 10 min; 4 $^{\circ}$ C 永久。扩增产物经 1% 琼脂糖凝胶电泳检测后, 使用 DNA 凝胶回收试剂盒将目的条带切胶回收。用 TA-cloning Kit 及 pMD18-T 载体与目的基因连接, 将阳性克隆送上海生工有限公司进行测序。

2.3 *SdDXS* 基因序列的生物信息学分析

将测序后得到的 DXS 基因序列通过美国国立生物技术信息中心 (NCBI) 中的 Blast 对比。利

用拟南芥在线数据库 TAIR-BLAST 寻找拟南芥中的同源蛋白；使用 NCBI 中的 Conserved Domain 预测蛋白保守域及超家族；使用蛋白结构和功能预测软件对 *StDXS* 编码的蛋白的理化性质 (ProtParam)、结构域 (Inter Pro Scan)、信号肽 (SignalP 4.1 Server)、亚细胞定位 (WoLF-PSORT II)、跨膜结构 (Tmhmm Server v.2.0)、糖基化、磷酸化位点 (NetNGlyc 1.0 Server、NetPhos 3.1

Server)、蛋白亲疏水性 (ProtScale)、蛋白结构 (Sopma, Swiss-Model) 等进行综合生物信息学分析。使用 DNAMAN 软件进行氨基酸序列比对，使用 MEME 在线软件进行氨基酸保守基序分析，利用 MEGA-X 10.0 的邻接法构建系统发育树 (Neighbor-Joining, NJ)，最后使用 CondoW 和 EMBOSS 软件对该基因的密码子偏好性进行分析。所用生物信息学在线程序见表 1。

表 1 生物信息学分析在线程序及网址

Table 1 Bioinformatics online programs and web sites

在线软件	网址
NCBI	https://www.ncbi.nlm.nih.gov/
ORF Finder	https://www.ncbi.nlm.nih.gov/orffinder/
TAIR-BLAST	https://www.arabidopsis.org/Blast/index.jsp
EMBOSS-cusp	http://www.bioinformatics.nl/cgi-bin/emboss/cusp
ProtParam	https://web.expasy.org/protparam/
Inter Pro Scan	https://www.ebi.ac.uk/interpro/search/sequence/
SignalP 4.1 Server	http://www.cbs.dtu.dk/services/SignalP-4.1/
WoLF-PSORT II	https://www.genscript.com/wolf-psort.html
TMHMM Server v. 2.0	http://www.cbs.dtu.dk/services/TMHMM/
NetNGlyc 1.0 Server	http://www.cbs.dtu.dk/services/NetNGlyc/
NetPhos 3.1 Server	http://www.cbs.dtu.dk/services/NetPhos/
ProtScale	https://web.expasy.org/protscale/
SOPMA	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html
SWISS-MODEL	https://swissmodel.expasy.org
MEME	http://meme-suite.org/tools/meme

3 结果与分析

3.1 荆芥 *StDXS* 基因 cDNA 克隆及序列分析

本实验提取得到的荆芥总 RNA 的 28S 和 18S 条带清晰，无弥散拖尾等现象，说明总 RNA 质量较好。以总 RNA 逆转录得到的 cDNA 为模板，利用特异性引物 DXS-F、DXS-R 进行全长扩增。扩增产物经琼脂糖凝胶电泳检测，发现在 2000 bp 左右出现特异片段，与该基因片段预期大小基本一致 (图 1)。

将 PCR 产物进行 TA 克隆测序，发现荆芥 *DXS* 基因序列全长为 2177 bp，命名为 *StDXS*。该序列经 BLAST 搜索比对后发现与唇形科植物丹参 *Salvia miltiorrhiza* Bunge. 的 *DXS* 基因序列相似性最高为 99%，说明该序列扩增正确。利用 NCBI 的 ORF Finder 查询功能对该序列进行分析，发现 *DXS* 基因包含一个长为 2157 bp 的开放阅读框 (open reading frame, ORF)，编码 718 个氨基酸残基 (图 2)。

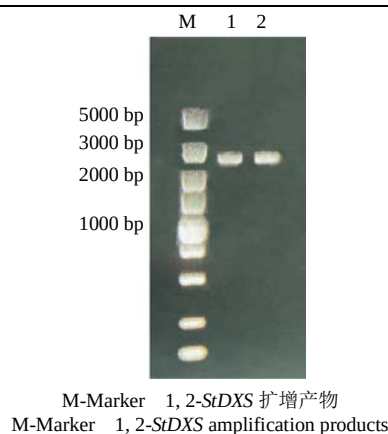


图 1 荆芥 *DXS* 基因扩增结果

Fig. 1 Amplification result of *StDXS* gene of *S. tenuifolia*

3.2 荆芥 *StDXS* 基因序列编码蛋白的生物信息学分析

利用 NCBI 中的 Conserved Domain 搜索后发现 *StDXS* 基因编码的蛋白质具有 PLN02582 活性位点 (45~718)，属于 PLN02582 超家族。利用 Inter Pro Scan 搜索发现 *StDXS* 蛋白质属于脱氧木酮糖磷酸

```

1   ATGGCTTTATGTCCATTTCATTTCTGGGAATTTAAGCAAAGGATTAGTAGCAGCAGATGCCAAAAACACCAAAATTTTGTCTCAA
1   M A L C P F A F S G N L S K G L V A A D A Q K H T N F C S Q
91  TGGCTTCATGGTCCAGATCTACCTTTCATCCCTTCTGCAAGAACAACAGGTAAGGAAAAGCTCAACAGGAATTTGTGCAACATTATCA
31  W L H G P D L P F H P F C K N N Q V R K S S T G I C A T L S
181 GAAAGAGGGGAATATTCTCAGAAAAGCTCCAACTCCCTTCTAGACACCATCAACTATCCGATTACATGAAAAACCTATCTACTAAG
61  E R G E Y F S Q K P P T P L L D T I N Y P I H M K N L S T K
271 GAAGTGAACAACCTTGCTGACGAACCTCCGGTCTGACGTCATCTTCAACGTGTCCAAGACTGGGGTTCATTGGGATCAAGCCTTGCTGTG
91  E L Q Q L A D E L R S D V I F N V S K T G G H L G S S L G V
361 ATTGAGCTAACTGTGGCTCTTCAATTATGTGTTCACTGCTCTCAAGATCGAATCTGTGGGATGTGGGACACAGGCTTATCCACACAAG
121 I E L T V A L H Y V F T A P Q D R I L W D V G H Q A Y P H K
451 ATTCTGACAGGAAGAAGAGACAAGATGCCGAGTTTAAAGACAGACCGGTGGTCTCTCCGGTTTACTAAGCGGTCTGAGAGCGACTATGAC
151 I L T G R R D K M P S L R Q T G G L S G F T K R S E S D Y D
541 TGCTTTGAGCCGGTCAAGCTCCACAACCATCTCTGAGGATAGAAATGGCTGTGGGAGAGATCTGAAAGGAAGAACCAACCAAGCTC
181 C F G A G H S S T T I S A G L E M A V G R D L K G R T N H V
631 GTGGCTGTATAGGAGACGGGGCTATGACAGCCGGTCAAGGCTTATGAAGCAATGAACAATGCGGGCTACCTCGACTCAGACATGATTGT
211 V A V I G D G A M T A G Q A Y E A M N N A G Y L D S D M I V
721 ATTCTCAATGACAATAAGCAAGTTTCTTACCCACTGCTAACCTGGATGGACCAACTGCTCCAGTGGGAGCCTTGAGCAGTGCCTTGAGT
241 I L N D N K Q V S L P T A N L D G P T A P V G A L S S A L S
811 AGGTGACAGTCGAACCGCCACTCAGAGACTAAGAGAAGTTGCAAGGAGTCAACCAAGCAGATTGGAGGCCCTACCGCATGAGCTTGCT
271 R L Q S N R P L R E L R E V A K G V T K Q I G G P T H E L A
901 GCAAAAGTCGATGAATATGCTCGTGGGCTGATCAGTGGTTCTGGATCAACGCTCTTTGAAGAGCTCGGACTTTATTACATTGGGCGAGT
301 A K V D E Y A R G L I S G S G S T L F E E L G L Y Y I G P V
991 GATGGTCACAATCTTGATGATCTCAGCAATTTCTAGAGAGGTAAGAGTACTAAAAACAAGGTCCTGGTGTGATCCATGTTGTGACT
331 D G H N L D D L T A I L R E V K S T K T T G P V L I H V V T
1081 GAGAAAGCAGGGGATATCCTTATGACAGAGAAAGCTGACAGATAAATATCATGGAGTGACCAAGTTCGATCCAGCAAGTGAAGCAGTTT
361 E K G R G Y P Y A E K A A D K Y H G V T K F D P A T G K Q F
1171 AAATCAAGTGCTCAACTCAGTCTTATACAACTACTTTGACAGAGGCTCTGATTGACAGAGCTGAAGAGGATAAGGATATTGTAGCAATC
391 K S S A P T Q S Y T T Y F A E A L I A E A E E D K D I V A I
1261 CATGCGCGATGGAGGTGGGACGGGTTGAACCTCTTCAACGTCGTTTCCCAACAAGGTGTTTCGATGTTGGGATAGCAGAGCAGCAT
421 H A A M G G G T T G L N L F Q R R F P T R C F D V G I A E Q H
1351 GCTGTAACCTTTGCTGACGGTTTGGCTGTGAAGGAATCGAACCTTCTGTGCAATCTACTCATCTCTTGCAGAGAGCTTACGACACAG
451 A V T F A A G L A C E G I E P F C A I Y S S F L Q R A Y D Q
1441 GTAGTGCAGATGTCGATTGACAGAAGTTGCCGGTATAGGTTTGTATGATAGAGCTGGCTTAGTGGGAGCAGACGCCGACACATTGT
481 V V H D V D L Q K L P V R F A M D R A G L V G A D G P T H C
1531 GGGGCGTTTGATGTTACTTTTATGGCTTGCTTCCCTAACATGGTGGTGGTGGACCTTCGGATGAGGCCGAATATTTCACATGGTTGCA
511 G A F D V T F M A C L P N M V V M A P S D E A E L F H M V A
1621 ACTGCTGCGCCATAGATGACAGACCAAGCTGCTCCGTTATCCGAGAGGTAATGGTATAGGTGAGGTGGCAGCTTCCACCTGGAAATAAAGGC
541 T A A A I D D R P S C F R Y P R G N G I G V E L P P G N K G
1711 AAACCTCTAGAGATCGGCAAGGCCGTATACTAATGAAGGGGAGAGTGGCTCTCTTGGGTTATGGGTGAGAGTTCAGAGCTGTTTG
571 K P L E I G K G R I L I E G E R V A L L G Y G S A V Q S C L
1801 GCTGACGCTGATGGTAGAAACCCGTTGTTTACGGTTGACAGTAGCTGATGCTCGTTTCTGCAAGCCTTTGGACCATGCCCTTATACGA
601 A A A A L V E T R G L R L T V A D A R F C K P L D H A L I R
1891 AGTTTGGCCAAATCACATGAGGTGTGATCACTGTGAAGAAGGGTCAATTGGTGGTTTGGATCTCATGTAGCTCAGTTCATGGCCTTG
631 S L A K S H E V L I T V E E G S I G G F G S H V A Q F M A L
1981 GATGGGCTTCTGATGCCAAATTAAGTGGAGACCATGGTTCTTCTGATCGGTATATCGATCATGGAGCCCGAGTCAATCAATAATG
661 D G L L D G K L K W R P L V L P D R Y I D H G A P V N Q I M
2071 GAAGCAGGACTACGCCCTTCTCATATTGGCGCAACTGTTTAAATATACTAGGAAAGCTAGGAGGCTCTGGAGATGATGCTCTAG
691 E A G L T P S H I A A T V F N I L G K A R E A L E M M S *

```

*为终止密码子；阴影部分为转酮酶结构域；下划线部分为 *DXS* 基因特征保守序列“DRAG”

*represents the Stop Codon; transketolase domain was marked by the shadow; the *DXS* signature sequence (DRAG) is underscored

图 2 荆芥 *DXS* 蛋白质编码区 (CDS) 及其氨基酸序列

Fig. 2 Coding sequence region (CDS) of *S. tenuifolia* and its amino acid sequence

合酶 (deoxyxylulose-5-phosphate synthase, IPR005477) 家族, 该家族主要参与萜类化合物的生物合成。*StDXS* 含有 3 个结构域, 分别为转酮醇酶 N 端结构域 (transketolase, N-terminal, 110~119), 类转酮醇酶嘧啶结合结构域 (transket-pyr, 397~562) 和转酮醇酶结构域 (transketolase_C, 576~699)。此外 *StDXS* 还具有 *DXS* 家族基因特征保守序列“DRAG”, 进一步证明基因扩增的正

确性。利用在线软件 ProtParam 对 *StDXS* 基因编码的氨基酸序列进行分析, 发现其编码 718 个氨基酸残基, 分子式为 $C_{3424}H_{5406}N_{952}O_{1023}S_{30}$, 相对分子质量为 77 240, 等电点为 6.32, 其半衰期较长, 为 30 h, 不稳定系数为 40.00, 属于稳定蛋白, 脂肪族氨基酸指数为 87.14, 亲水性总平均值为 -0.089, 属于亲水性蛋白。在 *StDXS* 基因编码的氨基酸分布中, 丙氨酸 (Ala) 含量最高为 10.6%, 其

次是亮氨酸 (Leu) 为 10.0%，半胱氨酸 (Cys) 和色氨酸 (Trp) 最少为 1.8% 和 0.4%，其中负电荷残基 (Asp+Glu) 有 77 个，正电荷残基 (Arg+Lys) 有 69 个。

利用 WoLF-PSORT II 在线软件进行 StDXS 蛋白的亚细胞定位预测，结果表明该蛋白定位于叶绿体、细胞核、胞质的 *K* 值分别为 8、3、3，说明 StDXS 最可能定位于叶绿体；利用 TMHMM Server v.2.0 在线软件预测 StDXS 蛋白的跨膜结构域，发现该蛋白无跨膜结构域，说明其不属于跨膜蛋白 (图 3-A)；利用 NetNGlyc 1.0 Server 在线软件预测 StDXS 蛋白的糖基化位点，发现该蛋白有 2 个糖基化位点 (11, 106) (图 3-B)；利用 NetPhos 3.1 Server 预测 StDXS 蛋白的磷酸化位点和蛋白激酶结合位点 (图 3-C)，发现该蛋白共有磷酸化位点 60 个，其中丝氨酸 (S) 28 个，苏氨酸 (T) 23 个，酪氨酸 (Y) 9 个，并存在蛋白激酶 C (PKC)、蛋白激酶 A (PKA)、络蛋白激酶 II (CKII)、细胞周期蛋白依赖性激酶抑制因子 (CKI)、细胞分裂周期

激酶 2 (CDC2)、P38 信号调节蛋白激酶 (p38MAPK)、unsp 等多种特异性蛋白激酶结合位点。利用 SignalP 4.1 Server 在线软件预测 StDXS 蛋白的信号肽，发现该蛋白不含信号肽，可能为非分泌蛋白 (图 3-D)；利用 ExPasy-ProtScale 分析 StDXS 蛋白亲/疏水性 (图 3-E)，发现疏水性最强出现在第 602 个氨基酸 (A) 处最高值为 2.300；亲水性最强出现在第 176 个氨基酸 (E) 处最低值为 -2.500。从整条肽链的氨基酸预测值来看，其中大部分为负值，由此可以推断，荆芥 DXS 蛋白是一种亲水性蛋白，与蛋白质的理化性质预测一致。

3.3 荆芥 StDXS 基因序列编码蛋白的二级、三级结构预测

利用 SOPMA 对 StDXS 蛋白的二级结构进行分析，结果表明该蛋白二级结构主要以 α -螺旋 (H) 和无规卷曲 (C) 为主，分别为 40.25% 和 37.47%，还存在 14.48% 的延伸链 (E) 和 7.80% 的 β -转角 (T)。使用 Swiss-Model 在线软件预测 StDXS 蛋白的三级结构，并使用 Pymol 对其进行可视化处理及优化 (图 4)。

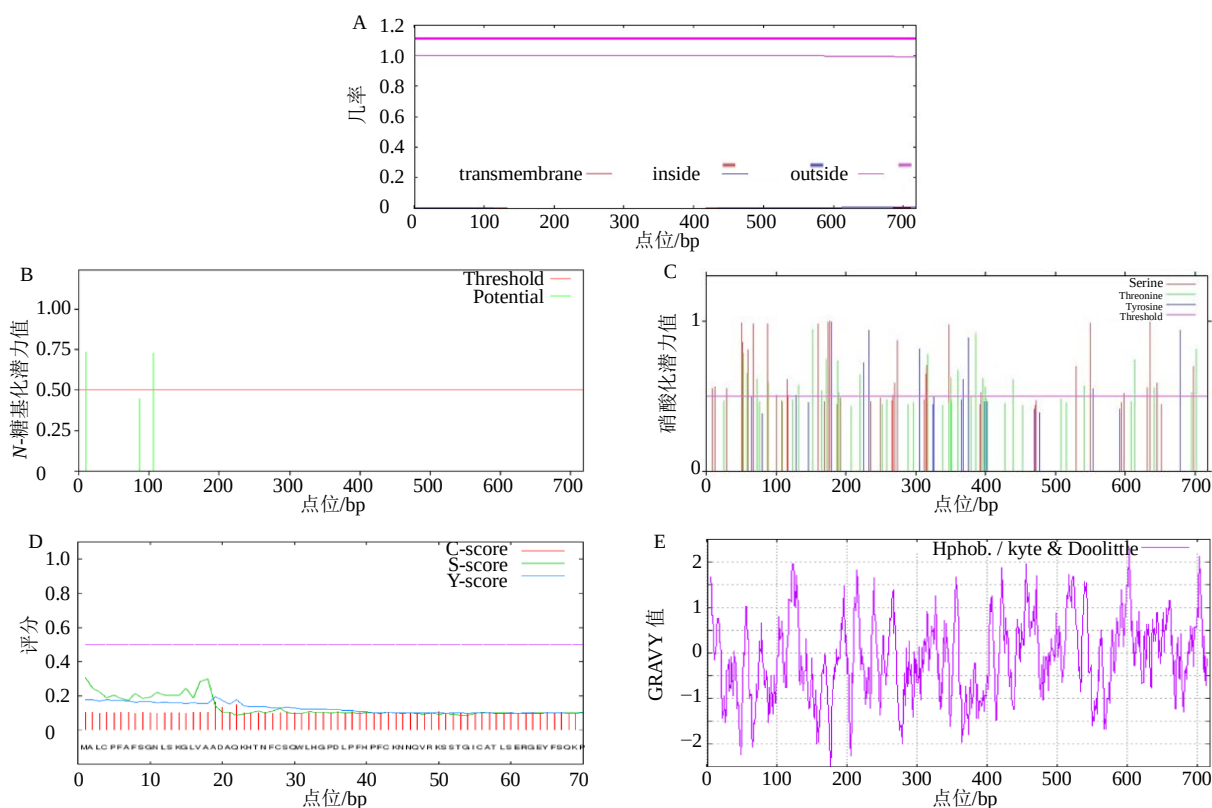


图 3 荆芥 StDXS 蛋白跨膜结构域 (A)、蛋白糖基化位点 (B)、蛋白磷酸化位点 (C)、蛋白信号肽 (D)、蛋白亲水性/疏水性 (E) 预测

Fig. 3 Prediction of potential protein transmembrane (A), glycosylation sites of protein (B), protein phosphorylation site (C), protein signaling peptide (D), and protein predicted hydrophilicity/hydrophobicity (E) of DXS protein in *S. tenuifolia*

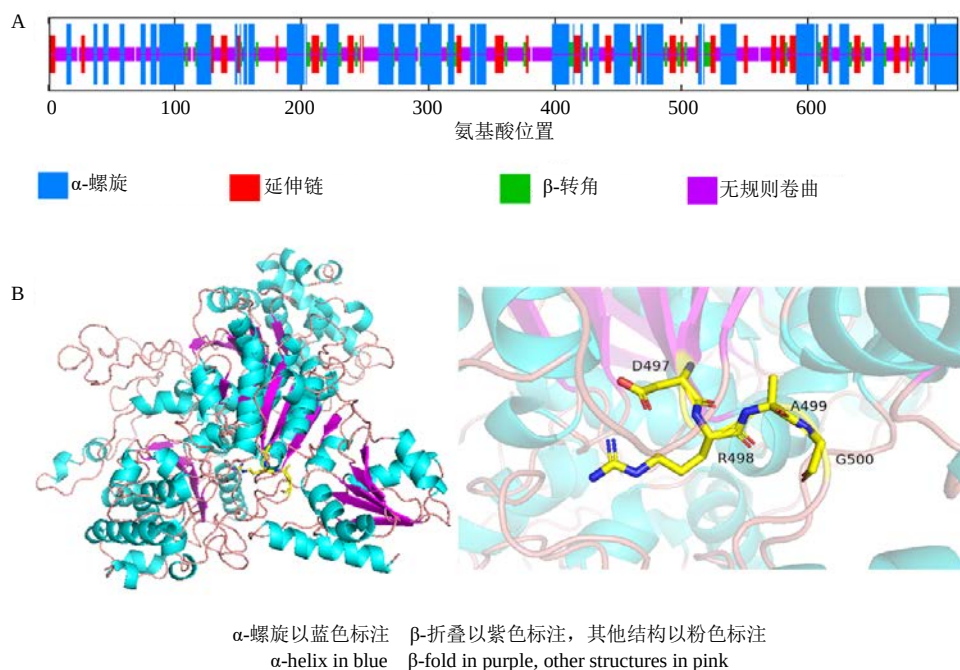


图 4 荆芥 StDXS 蛋白的二级结构 (A)、三级结构及其保守结构域 (B) 预测

Fig. 4 Prediction of secondary structure (A), tertiary structure and its conserved domain (B) of StDXS protein in *S. tenuifolia*

3.4 荆芥 StDXS 蛋白同源性及保守基序分析系统进化分析

利用 NCBI 中的 BLASTp 对 StDXS 基因编码的氨基酸与其他植物 DXS 蛋白的氨基酸序列进行同源比对, 发现其与丹参 (SmDXS, ACF21004.1)、冬凌草 *Isodon rubescens* (Hemsl.) Hara (IrDXS, AMM72794.1)、胡椒薄荷 *Mentha × piperita* L. (MpDXS, AAC33513.1)、蓖麻 *Ricinus communis* L. (RcDXS, XP_002533688.1)、番茄 *Solanum lycopersicum* L. (SlDXS3, XP_004245186.1) 具有较高相似性, 分别为 96.10%、72.02%、68.84%、59.17%、56.34%, 多序列分析表明荆芥、丹参、冬凌草和胡椒薄荷 DXS 氨基酸序列均含有 DXS 家族的高度保守的特殊结构域“DRAG”, 蓖麻与番茄 DXS 氨基酸序列含有特殊结构域“TSAG”(图 5)。利用 MEME 在线网站分析以上 6 种植物 DXS 氨基酸序列的保守基序 (motif), 在 E 值 <10 , P 值 <0.0001 条件下, 识别出 16 个共有的 motif, 长度为 50、43、41、29、21、15 个氨基酸各有 10、1、1、1、1、2 个, 这些 motif 可能是蛋白的功能位点 (图 6)。

3.5 荆芥 StDXS 基因系统进化分析

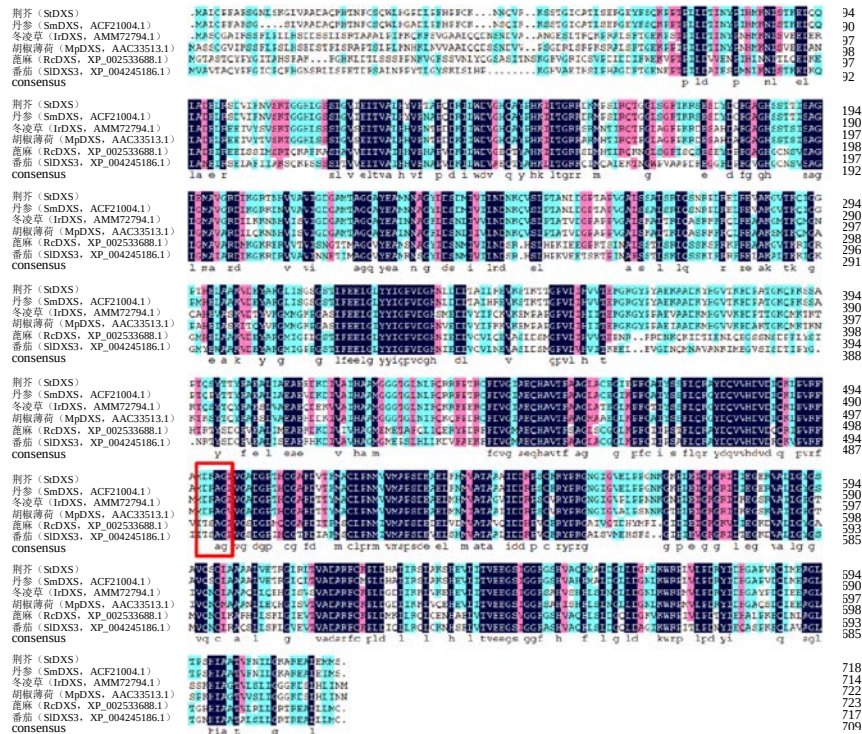
利用 MEGA-X 10.0 软件的 Neighbor-joining 模式建立 28 种植物的 StDXS 氨基酸序列系统进化树, 结果表明荆芥与同科植物鼠尾草 *Salvia*

officinalis L. 亲缘关系较高, 聚于一支 (图 7)。系统进化聚类显示植物 DXS 基因分为 3 个亚家族 DXS1、DXS2、DXS3, 与文献报道一致^[9,13]。

3.6 荆芥 StDXS 基因密码子偏好性分析

利用 CodonW 和 EMBOSS 软件分析荆芥 StDXS 基因的密码子适应指数 (codon adaptation index, CAI)、有效密码子数 (effective number of codons, ENC)、不同位置的 GC 含量 (GC1s、GC2s、GC3s)、总 GC 含量、同义密码子相对使用度 (relative synonymous codon usage, RSCU)。结果表明 CAI 为 0.3176、ENC 为 55.083, 说明密码子偏性较弱; 总 GC 值为 0.474 $7 < 0.5$, GC1s、GC2s 和 GC3s 分别为 0.5688、0.4451 和 0.4103, 说明该基因编码区的 A/T 含量高于 G/C 含量; 当 RSCU=1 时, 密码子没有偏好性; 当 RSCU>1 时, 密码子具有偏好性, 表 2 中 RSCU>1 的密码子共有 28 个, 其中 19 个以 A/T 结尾, 9 个以 G/C 结尾, 说明荆芥 StDXS 基因偏好于使用 A 或 T 结尾的密码子。

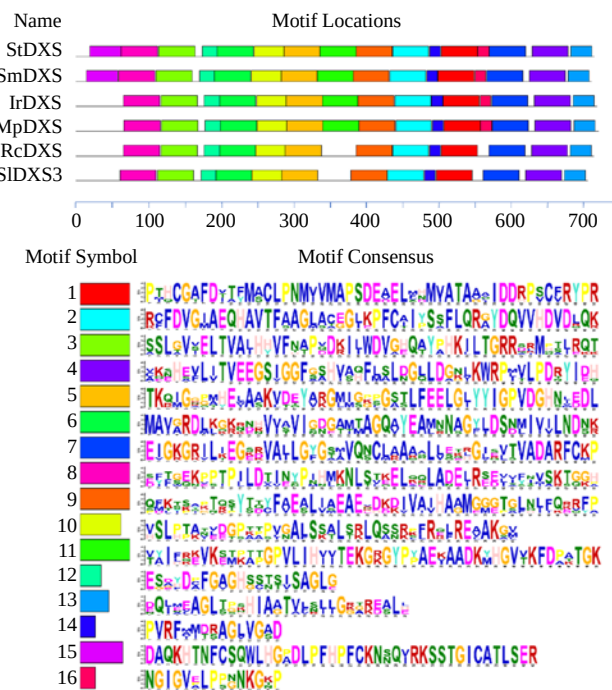
利用在线软件 EMBOSS-CHIP 分析荆芥 StDXS 基因密码子的使用频率, 在数据库 CodonUsage Database 查找拟南芥 *Arabidopsis thaliana* (L.) Heynh.、烟草 *Nicotiana glauca* L.、酵母 *Saccharomyces cerevisiae*、大肠杆菌 *Escherichia coli* 基因组的密码子偏好性, 荆芥的 StDXS 基因与拟南芥、烟草、酵母、大肠杆菌各密码子使用频



以相似性 50% 为阈值, 大于等于 50% 以蓝色标注, 大于等于 75% 以粉色标注, 100% 以深蓝色标注, 红框部分表示特殊结构域“DRAG”与“TSAG”
Similarity is set to 50% as threshold, greater than or equal to 50% is highlighted in blue, greater than or equal to 75% is in pink, and 100% is in dark blue, the red square indicates the “DRAG” and the “TSAG” special domain

图 5 荆芥 DXS 氨基酸序列与其他植物的多重比对

Fig. 5 Comparison of DXS amino acid sequences between *S. tenuifolia* and other plants



不同的基序用不同颜色的方框表示

Different motifs are represented by different colored boxes

图 6 荆芥 DXS 氨基酸序列与其他植物的保守基序分析

Fig. 6 Analysis of conserved motifs in DXS amino acid sequences of *S. tenuifolia* and other plants

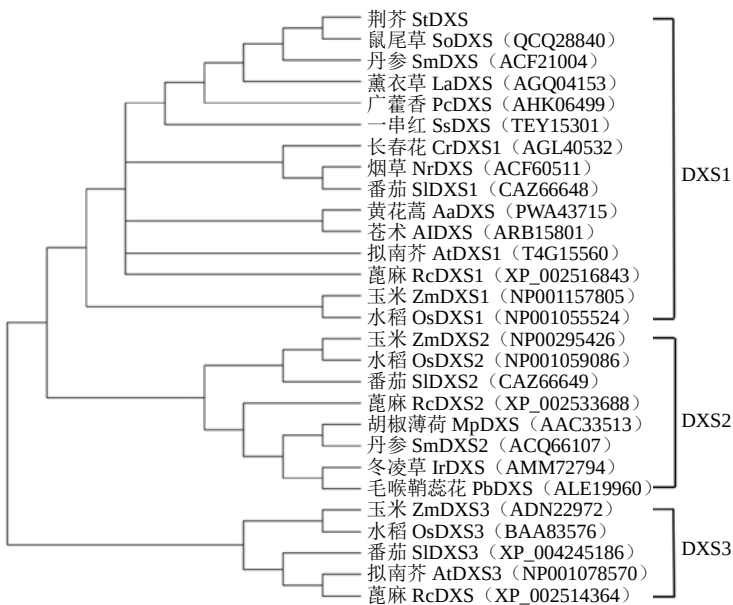


图 7 StDXS 氨基酸序列系统进化分析

Fig. 7 Phylogenetic analysis of StDXS amino acid sequence

表 2 荆芥 StDXS 基因中 RSCU>1 密码子

Table 2 RSCU > 1 of StDXS

氨基酸	密码子	数量	RSCU	氨基酸	密码子	数量	RSCU
苯丙氨酸	UUC	12	1.04	甘氨酸	GGU	20	1.23
亮氨酸	UUG	20	1.82		GGA	22	1.35
	CUU	12	1.09	组氨酸	CAU	13	1.24
	CUC	12	1.09	谷氨酰胺	CAG	12	1.20
异亮氨酸	AUU	14	1.27	天冬酰胺	AAC	9	1.06
缬氨酸	GUU	13	1.16	赖氨酸	AAG	16	1.07
	GUG	19	1.69	天冬氨酸	GAU	24	1.26
丝氨酸	UCU	8	1.26	谷氨酸	GAG	18	1.03
	UCA	9	1.42	半胱氨酸	UGU	5	1.11
	AGU	7	1.11	苏氨酸	ACU	17	1.74
精氨酸	AGA	14	2.55	脯氨酸	CCU	12	1.45
	CGU	7	1.27		CCA	13	1.58
丙氨酸	GCU	32	1.83	酪氨酸	UAU	14	1.47
	GCA	24	1.37	终止密码子	UAG	1	3.00

率的比值分别用 St/At、St/Ns、St/Sc、St/Ec 表示 (表 3)。若 2 物种间的密码子使用偏好性差异较小, 该比值常在 0.5~2.0, 不在此范围内则差异较大, 差异较大的密码子个数越少则说明 2 种物种密码子偏好性相似, 有利于基因的异源表达。荆芥 StDXS 基因与大肠杆菌、酵母、烟草、拟南芥偏好性差异较大的密码子个数依次为 16、15、7、9, 说明烟草最适合该基因的异源表达。

4 讨论

芳香类药用植物如荆芥、薄荷、紫苏等因含挥发性成分而具有特殊的气味, 药材记载常具有解表、化湿、行气、开窍等功能, 现代药理学表明此类挥发性成分具有抑菌、抗病毒、抗炎、抗肿瘤、抗氧化等药理活性, 在治疗发烧、炎症、病毒性感冒等疾病上具有显著疗效^[14-18]。而萜类化合物是挥发性成分发挥药理活性的主要物质基础。荆芥挥发

表 3 荆芥 *StDXS* 基因与大肠杆菌、酵母、烟草、拟南芥基因组密码子使用频率比较

Table 3 Comparison of codon usage frequency of *StDXS* gene among genome of *Escherichia coli*, *Saccharomyces cerevisiae*, *Nicotiana sylvestris* and *Arabidopsis thaliana*

氨基酸	密码子	密码子偏好的比值				氨基酸	密码子	密码子偏好的比值			
		St/Ec	St/Sc	St/Ns	St/At			St/Ec	St/Sc	St/Ns	St/At
丙氨酸	GCA	1.693	2.404	1.701	2.225	脯氨酸	CCA	2.293	1.140	0.888	1.296
	GCC	0.708	1.214	1.186	1.485		CCC	1.122	1.023	1.008	1.312
	GCG	0.264	0.897	1.184	0.618		CCG	0.480	1.312	1.449	0.809
	GCT	2.428	2.165	1.330	1.622		CCT	1.757	1.236	0.802	0.893
半胱氨酸	TGC	1.517	1.739	1.113	1.159	谷氨酰胺	CAA	0.966	0.509	0.647	0.717
	TGT	1.650	1.202	1.119	0.927		CAG	0.677	1.494	1.103	1.190
天冬氨酸	GAC	1.088	0.964	1.202	1.132	精氨酸	AGA	2.938	0.979	1.249	1.098
	GAT	1.073	0.962	1.016	0.988		AGG	2.086	0.907	0.605	0.759
谷氨酸	GAA	0.753	0.580	0.738	0.770		CGA	0.580	0.927	0.662	0.442
	GAG	1.290	1.304	0.897	0.777		CGC	0.000	0.000	0.000	0.000
苯丙氨酸	TTC	1.401	1.058	1.119	0.941	丝氨酸	CGG	0.880	4.091	1.480	1.419
	TTT	0.855	0.799	0.884	0.957		CGT	0.612	1.521	0.910	1.082
甘氨酸	GGA	2.454	3.062	1.274	1.379		AGC	0.778	1.135	1.135	0.985
	GGC	0.540	1.135	0.778	1.209		AGT	0.738	0.686	0.805	0.695
	GGG	1.922	3.941	2.111	2.318		TCA	1.168	0.818	1.048	0.836
	GGT	1.232	1.222	0.924	1.316		TCG	0.339	0.323	0.843	0.299
组氨酸	CAC	1.715	1.605	1.439	1.439	苏氨酸	TCT	1.062	0.592	0.635	0.552
	CAT	1.682	1.534	1.932	1.512		TCC	0.860	0.588	0.759	0.745
异亮氨酸	ATA	0.837	0.625	1.050	0.883		ACA	1.013	0.859	0.987	0.974
	ATC	0.789	0.889	1.101	0.827		ACC	0.589	0.876	1.050	1.080
	ATT	0.705	0.693	0.799	0.970		ACG	0.614	1.043	2.255	1.084
赖氨酸	AAA	0.636	0.564	0.833	0.768	缬氨酸	ACT	1.805	1.165	1.126	1.351
	AAG	1.545	0.768	0.730	0.723		GTA	0.955	1.061	1.128	1.264
亮氨酸	CTA	1.987	0.830	1.567	1.124		GTC	0.637	0.707	0.752	0.652
	CTC	1.757	3.091	1.403	1.037		GTG	1.328	2.447	1.519	1.519
	CTG	0.298	1.060	1.504	1.135		GTT	0.837	0.818	0.637	0.665
	CTT	1.247	1.470	0.712	0.750	色氨酸	TGG	0.311	0.401	0.331	0.334
	TTA	0.879	0.584	1.275	1.205		TAC	0.594	0.470	0.467	0.508
	TTG	2.156	1.023	1.194	1.331	酪氨酸	TAT	0.966	1.110	1.234	1.429
甲硫氨酸	ATG	0.998	1.131	0.977	0.965		TAA	0.000	0.000	0.000	0.000
天冬酰胺	AAC	0.754	0.617	0.905	0.732		TAG	4.637	2.782	2.782	2.782
	AAT	0.427	0.351	0.551	0.561		TGA	0.000	0.000	0.000	0.000

性成分中萜类物质占 70%，但目前荆芥精油的获得主要依赖于新鲜植物的水蒸气蒸馏提取，提取效率较低，且耗时较长，造成荆芥资源的浪费。有研究表明荆芥体内的胡薄荷酮和薄荷酮含量随荆芥的生长发育呈现规律性的变化^[19]，这与荆芥萜类成分生物合成途径中关键酶的表达调控有着必

然的联系。因此利用分子生物学和基因编辑手段提高荆芥体内萜类物质的含量可以极大地减少荆芥资源的消耗，使其充分利用，从而为荆芥药材资源的可持续发展提供保障。

本研究从荆芥中成功克隆了荆芥萜类化合物生物合成途径上的 *StDXS* 基因，该基因 cDNA 全长

2177 bp, 包含一个长度为 2157 bp 的开放阅读框。生信学分析表明 *StDXS* 基因编码 718 个氨基酸残基, 理论预测蛋白质相对分子质量为 77 240, 等电点为 6.32, 属于 PLN02582 超家族, 含有 3 个转酮醇酶结构域和 *DXS* 家族特征性保守序列“DRAG”。预测该基因的亚细胞定位于叶绿体, 编码一种非跨膜、无转运肽的亲水性蛋白。

在荆芥与其他 5 种植物的氨基酸序列保守基序分析中发现, 其保守基序的种类和位置非常相似, 同一亚组之间保守基序几乎相同, 不同亚组之间保守基序存在一定差异。系统进化树分析发现荆芥 *StDXS* 基因属于 *DXS1* 亚家族, 同科的冬凌草 *IrDXS* 和胡椒薄荷 *MpDXS* 基因属于 *DXS2* 亚家族, 而蓖麻 *RcDXS* 和番茄 *SIDXS3* 属于 *DXS3* 亚家族, *DXS3* 亚家族与其他 2 个亚家族进化关系较远, 它们的氨基酸序列存在一定差异, 从而造成了功能上的差异。有研究表明 *DXS* 基因的 3 个亚家族中, *DXS1* 亚家族具有看家基因的功能, 其所编码的酶催化前体物质形成叶绿素、类胡萝卜素等与初生代谢有关的化合物, 如拟南芥中的 *AtDXS1* 和黄花蒿中的 *AaDXS1*^[20]; *DXS2* 亚家族主要为一些特殊次生代谢物的合成提供异戊二烯骨架, 起到防御草食系动物、吸引昆虫授粉和抵御微生物等作用, 如可防御草食性动物取食的番茄 *SIDXS2*^[21] 和参与抗病反应的烟草 *NtDXS2*^[22]; *DXS3* 亚家族通常存在于植物基因组中, 目前研究较少, 可能主要参与了植物激素如 ABA (Absciscic acid 和 GA3 (Gibberellic acid) 的合成^[6,23]。因此, 推测属于 *DXS1* 家族的 *StDXS1* 基因较为保守, 很可能作为看家基因在初生代谢产物的合成与积累过程中发挥重要作用, 其在荆芥植物体内部的具体功能有待于进一步探究。另外, 在观察玉米、水稻、拟南芥等植物的 *DXS* 亚家族聚类结果时, 发现同种植物可能同时具有不同亚家族的 *DXS* 基因。由此推测在荆芥中, 除了在本研究中发现的属于 *DXS1* 亚家族的 *StDXS* 基因外, 极可能还具有与萜类等次生代谢产物生物合成相关的 *DXS2* 亚家族的基因序列。

通过使用大肠杆菌、酵母、烟草、拟南芥等模式生物来研究基因表达的技术已日趋成熟。密码子偏好性的研究有助于人们更好地利用异源表达系统, 克服异源表达的困难, 选择最合适的表达宿主。本研究发现荆芥 *StDXS* 基因偏好使用 A 或

T 结尾的密码子, 且编码区 A/T 含量高于 G/C 含量。分析大肠杆菌、酵母、烟草、拟南芥与荆芥 *StDXS* 基因的密码子使用偏好性, 发现荆芥与烟草具偏好性差异较大的密码子个数最少, 说明该基因更适合在烟草中表达。

本研究为今后 *StDXS* 蛋白功能的鉴定及定向优化荆芥药材品质提供了理论基础, 同时 *StDXS* 作为 MEP 途径中的第 1 个限速酶, 也为进一步探究荆芥萜类化合物生物合成的分子机制奠定基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 王凤, 温桃群, 桑文涛, 等. 荆芥挥发油化学成分及药理作用研究现状 [J]. 中南药学, 2017, 15(3): 312-318.
- [2] 刘玥欣. 荆芥不同煎煮时间对荆芥汤中胡薄荷酮含量及药效作用的影响 [D]. 长春: 长春中医药大学, 2018.
- [3] 何婷, 汤奇, 曾南, 等. 荆芥挥发油及其主要成分抗流感病毒作用与机制研究 [J]. 中国中药杂志, 2013, 38(11): 1772-1777.
- [4] Chen C, Zheng Y, Zhong Y, et al. Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora* [J]. *BMC Genomics*, 2018, 19(1): 550.
- [5] Turner G W, Croteau R. Organization of monoterpene biosynthesis in mentha. immunocytochemical localizations of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase [J]. *Plant Physiol*, 2004, 136(4): 4215-4227.
- [6] Rodríguez-Concepción M, Boronat A. Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis [J]. *Curr Opin Plant Biol*, 2015, 25: 17-22.
- [7] Wright L P, Rohwer J M, Ghirardo A, et al. Deoxyxylulose 5-phosphate synthase controls flux through the methylerythritol 4-phosphate pathway in *Arabidopsis* [J]. *Plant Physiol*, 2014, 165(4): 1488-1504.
- [8] 查良平, 杨健, 刘爽, 等. 厚朴 1-脱氧-D-木酮糖-5-磷酸合成酶基因的鉴定及挥发油成分分析 [J]. 中草药, 2015, 46(24): 3734-3739.
- [9] Zhang F Y, Liu W H, Xia J, et al. Molecular characterization of the 1-deoxy-D-xylulose 5-phosphate synthase gene family in *Artemisia annua* [J]. *Front Plant Sci*, 2018, 9: 952.
- [10] 张海晨, 李彩霞, 王元忠, 等. 滇龙胆 1-脱氧-D-木酮糖 5-磷酸合酶基因的克隆与表达分析 [J]. 生物技术通

- 报, 2016(4): 128-136.
- [11] 朱昀昊, 苏秀红, 董诚明, 等. 冬凌草 1-脱氧-D-木酮糖-5-磷酸合成酶基因克隆与表达分析 [J]. 广西植物, 2016, 36(12): 1476-1482.
- [12] Maruyama T, Ito M, Kiuchi F, *et al.* Molecular cloning, functional expression and characterization of d-limonene synthase from *Schizonepeta tenuifolia* [J]. *Biol Pharm Bull*, 2001, 24(4): 373-377.
- [13] 张浩宇, 樊俊苗, 王婷, 等. 植物萜类合成关键基因 DXS 研究进展 [J]. 生物技术通报, 2018(3): 1-8.
- [14] 张建红, 刘琬菁, 罗红梅. 药用植物萜类化合物活性研究进展 [J]. 世界科学技术—中医药现代化, 2018, 20(3): 419-430.
- [15] Shan M Q, Qian Y, Yu S, *et al.* Anti-inflammatory effect of volatile oil from *Schizonepeta tenuifolia* on carrageenin-induced pleurisy in rats and its application to study of appropriate harvesting time coupled with multi-attribute comprehensive index method [J]. *J Ethnopharmacol*, 2016, 194: 580-586.
- [16] Chen S G, Cheng M L, Chen K H, *et al.* Antiviral activities of *Schizonepeta tenuifolia* Briq. against Enterovirus 71 *in vitro* and *in vivo* [J]. *Sci Rep*, 2017, 7: 935.
- [17] Fung D, Lau C B S. *Schizonepeta tenuifolia*: Chemistry, pharmacology, and clinical applications [J]. *J Clin Pharmacol*, 2002, 42(1): 30-36.
- [18] 黄晓巍, 刘玥欣, 刘轶蕾, 等. 荆芥化学成分及药理作用研究进展 [J]. 吉林中医药, 2017, 37(8): 817-819.
- [19] Liu C C, Srividya N, Parrish A N, *et al.* Morphology of glandular trichomes of Japanese catnip (*Schizonepeta tenuifolia* Briquet) and developmental dynamics of their secretory activity [J]. *Phytochemistry*, 2018, 150: 23-30.
- [20] 刘万宏. 青蒿 DXS 基因家族功能分化及低温促进青蒿素合成的分子机制研究 [D]. 重庆: 重庆大学, 2016.
- [21] Paetzold H, Garms S, Bartram S, *et al.* The isogene 1-deoxy-D-xylulose 5-phosphate synthase 2 controls isoprenoid profiles, precursor pathway allocation, and density of tomato trichomes [J]. *Mol Plant*, 2010, 3(5): 904-916.
- [22] 罗朝鹏, 李泽锋, 张剑锋, 等. 烟草 NtDXS2 基因的克隆及表达分析 [J]. 烟草科技, 2018, 51(11): 1-7.
- [23] Cordoba E, Porta H, Arroyo A, *et al.* Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize [J]. *J Exp Bot*, 2011, 62(6): 2023-2038.

[责任编辑 时圣明]