

滇龙胆草中香叶醇 10-羟化酶基因的克隆、蛋白结构及表达分析

王迎夏^{1,2}, 赵赛静^{1,2}, 田维圣^{1,2},
崔晓雪^{1,2}, 黄文倩^{1,2}, 屠鹏飞¹, 史社坡^{1*}, 刘 晓^{1*}

¹北京中医药大学中药学院中药现代研究中心; ²北京中医药大学中药学院, 北京 100029

摘要:本研究基于滇龙胆草转录组数据分析, 使用 RT-PCR 技术克隆获得两条香叶醇 10-羟化酶基因, 命名为 *GrG10H1* 和 *GrG10H2*。使用在线软件对其进行了生物信息学分析, 两条基因全长分别为 1 548 bp、1 482 bp。系统进化分析显示该基因与细胞色素 P450 家族中 CYP76B 亚家族亲缘关系最近。通过对该蛋白进行三维结构预测、同源模建和分子对接, 分析了 GrG10H 蛋白活性口袋中可能参与催化功能发挥的关键氨基酸残基。荧光定量 PCR 表明 *GrG10H1* 基因在叶中表达水平最高, 茎次之, 根最低。而 *GrG10H2* 基因在根和叶中表达水平较高, 在茎中几乎不表达。本研究还建立了能够生产龙胆苦苷的愈伤组织细胞培养体系, 为进一步解析龙胆苦苷生物合成机制奠定了基础。

关键词:滇龙胆草; 龙胆苦苷; 生物信息学分析; 香叶醇 10-羟化酶; 基因克隆

中图分类号: R285

文献标识码: A

文章编号: 1001-6880(2024)9-1542-12

DOI: 10.16333/j.1001-6880.2024.9.010

Cloning, protein structure, and expression analysis of geraniol 10-hydroxylase gene from *Gentiana rigescens*

WANG Ying-xia^{1,2}, ZHAO Sai-jing^{1,2}, TIAN Wei-sheng^{1,2},
CUI Xiao-xue^{1,2}, HUANG Wen-qian^{1,2}, TU Peng-fei¹, SHI She-po^{1*}, LIU Xiao^{1*}

¹Modern Research Center for Traditional Chinese Medicine, Beijing Research Institute of Chinese Medicine, Beijing University of Chinese Medicine; ²School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100029, China

Abstract: In this study, two geraniol 10-hydroxylase (G10H) genes, namely *GrG10H1* and *GrG10H2*, were cloned from *Gentiana rigescens* through RT-PCR based on transcriptome data analysis. *In silico* bioinformatics analysis was performed towards these two genes whose open reading frame were 1 548 bp and 1 482 bp in length, respectively. Phylogenetic analysis revealed that the two genes are phylogenetically closed to the CYP76B subfamily of the cytochrome P450 enzymes. Homology modeling was employed to construct the three-dimensional structures of GrG10H1 and GrG10H2. Combined with molecular docking, the key residues in the substrate binding pocket of GrG10H1 and GrG10H2 protein were analyzed. Quantitative PCR demonstrated that the expression level of the *GrG10H1* gene was highest in leaves, followed by stems and lowest in roots. However, the expression level of the *GrG10H2* gene was high in roots and leaves but almost absent in stems. This study also established a callus cell culture system for *G. rigescens* to achieve stable production of gentiopicroside, which provides a foundation for further investigation into gentiopicroside biosynthesis mechanism.

Key words: *Gentiana rigescens*; gentiopicroside; bioinformatics analysis; geraniol 10-hydroxylase; gene cloning

收稿日期: 2024-01-26 接受日期: 2024-06-24

基金项目: 国家自然科学基金 (82173922, 81402809); 中央高校基本科研业务费专项 (2023-JYB-JBQN-054); 北京市自然科学基金 (7192112)

* 通信作者 Tel: 86-10-53911868; E-mail: felixiao@163.com, shishepo@163.com;

龙胆苦苷 (gentiopicroside) 是龙胆、秦艽、当药、獐牙菜等龙胆属植物的活性成分之一^[1]。从结构上来看, 龙胆苦苷属于裂环烯醚萜类化合物, 由环烯醚萜苷的母核环戊烷 C7-C8 键断裂衍生而成^[2]。龙胆苦苷具有抗炎镇痛^[3], 保肝^[4], 抗氧化, 抗病原

微生物等多种药理活性,临床应用价值广泛^[5]。研究表明龙胆苦苷主要在龙胆属植物的叶片中合成,经过茎,最后运输到根部储藏^[6]。目前从药用植物中直接提取龙胆苦苷的方法受到提取效率低,含量低、资源匮乏(龙胆作为 III 级保护植物物种收录于《国家重点保护野生药材物种名录》中)等多种因素限制,而化学合成法涉及到活性基团的保护与去保护,程序繁琐,不易大规模工业化制备,且环境不友好^[7]。相比之下,酶法合成的方式绿色高效且专一性强,为龙胆苦苷的快速获得提供了新的思路。因此,深入解析并挖掘龙胆苦苷生物合成途径中涉及到的相关基因并对其进行功能验证已成为了目前研究的热点^[8]。

作为一类特殊的单萜,龙胆苦苷的生物合成途径也遵循萜类化合物的一般合成规律,主要分为前体物质合成、萜类骨架形成以及后修饰三个阶段^[9]。第一阶段通过甲羟戊酸 (mevalonic acid, MVA) 途径和 2-甲基-赤藓醇-4-磷酸 (2-C-methyl-D-erythritol-4-phosphate, MEP) 途径生成前体物质二甲烯丙基焦磷酸 (dimethylallyl pyrophosphate, DMAPP) 和异戊烯基焦磷酸 (isopentenyl pyrophosphate, IPP)。第二阶段以牻牛儿基二磷酸 (Geranyl diphosphate, GPP) 为前体,经过脱磷酸化反应生成环烯醚萜类化合物生物合成途径的重要组分香叶醇 (geraniol),该化合物后续在多种酶例如香叶醇 10-羟化酶 (geraniol 10-hydroxylase, G10H)、香叶醇-10 羟基氧化还原酶 (10-hydroxygeraniol oxidoreductase, 10-HGO)、环

烯醚萜合酶 (iridoid synthase, IS)、7-脱氧马钱苷酸葡萄糖基转移酶 (7-deoxyloganetic acid glucosyl transferase, 7-DLGT)、7-脱氧番木鳖酸-7-羟化酶 (7-deoxyloganin 7-hydroxylase, 7-DLH) 及裂环番木鳖酸合酶 (secologanin synthase, SLS) 的作用下经过氧化、还原、甲基化、糖基化等反应生成关键中间体裂环番木鳖苷。目前,基于转录组测序技术的发展,龙胆苦苷生物合成途径的第一、二阶段已经相对清楚。但是经由裂环番木鳖苷生成龙胆苦苷第三阶段的相关酶和多个关键步骤还未被揭示。

在龙胆苦苷的生物合成途径中,香叶醇 (geraniol) 的氧化被认为是第一个关键限速步骤^[10] (见图 1)。此反应由香叶醇 10-羟化酶 (geraniol 10-hydroxylase, G10H) 催化香叶醇在 C-10 位羟基化,生成重要的中间产物 10-羟基香叶醇 (10-hydroxygeraniol)^[11,12]。该酶属于细胞色素 P450 单加氧酶 (cytochrome P450 monooxygenase) 家族中 CYP76B 亚家族的成员,在发挥催化作用时需 NADPH 的辅助^[13]。目前可从广藿香^[14]、长春花、川西獐芽菜、秦艽^[15]、蜘蛛香^[16] 等少数植物中克隆得到。已有的研究表明提高香叶醇 10-羟化酶基因在蜘蛛香^[16]、川西獐芽菜^[17] 等植物中的表达量可促进包括獐芽菜苦苷等环烯醚萜类成分的积累,除此之外,该基因在长春花萜类吲哚生物碱合成上游途径中也起着关键调控作用^[18]。因此,深入解析 G10H 发挥催化作用的机理有助于为阐明包括龙胆苦苷在内的环烯醚萜类化合物的生物合成途径解析奠定基础。

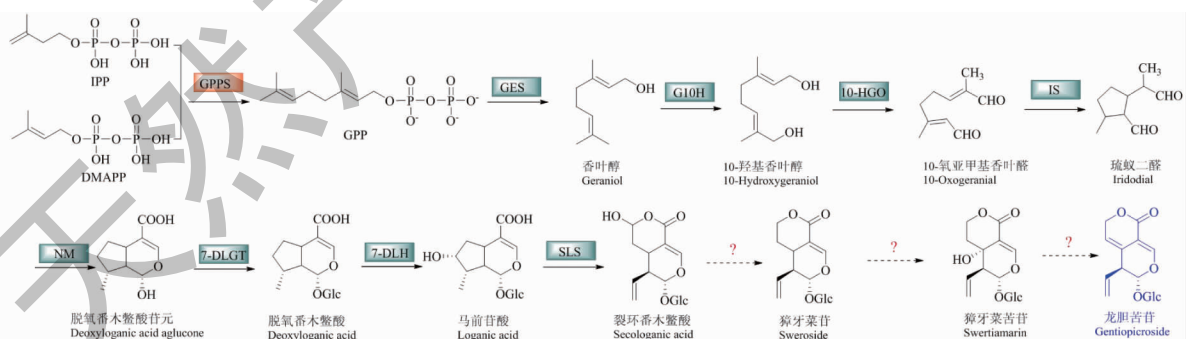


图 1 龙胆苦苷生物合成途径推测

Fig. 1 The hypothetical biosynthetic pathway of gentiopicroside

本研究对滇龙胆草进行了转录组测序和比较转录组学分析,利用 RT-PCR 技术从转录组数据库中克隆获得两条 G10H 基因,对其进行了生物信息学分析,并通过同源建模对蛋白进行了三维结构预测,

最后利用荧光定量 PCR 技术对克隆获得的两条 G10H 基因在滇龙胆草不同部位的表达模式进行了评价。此外,本研究还建立了能够生产龙胆苦苷的滇龙胆草愈伤组织细胞培养体系。研究结果为龙胆

苦苷合成途径关键酶分析与挖掘提供了参考。

1 材料与方法

1.1 植物材料

本实验所用材料滇龙胆草采自于云南省保山市,经由北京中医药大学中药现代研究中心刘晓副研究员鉴定为龙胆科龙胆属植物滇龙胆草(*Gentiana rigescens*)。选取滇龙胆草不同部位:根、茎、叶作为比较转录组测序的材料。滇龙胆草幼嫩茎作为愈伤组织诱导的材料。

1.2 菌种与载体

大肠杆菌菌株 *Fast1-T1* 及载体 pCE2-TA-Blunt-Zero 均购自北京诺维赞生物科技有限公司。

1.3 试剂

RNA 提取试剂盒: Plant RNA Kit (OMEGA, 批号: R6827010000L25W135); 反转录试剂盒: TransScript[®] One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen, 批号: AE311-02); 胶回收试剂盒: FastPure Gel DNA Extraction Mini Kit (Vazyme, 批号: DC301-01); 质粒提取试剂盒 FastPure Plasmid Mini Kit (Vazyme, 批号: DC201-01); 连接酶试剂盒: 5 min TM TA/Blunt Zero Cloning Kit (Vazyme, 批号: C601-02); PCR 扩增酶: Planta[®] Max Super-Fidelity (Vazyme, 批号: P505-d2); 2 × Rapid Taq Master Mix (Vazyme, 批号: P222-02)。

甲醇(色谱纯,批号:225397,北京环科智仪有限公司);乙腈(色谱纯,批号:222372,北京环科智仪有限公司);甲酸(色谱纯,批号:223168,北京环科智仪有限公司);MS 盐(批号:SLCM3143)、肌醇(批号:20210919)、蔗糖(批号:20240307)、KH₂PO₄(批号:K130BA0014)(分析纯,北京拜尔迪生物技术有限公司);SHESHIDO C₁₈ Column 分析型色谱柱(4.6 mm × 250 mm, 5 μm, 日本 SHESHIDO 公司)。

1.4 仪器

Agilent 1 260 高效液相色谱仪(型号: Agilent G1316A, 美国 Agilent 公司); PCR 仪(型号: Eppendorf Mastercycler[®] nexus, 艾本德中国有限公司); qPCR 仪(型号: BIO-RAD CFX Connect[™] Optics Module, 上海伯乐生命医学产品有限公司)

1.5 滇龙胆草不同部位化学成分分析

取清洗干净的滇龙胆草新鲜根、茎和叶片置于冻干机中冷冻干燥,将干燥后的不同部位的样品于研钵中充分研碎,准确称量 50 mg 粉末,按照质量(mg): 体积(mL) = 2.5: 1 的比例加入 70% 甲醇,30

°C 超声提取 40 min 后离心取上清液过 0.22 μm 滤膜进行成分分析及含量测定。使用 Agilent 1 260 HPLC 进行检测,色谱条件如下:柱温:30 °C,流动相 A:0.1% 甲酸水,流动相 B:乙腈,流速:1.0 mL/min,梯度洗脱程序为:0 ~ 20 min,5% → 15% B;20 ~ 50 min,15% → 30% B;50 ~ 60 min,30% → 50% B;60 ~ 70 min,50% → 95% B;70 ~ 75 min,95% B,进样体积为 8 μL,在 275 nm 紫外吸收波长下测定龙胆苦苷的峰面积,使用标准曲线法计算其含量。

1.6 滇龙胆草愈伤组织体系建立及诱导培养

选取滇龙胆草植株新鲜叶、嫩茎及根,洗净后放入广口瓶中,浸没于 75% 乙醇灭菌约 30 s 后迅速取出浸于无菌水中,洗涤 3 次;之后使用 2.5% NaClO 溶液振荡灭菌 15 min,用无菌水洗涤 3 次,将所得外植体在超净工作台中裁剪成约 2 ~ 4 cm 组织培养材料,将切面置于添加 2,4-二氯苯氧乙酸(2,4-dichlorophenoxyacetic acid, 2,4-D) 和 6-苄氨基嘌呤(6-benzylaminopurine, 6-BA) 两种植物激素的 MS 基本培养基(由 MS 盐、蔗糖、KH₂PO₄ 等组分配制而成)中,室温条件下避光培养,观察愈伤组织生成情况^[19]。

1.7 滇龙胆草不同部位比较转录组测序与 G10H 基因挖掘

按照 OMEGA Plant RNA Extraction Kit (OMEGA, USA) 试剂盒的使用说明,以新鲜滇龙胆草根、茎、叶为材料分别进行 RNA 的提取,每组设置 3 个平行。选取完整性较好(在琼脂糖凝胶电泳检测 28S rRNA 亮度明高于 18S rRNA), OD₂₆₀/OD₂₈₀ 值在 1.8 ~ 2.0 的 RNA 样本使用 DNBSEQ 平台进行转录组测序(交由北京六合华大基因科技有限公司完成),从测序产生的原始数据中过滤掉低质量、接头污染以及未知碱基 N 含量过高的 reads 后组装得到 Unigene,之后对 Unigene 进行七大功能数据库包括京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)、非冗余蛋白质数据库(Non-Redundant Protein Database, NR)、核酸序列数据库(Nucleotide Sequence Database, NT)、经过注释的蛋白质序列数据库(Swiss-Prot Protein Sequence Database, SwissProt)、GO 数据库(Gene Ontology, GO)、蛋白质家族数据库(The Protein Family Database, Pfam)、真核生物同源组(Eukaryotic Orthologous Groups, KOG)注释。使用转录本定量数据(RNA-Seq by expectation-maximization, RSEM)计算

各个样品的基因表达水平,用基因表达丰度数据 (fragments per kilobase of exon model per million mapped reads, FPKM) 对表达量数据用 z-score 法进行标准化处理。

利用已鉴定功能的植物来源的香叶醇 10-羟化酶的氨基酸序列为模板进行本地 Blast 搜索,结合转录组数据中基因注释和不同基因的表达丰度数据 FPKM 值,筛选可能参与龙胆苦苷生物合成的 G10H 候选基因。

1.8 *GrG10H1*、*GrG10H2* 基因的扩增

以滇龙胆草 RNA 为模板,依照反转录试剂盒 TransScript® One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen) 说明书合成第一链 cDNA。利用转录组数据分析挖掘出的 *GrG10H* 基因序列完整开放读码框信息设计两端特异引物(见表

1)。采用高保真酶 Planta® Max Super-Fidelity (Vazyme) 进行 DNA 扩增,PCR 条件为:95 °C 3 min;95 °C 30 s,56 °C 30 s,72 °C 30 s,扩增 35 个循环;72 °C 5 min。PCR 扩增产物通过 1.0% 琼脂糖凝胶电泳检测后,将目的片段利用 FastPure Gel DNA Extraction Mini Kit (Vazyme) 胶回收试剂盒进行产物纯化。PCR 回收产物与 pCE2 TA/Blunt-Zero 克隆载体进行连接。PCR 反应体系:PCR 纯化产物 60~110 ng,5 × TA/Blunt-Zero Cloning Mix 1 μL, ddH₂O 补足到 5 μL,轻弹混匀,37 °C 反应 5 min。反应结束后,将离心管置于冰上。取 5 μL 连接产物加入冰上融化的 *Fast-T1* 化学感受态细胞,涂布于含 Amp 抗性的固体 LB 平板,37 °C 培养过夜。利用 2 × Rapid Taq Master Mix 进行菌落 PCR 筛选阳性克隆,测序验证序列。

表 1 *GrG10H* 基因扩增所用引物

Table 1 Primers used for *GrG10H* gene amplification

序号 No.	引物作用 Primer usage	引物名称 Primer name	引物序列 Primer sequence(5'→3')
1	序列扩增	<i>GrG10H1</i> -F	ATGCTTGGCAGCAATAAATCCTCG
2	序列扩增	<i>GrG10H1</i> -R	TCATAGAGATGTTGGAATAGCAGC
3	序列扩增	<i>GrG10H2</i> -F	ATGGATTTTCTTACCATTGCAGTCGG
4	序列扩增	<i>GrG10H2</i> -R	TTACAAGGGAATGGGAACAACATG
5	qPCR	q <i>G10H1</i> -F	GCTCTTGGATTCTCTTCGC
6	qPCR	q <i>G10H1</i> -R	GCTAAGGATTTGTGAGGTTGG
7	qPCR	q <i>G10H2</i> -F	CCAAGATGAGTTAGCACAAAGTC
8	qPCR	q <i>G10H2</i> -R	CCAGGTGAGTGAATCGTAGAG
9	qPCR	<i>GrACTIN</i> -F	GCGGATCGTATGAGCAAGGA
10	qPCR	<i>GrACTIN</i> -R	GGGCGCGGATTCTCTCGTATT

1.9 *GrG10H1*、*GrG10H2* 基因生物信息学分析

利用 NCBI 中的在线 Blast 工具 (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) 对各基因编码的氨基酸序列进行同源性比较分析;利用 ExPASy 在线服务器 (<http://web.expasy.org/protparam/>) 预测基因编码蛋白的分子量、等电点、不稳定系数等特征;利用 SignalP-5.0 在线软件 (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>) 进行蛋白的信号肽预测,利用 TMHMM-2.0 在线软件 (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>) 预测蛋白的跨膜结构域;使用 InterPro 6.0 在线工具 (<https://www.ebi.ac.uk/interpro/>) 及 NCBI 中的在线工具 (<https://www.ncbi.nlm.nih.gov/Structure/>

[cdd/cdd.shtml](https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml)) 对 *GrG10H* 蛋白质进行归类 and 结构域预测;运用 Predict Protein (<https://predictprotein.org/>) 及 Softberry 软件 (<http://www.softberry.com/>) 预测蛋白的亚细胞定位;利用 ESript 3.0 软件对各基因编码氨基酸序列进行多重序列比对分析;利用 MEGA 11.0 软件中的 Clustal W 进行序列比对,Neighbor-joining 法进行聚类分析,重复次数为 1 000 次,将目的序列与已报道的其他对应功能序列进行比对并构建系统进化树。

1.10 *GrG10H1*、*GrG10H2* 蛋白三维结构预测及分子对接分析

利用在线分析软件 SWISS-MODEL 对 *GrG10H1* 和 *GrG10H2* 蛋白进行三维结构预测。利用

AutoDock Tools 1.5.6 软件对配体分子进行加氢以及加电荷处理。以已报道蛋白三维结构的高粱 (*Sorghum bicolor*) 中的肉桂酸 4-羟化酶 (PDB ID: 6VBY) 为模板, 设置亚铁血红素 (heme) 与香叶醇结合腔中心坐标。活性空腔周围的氨基酸残基侧链设置为柔性。格点 box 设置为 4 nm × 4 nm × 4 nm, 格点间距 0.037 5 nm。分子对接在 AutoDock Vina 中进行。采用 PyMOL 2.0 软件对亲和力 (kcat/mol) 最佳的模型进行可视化分析。

1.11 *GrG10H* 基因表达分析

根据转录组数据和同属植物中管家基因的报道, 选择滇龙胆草转录组中 *GrACTIN* 基因 (GenBank 序列号: KM061808)^[20] 为内参基因设计实时荧光定量 PCR 引物 (见表 1)。根据 *GrG10H1* 和 *GrG10H2* 基因的 DNA 序列设计目的基因的实时荧光定量 PCR 引物 (见表 1)。使用嵌合荧光检测试剂盒进行 qPCR, qPCR 程序如下: 94 °C 30 s; 94 °C 5 s, 60 °C 30 s, 共 40 个循环 (每次循环后采集荧光); 循环结

束后, 在 60 ~ 95 °C 温度范围内记录溶解曲线, 每个温度以 0.5 °C 上升, 每次停留 5 s。扩增曲线、溶解曲线、标准曲线由 qPCR 仪软件自动生成。使用内参基因 *GrACTIN* 表达校准后, 采用比较 Ct 值的 $2^{-\Delta\Delta Ct}$ 方法计算出根、茎、叶中 *GrG10H* 基因的相对表达量。实验设置三个生物学重复和三个实验重复, 结果用 Graph Pad Prism 5.0 软件作图表示。

2 结果与分析

2.1 滇龙胆草愈伤组织诱导及不同组织部位中龙胆苦苷含量分析

以滇龙胆草新鲜根、茎、叶为材料诱导愈伤组织, 以添加激素配比为 1.5 mg/L 2,4-D, 0.2 mg/L 6-BA 的 MS 培养基诱导培养 10 d 后, 在外植体周围及切口处观察到愈伤组织出现, 继续培养至 14 d 后, 切口处可见明显黄色透明愈伤组织, 其中由茎段诱导出愈伤组织的外植体数量远高于根和叶, 证明茎是诱导滇龙胆草愈伤组织的最适组织 (见图 2A)。

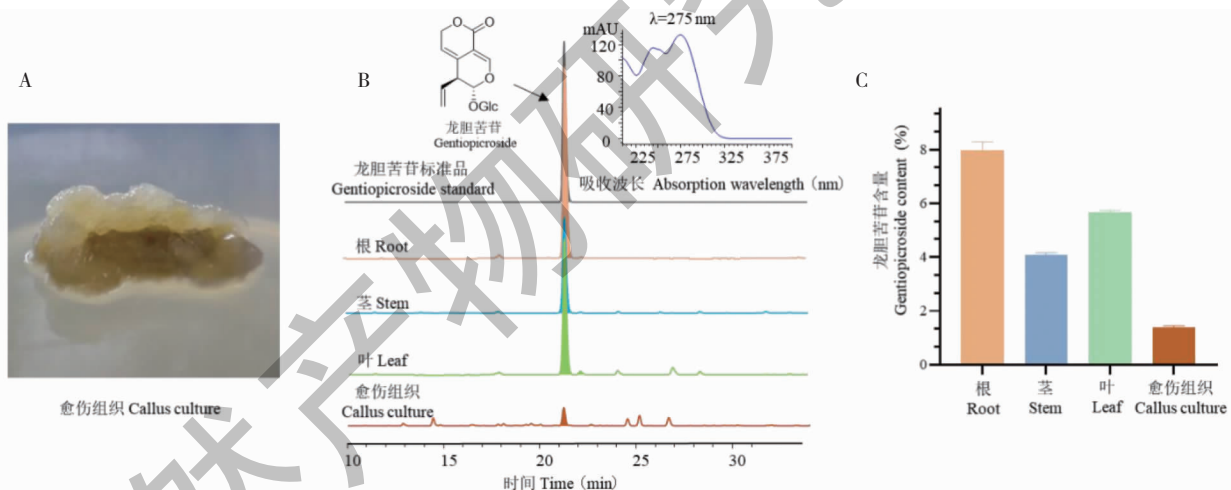


图 2 滇龙胆草不同部位化学成分及龙胆苦苷含量分析

Fig. 2. Chemical components analysis and gentiopicroside content determination in different parts of *G. rigescens*

注: A: 滇龙胆草愈伤组织诱导; B: 滇龙胆草不同部位及愈伤组织化学成分分析; C: 滇龙胆草不同部位及愈伤组织龙胆苦苷含量分析。Note:

A: The induced callus culture of *G. rigescens* plant; B: Chemical components analysis of different parts of *G. rigescens*; C: Content determination of gentiopicroside of different parts and callus culture of *G. rigescens*.

以不同含量龙胆苦苷标准品 (0.625 ~ 40 mg/L) 在 275 nm 波长下的紫外吸收峰面积绘制龙胆苦苷含量标准曲线, 回归方程为 $y = 527.57x - 249.71$ ($R^2 = 0.9987$)。在相同分析条件下分别对滇龙胆草根、茎、叶和诱导愈伤组织中的化学成分进行分析, 色谱图见图 2B。结果发现, 龙胆苦苷在滇龙胆草植株根、茎、叶部位中均为最主要成分; 愈伤组织

也明显检测到龙胆苦苷的生成, 但其在愈伤组织中的相对含量远低于植株; 进一步对植株不同部位及愈伤组织的龙胆苦苷含量进行了测定, 根据标准曲线公式计算求得对应龙胆苦苷含量, 以组织干重质量分数表示。结果发现, 龙胆苦苷含量在不同组织部位表现出明显的差异 (见图 2C), 具体含量为: 根 (7.8%) > 叶 (5.6%) > 茎 (4.1%) > 愈伤组织

(1.4%),其中龙胆苦苷在根中的含量是愈伤组织中的 5.6 倍。

2.2 G10H 基因的筛选

通过 DNBSEQ 平台测序,总计产出 76.39 Gb 二代转录组测序数据。组装并去冗余后得到 174 890 个 Unigene。根据 KEGG、NR、NT、SwissProt、GO、Pfam、KOG 七大功能数据库注释,从滇龙胆草转录组数据中共获得了 118 条潜在的香叶醇 10-羟化酶(geraniol 10-hydroxylase, G10H)基因。进一步以已经报道的、完成功能鉴定的香叶醇 10-羟化酶 *SmG10H*(GenBank 序列号: ACZ48680.1), *CrG10H*(GenBank 序列号: Q8VWZ7.1)的氨基酸序列为模板通过本地 Blastp 搜索滇龙胆草转录组数据库,去掉重复冗余以及氨基酸长度 < 200 bp 的序列后,共获得 20 条候选序列;进一步将这 20 条基因在滇龙胆草不同组织部位比较转录组数据库中的表达水平进行了差异分析(见图 3)。结果显示:有 7 条基因

在根中高表达(见图 3 中标记区域 I),7 条基因在叶片中高表达(见图 3 中标记区域 II),其余 6 条基因在茎中高表达(见图 3 中标记区域 III)。结合上述不同部位化学成分分析结果显示龙胆苦苷在滇龙胆草根中含量较高,因此将转录组差异分析后根部高表达的候选序列进一步与已知功能基因序列进行相似度比对,筛选获得与同源蛋白序列相似度较高的 *TRINITY_DN240_c0_g1_i7-leaf_3* 作为候选基因 1(命名为 *GrG10H2*, GenBank 序列号: PP740721)。另一方面,依据文献报道,有学者认为龙胆苦苷主要是在植物叶片中合成,然后经过茎的转运,最后在根中积累,因此不能排除叶中高表达基因在龙胆苦苷生物合成中发挥重要作用的可能性^[6]。*TRINITY_DN25850_c0_g1_i1-leaf_2* 为在叶片中高表达候选序列中与目前已报道功能蛋白相似度较高的序列,故将其同时作为候选基因 2(命名为 *GrG10H1*, GenBank 序列号: PP740720)进行后续研究。

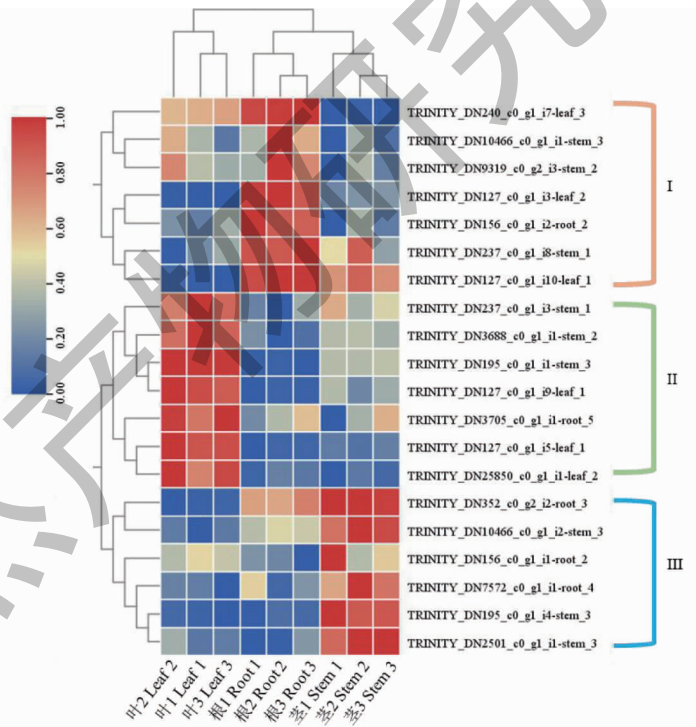


图 3 滇龙胆草不同部位中差异表达基因热图

Fig. 3 Heatmap of differentially expressed genes in different parts of *G. rigescens*

2.3 龙胆草中 *GrG10H1* 和 *GrG10H2* 基因的克隆

以滇龙胆草总 RNA 为模板进一步反转录合成 cDNA,根据实验前期筛选获得的两条具有完整开放阅读框的 G10H 基因 *GrG10H1* 和 *GrG10H2* 的序列设计特异性引物,以滇龙胆草 cDNA 为模板,进行

PCR 扩增后在 1 500 bp 处均检测到单一大小条带(见图 4)。将 PCR 扩增产物构建到 pCE2 TA/Blunt-Zero 克隆载体中进行测序验证序列,最终得到两条基因 *GrG10H1* 和 *GrG10H2*,基因长度分别为 1 548 bp、1 482 bp,与转录组数据分析结果一致。

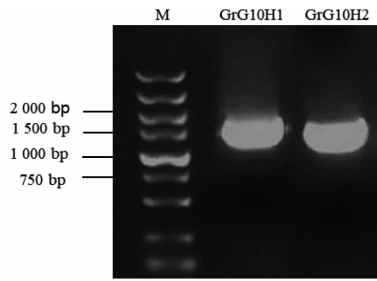


图4 候选 G10H 基因的琼脂糖凝胶电泳图

Fig. 4 Agarose gel electrophoresis of candidate G10H gene
注: M 表示 DNA Marker。Note: M represents DNA Marker.

2.4 滇龙胆草 *GrG10H1* 及 *GrG10H2* 基因的生物信息学分析

2.4.1 *GrG10H1* 和 *GrG10H2* 基因编码氨基酸理化性质分析

利用 ExPASy 的在线工具 ProtParam tool 分析 *GrG10H1* 和 *GrG10H2* 所编码蛋白的理化性质, 结果

显示 *GrG10H1* 蛋白分子量为 57.66 kDa, 化学方程式为 $C_{2594}H_{4120}N_{708}O_{735}S_{21}$, 不稳定性指数较低 (38.64), 为稳定性蛋白。总亲水性平均系数 (grand average of hydropathicity, GRAVY) 为 -0.201, 为亲水性蛋白。*GrG10H2* 蛋白分子量为 55.13 kDa, 化学方程式为 $C_{2480}H_{3987}N_{667}O_{703}S_{23}$, 不稳定性指数较低 (33.05), 为稳定性蛋白, GRAVY 为 0.008 也为亲水性蛋白。利用 <https://web.expasy.org/protscale/> 网站分析 *GrG10H1* 和 *GrG10H2* 蛋白的亲/疏水性, 结果显示 *GrG10H1* 蛋白中亲水性最强的为 344 位的赖氨酸, 其最低分值为 -2.444, 疏水性最强的为 32 位的苯丙氨酸, 其最高分值为 2.744。*GrG10H2* 亲水性最强的为 145 位的亮氨酸, 其最低分值为 -2.556, 疏水性最强的为 8 位的缬氨酸和 10 位的亮氨酸, 其最高分值为 2.622, 也证实该类蛋白质为一种亲水性蛋白 (见图 5)。

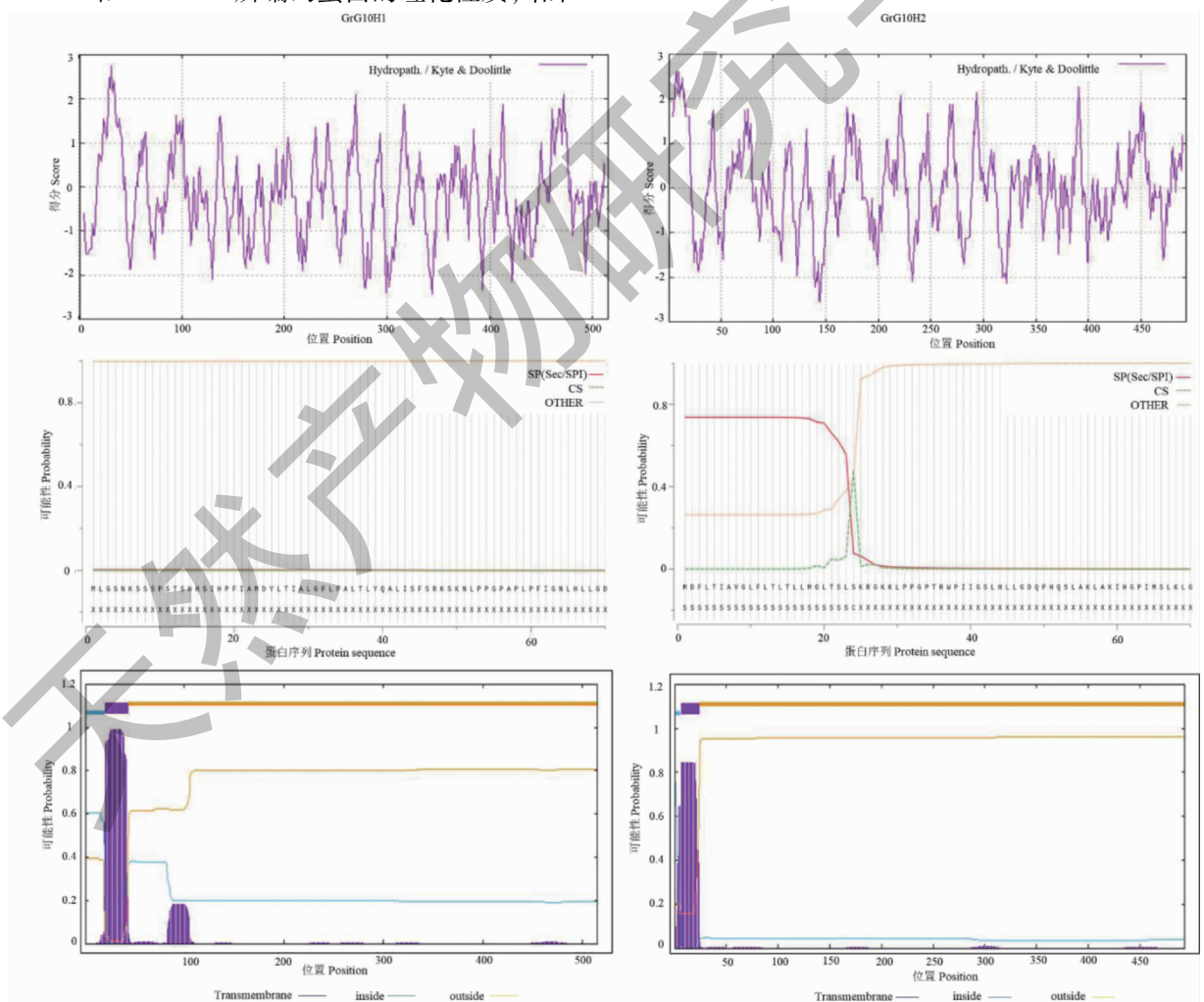


图5 *GrG10H1* 和 *GrG10H2* 蛋白的亲/疏水性 (左)、信号肽分析 (中) 及跨膜区预测 (右)

Fig. 5 Hydrophilicity/hydrophobicity (left line) and signal peptide (middle line) analysis, and transmembrane region (right line) prediction of *GrG10H1* and *GrG10H2* protein

2.4.2 GrG10H1 和 GrG10H2 蛋白质信号肽分析和跨膜结构域预测

利用 SignalP 5.0 服务器分析 GrG10H1 和 GrG10H2 蛋白质的信号肽,结果显示 GrG10H1 蛋白存在信号肽的可能性为 0.0028,表明 GrG10H1 蛋白质无信号肽,为非分泌型蛋白质,而 GrG10H2 蛋白质存在信号肽可能性为 0.736,在蛋白 N 端的 1~28 位存在信号肽,推测 GrG10H2 蛋白属于分泌型蛋白(见图 5)。利用 TMHMM2.0 工具预测 GrG10H1 和 GrG10H2 蛋白质的跨膜螺旋区,结果表明:GrG10H1 蛋白质在 21~43 位氨基酸含有跨膜区域,GrG10H2 蛋白质在 7~26 位氨基酸含有跨膜区域(见图 5)。

2.4.3 GrG10H1 和 GrG10H2 蛋白分类及结构域预测

使用 InterPro 6.0 在线工具对 GrG10H1 和 GrG10H2 蛋白质进行分析和归类,结果表明,GrG10H1 及 GrG10H2 蛋白均属于细胞色素 P450 超家族 E 类、类群 I (IPR002401)的成员,包含细胞色素 P450 保守结构域(IPR036396),属于 CYP76 亚家族。分子功能 GO 预测结果表明,GrG10H1 及 GrG10H2 蛋白质均可与血红素亚铁离子结合,可能具有单加氧酶活性、氧化还原酶活性。通过 NCBI 的 Conserved Domain Database 工具预测两种 GrG10H 蛋白质的保守功能域,也证实了 GrG10H1 和 GrG10H2 蛋白属于细胞色素 P450 家族。

2.4.4 GrG10H1 和 GrG10H2 蛋白质亚细胞定位分析

运用 Predict Protein 在线软件预测 GrG10H1 和

GrG10H2 蛋白的亚细胞定位,结果显示 GrG10H1 及 GrG10H2 均定位在内质网,可信度分别为 78% 和 56%;使用 Softberry 数据库 Prot Comp 9.0 软件预测结果显示 GrG10H1 和 GrG10H2 亚细胞定位结果与 Predict Protein 预测结果一致,两种蛋白的亚细胞均定位于内质网。

2.5 GrG10H1 和 GrG10H2 蛋白序列比对分析

利用 NCBI 网站的 BLASTP/程序分别对 GrG10H1 和 GrG10H2 蛋白进行序列比对分析,结果显示,GrG10H1 和 GrG10H2 相互之间的蛋白序列相似度为 63.29%。其中 GrG10H1 蛋白与同属植物龙胆草(*G. rigescens*)中 geraniol 10-hydroxylase 蛋白相似度最高(99.61%),GrG10H2 蛋白与同属植物秦艽(*Gentiana macrophylla*)中 geraniol 10-hydroxylase 蛋白相似度最高(66.95%)。其次二者与川西獐牙菜(*Swertia mussotii*)、长春花(*Catharanthus roseus*)、欧防风(*Pastinaca sativa*)、细辛叶獐牙菜(*Swertia asarifolia*)等其他植物物种来源的蛋白序列同源性也相对较高,故选择以上序列利用 ESprict 3.0 软件进行多序列比对分析^[29],结果显示 GrG10H1 和 GrG10H2 具有大多数植物来源 P450 蛋白的高保守结构域,包括 N 端富集脯氨酸的区域即 PPGPXPLP,与底物和铁结合氧互作的中央螺旋区(central helix I)即 AGTDTT 以及血红素结合域(heme-binding)即 PFGXGRRXCXG^[21](见图 6)。

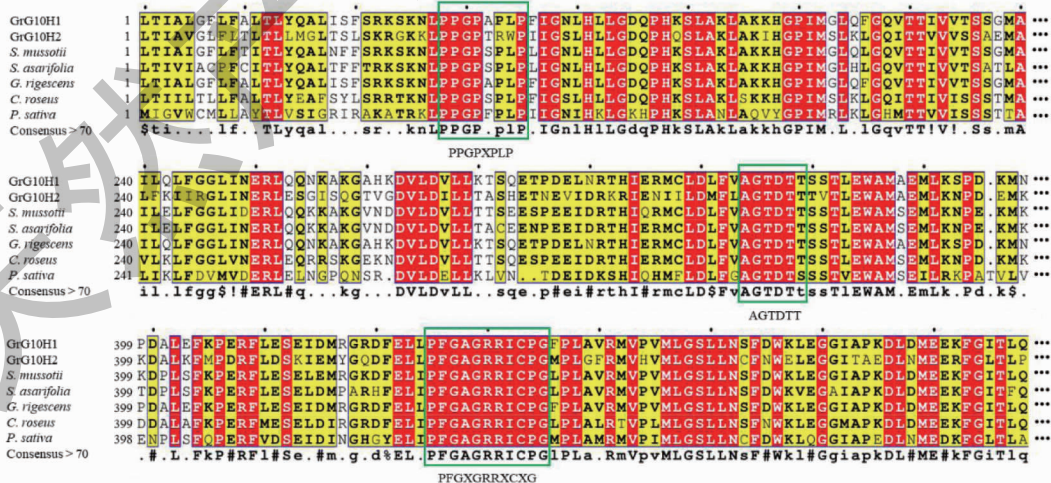


图 6 GrG10H1 和 GrG10H2 基因编码蛋白与其他植物来源 G10H 蛋白多序列比对分析

Fig. 6 Multiple amino acid sequence alignment of GrG10H1 and GrG10H2 with other G10H proteins from other plants

2.6 GrG10H1 和 GrG10H2 蛋白系统进化分析

将 GrG10H1 和 GrG10H2 的氨基酸序列与细胞

色素 P450 家族中 CYP76 其他亚家族氨基酸序列构建系统发育树(见图 7)。系统进化分析结果显示:

G10H 基因按照不同 CYP450 亚家族的分类处于不同的进化分支,而 GrG10H1 和 GrG10H2 均处于 CYP76B 亚家族的分支中,分别与来自川西獐牙菜

(*S. mussotii*) 的香叶醇 10-羟化酶 CYP76B10 和来自长春花(*C. roseus*) 中的香叶醇 10-羟化酶 CYP76B6 亲缘关系最为相近。

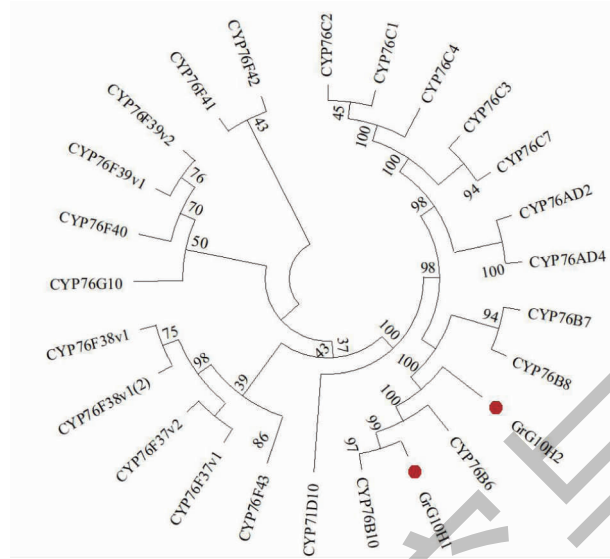


图 7 GrG10H1 和 GrG10H2 基因编码蛋白系统进化分析

Fig. 7 Phylogenetic analysis of GrG10H1 and GrG10H2 protein

2.7 GrG10H1 和 GrG10H2 蛋白三维结构研究及关键位点分析

运用在线工具 SOPMA 对 GrG10H 的蛋白二级结构进行了预测,其中 GrG10H1 蛋白二级结构中 α -螺旋占 47.96%, β -折叠占 3.69%, 延伸带占 12.23%, 无规则卷曲占 36.12% (见图 8A); GrG10H2 蛋白质二级结构中 α -螺旋占 50.30%, β -折叠占 5.07%, 延伸带占 11.97%, 无规则卷曲占 32.66% (见图 8B)。以丹参 (*Salvia miltiorrhiza*) 中报道的 P450 酶 CYP76AH1 (PDB ID: 5YLM) 为模板, 利用在线分析软件 SWISS-MODEL 对 GrG10H1 和 GrG10H2 的蛋白三维结构进行预测, 结果如图 8C 所示, GrG10H1 和 GrG10H2 与 CYP76AH1 的氨基酸序列相似度分别为 42% 与 43%, 三维结构叠合结果显示其空间结构重合度较好, GrG10H1 与 CYP76AH1 蛋白叠合后均方根偏差 (root mean square deviation, RMSD) 值为 0.006 5 nm, GrG10H2 和与 CYP76AH1 蛋白叠合后均方根偏差 RMSD 值为 0.006 3 nm, 这也与同源序列比对及系统进化分析结果中 GrG10H 蛋白属于 CYP76 亚家族的结果是一致的 (见图 8C)。

通过 NCBI BLAST 比对分析, 选择同源性较高且同时含有两个配体 (heme 与底物) 的来自高粱 (*S. bicolor*) 中的 P450 酶名为肉桂酸 4-羟化酶 (cin-

namate 4-hydroxylase, SbC4H1) 的蛋白晶体复合物 (PDBID 6VBY) 作为比对模板, 分别对 GrG10H1 和 GrG10H2 蛋白与血红素及香叶醇配体结合的区域进行分析。采用 AutoDock vina 程序进行分子对接分析, 结果如见图 8E-8G 所示。从氨基酸类型及空间分布与构型来看, GrG10H1 (绿色) 和 GrG10H2 (蓝色) 二者与血红素的结合口袋都极为相似, 重叠度较高 (见图 8D)。图中同时观察到多个 P450 家族中与 heme 结合相关的保守氨基酸, 包括 GrG10H1 (GrG10H2) 中的 R97 (R119)、W122 (W144)、R126 (R148)、T301 (T323)、T302 (T324)、P428 (P450)、F429 (F451)、R434 (R456)、C436 (C458)、P437 (P459) 和 G438 (G460)。而在底物香叶醇结合口袋中 (见图 8F), GrG10H1 (GrG10H2) 的几处关键氨基酸位点有明显差异, 包括 A103 (L125)、L296 (V318)、G362 (V384)、L478 (I500), 同时 W115 (W137) 位点的色氨酸在空间构象上也有明显差异, 这些变化导致香叶醇在 GrG10H1 和 GrG10H2 的结合构型也存在差异。在 GrG10H1 中, 香叶醇中的醇羟基与底物口袋中的氨基酸残基 A297、D300 和 T301 形成分子间氢键, C10 位与血红素中 Fe 离子在空间上靠近, 距离为 0.2 nm (见图 8E)。在 GrG10H2 中, 香叶醇中的醇羟基与底物口

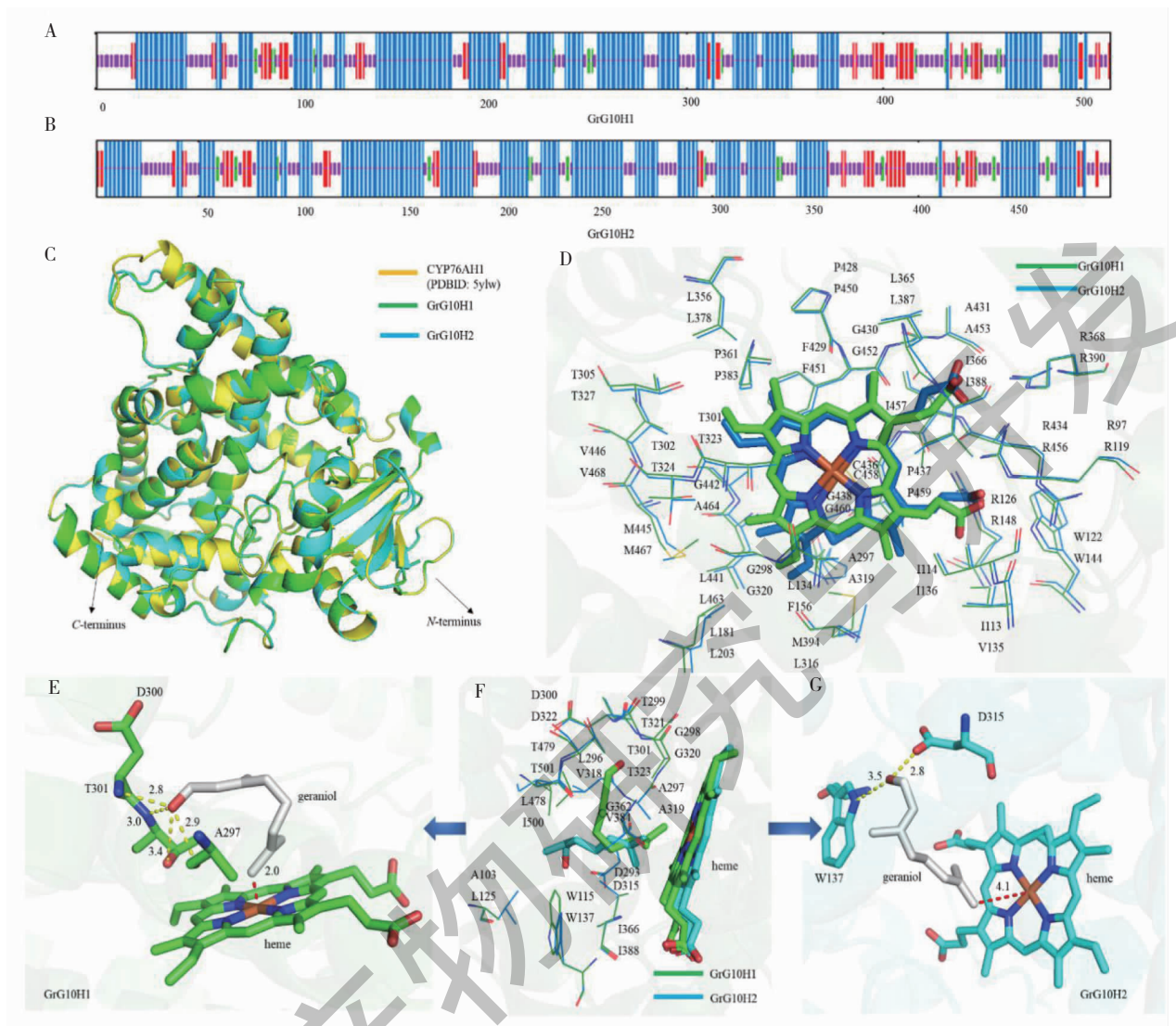


图 8 GrG10H 蛋白三维结构及分子对接研究

Fig. 18 3D structures and molecular docking of Cg10H protein

注: A: GrG10H1 蛋白二级结构预测; B: GrG10H2 蛋白二级结构预测; C: 以 CYP76A11 为模板预测 GrG10H1 及 GrG10H2 蛋白三维结构; D: GrG10H1、GrG10H2 与血红素分子对接结果; E: GrG10H1 与香叶醇及血红素对接结果; G: GrG10H2 与香叶醇及血红素对接结果; F: GrG10H1、GrG10H2 与香叶醇及血红素对接结果叠合分析。Note: A: Predicted secondary structure of GrG10H1 using SOPMA; B: Predicted secondary structure of GrG10H2 using SOPMA; C: Three-dimensional structure alignment of GrG10H1, GrG10H2 and the template CYP76A11; D: Key residues in the heme iron binding pocket of GrG10H1 and GrG10H2. Heme was shown as sticks while key residues were shown as spheres; E: Binding conformations of geraniol and heme in GrG10H1; G: Binding conformations of geraniol and heme in GrG10H2; F: Binding conformations of geraniol and heme in GrG10H1, GrG10H2 and their alignment.

袋中的氨基酸残基 W137 和 D315 间形成 2 个氢键, C10 位与血红素中 Fe 离子在空间上靠近, 距离为 0.41 nm (见图 8G)。

2.8 GrG10H1 和 GrG10H2 基因在根茎叶中的表达分析

通过 qRT-PCR 检测 GrG10H1 和 GrG10H2 基因在根茎叶中的表达情况。结果显示 (见图 9),

GrG10H1 基因在叶中表达量最高, 在根中表达量最低, 仅为叶的 1/3, 该表达模式与川西獐牙菜 SmG10H 基因在根茎叶中的表达模式类似^[22]。而 GrG10H2 基因在根和叶中表达量较高, 在茎中几乎不表达, GrG10H2 基因在根中的表达量是茎的 10 倍, 这与长春花中 CrG10H 基因主要在根和叶中高表达的模式一致。

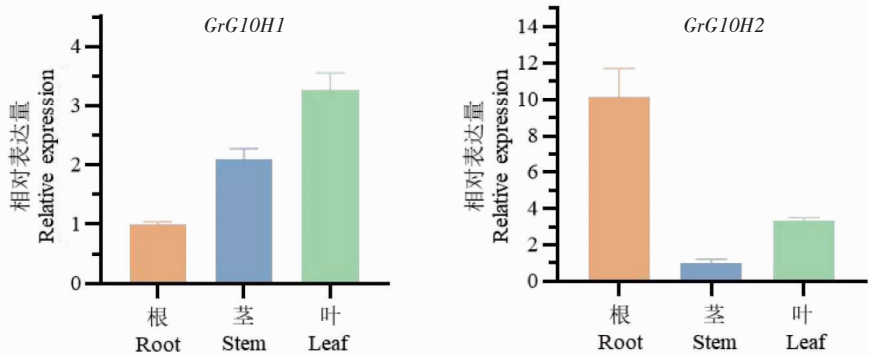


图9 *GrG10H1*、*GrG10H2* 基因在滇龙胆草根茎叶中的表达分析

Fig. 9 Relative expression level of *GrG10H1* and *GrG10H2* in *G. rigescens* roots, stems and leaves

3 讨论与结论

龙胆苦苷是龙胆、秦艽、川西獐牙菜等龙胆科植物的主要活性成分,具有抗炎镇痛、保肝利胆、抑制病原微生物等诸多药理学作用,目前关于调控龙胆苦苷生物合成途径的中下游阶段的关键酶尚未完全挖掘。香叶醇 10-羟化酶 (geraniol 10-hydroxylase, G10H) 是一种细胞色素 P450 单加氧酶,是调控龙胆苦苷生物合成途径的关键限速酶,该酶可催化香叶醇在 C-10 位羟化,产生 10-羟基香叶醇^[23],随后该前体物质可经过氧化、还原、甲基化、糖基化反应生成关键中间体裂环番木鳖苷,最后裂环番木鳖苷经尚未阐明的途径生成龙胆苦苷。目前 G10H 仅在长春花、川西獐牙菜、秦艽等少数物种中被克隆。

本研究以云南道地药材滇龙胆草为材料,测定了龙胆苦苷在不同组织部位的成分含量,发现龙胆苦苷在根部含量最多,茎和叶次之,在愈伤组织中含量最低。实验后期获得能够稳定传代的愈伤组织后,进行不同条件的胁迫处理,从而探究龙胆苦苷在不同处理条件下愈伤组织中的积累量。本研究对滇龙胆草不同部位比较转录组数据进行了深入分析,结合与已知功能序列同源比对,筛选获得了 2 条 G10H 候选基因分别命名为 *GrG10H1*、*GrG10H2*,并进行分子克隆,生物信息学分析发现两条基因所编码的蛋白均为亲水性蛋白,具有跨膜结构域。通过保守结构域分析发现,*GrG10H1* 和 *GrG10H2* 具有大多数植物 P450 的保守结构域,进化树分析表明二者均属于细胞 P450 家族中 CYP76B 亚家族。本研究还对 *GrG10H1* 和 *GrG10H2* 的蛋白三维结构进行了同源建模和分子对接分析,在对接结果中观察到多个 P450 酶中与 heme 结合相关的保守氨基酸,且底物香叶醇的构象及活性位点与血红素亚铁离子的空

间距离均支持 C10 位羟化反应的发生,对接结果提示了 *GrG10H1* 和 *GrG10H2* 具有催化香叶醇发生羟化的可能性。荧光定量 PCR 分析 *GrG10H1* 和 *GrG10H2* 在滇龙胆草不同部位的表达模式发现,*GrG10H1* 在叶中相对表达量最高,茎和根中较低,而 *GrG10H2* 在根和叶中的表达量最高。同时,滇龙胆草愈伤组织细胞培养体系的获得也为后续参与龙胆苦苷生物合成途径其他关键基因的克隆及表达调控提供了可持续利用的生物材料。本研究后续将构建酵母真核表达系统,实现对两条基因催化功能的验证,为阐明龙胆苦苷的生物合成途径奠定基础。

参考文献

- Zhang XD, Allan AC, Li CX, et al. De novo assembly and characterization of the transcriptome of the Chinese medicinal herb, *Gentiana rigescens* [J]. *Int J Mol Sci*, 2015, 16: 11550-11573.
- Yang JQ. Structural modification and activity study of gentiopicroside [D]. Kunming: Yunnan University (云南大学), 2016.
- Xie XQ, Li H, Wang YL, et al. Anti-inflammatory effect of gentiopicroside in adjuvant-induced arthritis rats [J]. *Chin J Exp Tradit Med Form* (中国实验方剂学杂志), 2020, 26: 58-63.
- Ding J. Effects of gentiopicroside on anit-induced cholestasis in rats [J]. *China Pharm* (中国药师), 2020, 23: 2133-2137.
- Liu B, Pang F, Bi H, et al. Regulatory mechanisms of gentiopicroside on human diseases: a brief review [J]. *Naunyn Schmiedebergs Arch Pharmacol*, 2024, 397: 725-750.
- Zhang XD, Li CX, Li S, et al. Cloning and expression analysis of geraniol 10-hydroxylase gene in *Gentiana rigescens* [J]. *Southwest China J Agric Sci* (西南农业学报), 2017, 30: 1499-1506.

- 7 Coscia CJ, Guarnaccia R. Biosynthesis of gentiopicroside, a novel monoterpene [J]. J Am Chem Soc, 1967, 89: 1280-1281.
- 8 Zhang XD, Allan AC, Li C, et al. De novo assembly and characterization of the transcriptome of the Chinese medicinal herb, *Gentiana rigescens* [J]. Int J Mol Sci, 2015, 16: 11550-11573.
- 9 Kang H, Zhao ZL, Ni LH, et al. Transcriptome analysis and exploration of genes involved in the biosynthesis of iridoids in *Gentiana crasicaulis* (Gentianaceae) [J]. Acta Pharm Sin (药学学报), 2021, 56: 2005-2014.
- 10 Collu G, Unver N, Peltenburg-Looman AM, et al. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis [J]. Febs Lett, 2001, 508: 215-220.
- 11 Miettinen K, Dong L, Navrot N, et al. The seco-iridoid pathway from *Catharanthus roseus* [J]. Nat Commun, 2014, 5: 3606.
- 12 Fu N, Yang ZL, Pauchet Y, et al. A cytochrome P450 from the mustard leaf beetles hydroxylates geraniol, a key step in iridoid biosynthesis [J]. Insect Biochem Mol Biol, 2019, 113: 103212.
- 13 Munkert J, Pollier J, Miettinen K, et al. Iridoid synthase activity is common among the plant progesterone 5 β -reductase family [J]. Mol Plant, 2015, 8: 136-152.
- 14 Ouyang PY, Li YP, Liang YY, et al. Cloning and expression analysis of geraniol-10 hydroxylase gene in different cultivated varieties of *Patchouli* [J]. J Chin Med Mater (中药材), 2017, 40: 315-319.
- 15 Hua WP, Wang JZ. Cloning and sequence analysis of G10H gene from *Gentiana macrophylla* [J]. Genomics Appl Biol (基因组学与应用生物学), 2013, 32: 510-515.
- 16 Zhao S, Dong TT, Tang H. Cloning of geraniol-10-hydroxylase gene from *Valeriana jatamansi* and expression analysis [J]. J Chin Med Mater (中药材), 2019, 42: 1007-1011.
- 17 Li SM, Unsöld IA. Post-genome research on the biosynthesis of ergot alkaloids [J]. Planta Med, 2006, 72: 1117-1120.
- 18 Sung PH, Huang FC, Do YY, et al. Functional expression of geraniol 10-hydroxylase reveals its dual function in the biosynthesis of terpenoid and phenylpropanoid [J]. J Agric Food Chem, 2011, 59: 4637-4643.
- 19 Hou Q, Guo M, Zhang F, et al. Callus cultivation and determination of gentiopicroside from *Gentiana macrophylla* [J]. J Chin Med Mater (中药材), 2012, 35: 675-678.
- 20 Zhang XD, Li CX, Wang YZ, et al. Cloning and expression analysis of the 7-deoxygenate-7-hydroxylase gene from *Gentiana rigescens* [J]. Jiangsu Agric Sci (江苏农业科学), 2019, 47: 76-81.
- 21 Zhang XD, Li CX, Li S, et al. Cloning and expression analysis of geraniol 10-hydroxylase gene in *Gentiana rigescens* [J]. Southwest China J Agric Sci (西南农业学报), 2017, 30: 1499-1506.
- 22 Wang J, Liu Y, Cai Y, et al. Cloning and functional analysis of geraniol 10-hydroxylase, a cytochrome P450 from *Swertia mussotii* Franch [J]. Biosci Biotechnol Biochem, 2010, 74: 1583-1590.
- 23 Sintupachee S, Promden W, Ngamrojanavanich N, et al. Functional expression of a putative geraniol 8-hydroxylase by reconstitution of bacterially expressed plant CYP76F45 and NADPH-cytochrome P450 reductase CPR I from *Croton stellatopilosus* Ohba [J]. Phytochemistry, 2015, 118: 204-215.