

## 连翘 *FsUGTs* 基因家族鉴定与分析

李文文<sup>1,2</sup>, 闫雪娇<sup>1,2</sup>, 许家齐<sup>1,2</sup>, 李义龙<sup>1,2</sup>, 张力予<sup>1,2</sup>, 许高龙<sup>1,2</sup>, 马占强<sup>1,2</sup>, 吕淑芳<sup>1,2\*</sup>, 侯典云<sup>1,2\*</sup>

1. 河南科技大学农学院, 河南 洛阳 471023

2. 河南省药食兼用资源评价与创新利用工程研究中心, 河南 洛阳 471023

**摘要:** **目的** 基于转录组数据筛选鉴定分析连翘 *Forsythia suspensa* *FsUGTs* 家族。 **方法** 基于转录组与基因组鉴定连翘 *FsUGTs* 基因家族, 运用生物信息学技术对其展开综合分析, 深入探究其蛋白的理化特性、亚细胞定位。通过系统发育树构建、染色体定位、顺式作用元件分析、保守基序、结构域和基因结构分析, 进一步揭示这些基因家族成员间的进化关系。此外通过 RT-qPCR 检测 *FsUGTs* 在不同组织(根、茎、叶、果)的表达状况, 并将其与连翘苷的组织分布结合进行相关性分析。 **结果** 在连翘的转录组数据中经过筛选、鉴定, 最终得到 82 个 *FsUGTs*, 其氨基酸长度为 385~558 aa, 相对分子质量在 43 140~60 750, 理论等电点范围为 4.81~8.26, 主要定位在质膜上, 均含有 PSPG box 结构域。系统发育分析将 82 个 *FsUGTs* 分为 17 组, 其中 L 组的 UGT 最多, 有 16 条, 其次是 A 组和 E 组, 分别为 10 条和 9 条, Q 组没有 *FsUGTs* 分布。染色体定位结果显示 16 个 *FsUGTs* 定位于 Chr10, 而 Chr11、Chr12、Chr13 上各有 7 个 *FsUGTs*, Chr3 上的 *FsUGTs* 数量最少, 只有 2 个。顺式作用元件分析表明 *FsUGTs* 的表达受光的影响最大, 茉莉酸甲酯的影响次之。结构域分析显示 82 个 *FsUGTs* 中都有 3 个保守的结构域, 分别为 GT1\_Gtf-like、PLN02448 和 Glycosyltransferase\_GTB-type superfamily。组织差异性表达与连翘苷在不同组织中含量分布的相关性分析结果显示 *FsUGT1*~*FsUGT3*、*FsUGT9*、*FsUGT12*~*FsUGT16*、*FsUGT18*、*FsUGT21*~*FsUGT24* 和 *FsUGT31* 可能在连翘脂素糖基化形成连翘苷的过程中发挥作用。 **结论** 鉴定并分析了连翘 *FsUGTs* 家族的理化性质、顺式作用元件、蛋白结构、系统发育关系、组织差异性表达等, 对可能参与到连翘脂素糖基化形成连翘苷的 *FsUGTs* 进行了筛选, 为后续的深入寻找连翘脂素糖基转移酶及理解连翘药用成分合成的调控机制奠定了基础。

**关键词:** 连翘; *UGT* 基因家族; 连翘苷; 组织差异性表达; 染色体定位

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2026)05-1853-11

DOI: 10.7501/j.issn.0253-2670.2026.05.022

## Identification and analysis of *FsUGTs* gene family in *Forsythia suspensa*

LI Wenwen<sup>1,2</sup>, YAN Xuejiao<sup>1,2</sup>, XU Jiaqi<sup>1,2</sup>, LI Yilong<sup>1,2</sup>, ZHANG Liyu<sup>1,2</sup>, XU Gaolong<sup>1,2</sup>, MA Zhanqiang<sup>1,2</sup>, LYU Shufang<sup>1,2</sup>, HOU Dianyun<sup>1,2</sup>

1. College of Agricultural, Henan University of Science and Technology, Luoyang 471023, China

2. Henan Engineering Research Center for Evaluation and Innovative Utilization of homology of medicine and food, Luoyang 471023, China

**Abstract: Objective** To screen, identify, and analyze the *FsUGTs* family from *Forsythia suspensa* based on transcriptomic data.

**Methods** The *FsUGTs* gene family of *F. suspensa* was identified from transcriptomic and genomic data. Bioinformatics techniques were applied for comprehensive analysis to explore the physicochemical properties and subcellular localization of the proteins encoded by these genes. The evolutionary relationships among these gene family members were further revealed through phylogenetic tree construction, chromosomal localization, cis-acting element analysis, conserved motif analysis, domain analysis, and gene structure analysis. In addition, RT-qPCR was used to detect the expression patterns of *FsUGTs* in different tissues (roots, stems, leaves, and fruits), and correlation analysis was performed by integrating these results with the tissue distribution of phillyrin. **Results** After screening and identification in the *F. suspensa* transcriptome data, a total of 82 *FsUGTs* were finally obtained. Their amino acid lengths

收稿日期: 2025-11-02

基金项目: 中央本级重大增减支项目“名贵中药资源可持续利用能力建设项目”(2060302); 河南省中央引导地方科技发展资金项目(Z20241471030); 河南省科技攻关项目(242102110325); 河南省中药材产业科技特派员服务团项目

作者简介: 李文文(1998—), 硕士研究生, 主要从事药用植物分子生物学研究。E-mail: 15037099335@163.com

\*通信作者: 侯典云(1975—), 教授, 博士生导师, 主要从事药用植物资源评价与利用。Tel: (0379)64282340 E-mail: dianyun518@163.com

ranged from 385 to 558 amino acids (aa), relative molecular masses from 43 140—60 750, and theoretical isoelectric points (pI) from 4.81 to 8.26. These proteins are primarily localized to the plasma membrane and all contain the PSPG box domain. Phylogenetic analysis classified the 82 *FsUGTs* into 17 groups: Group L contained the highest number of UGTs (16 members), followed by Groups A and E with ten and nine respectively, while no *FsUGTs* were distributed in Group Q. Chromosomal localization results showed that 16 *FsUGTs* were located on Chr10, with seven each on Chr11, Chr12, and Chr13, while Chr3 contained the fewest *FsUGTs* (only two). Analysis of cis-acting elements indicated that the expression of *FsUGTs* was most strongly influenced by light, followed by methyl jasmonate. Domain analysis revealed that all 82 *FsUGTs* contained three conserved domains, namely GT1\_Gtf-like, PLN02448, and Glycosyltransferase\_GTB-type superfamily. Correlation analysis between tissue-specific expression of *FsUGTs* and phillyrin content in different tissues indicated that *FsUGT1-FsUGT3*, *FsUGT9*, *FsUGT12-FsUGT16*, *FsUGT18*, *FsUGT21-FsUGT24*, and *FsUGT31* may play roles in the glycosylation of phillygenin to form phillyrin. **Conclusion** This study identified and analyzed the physicochemical properties, cis-acting elements, protein structures, phylogenetic relationships, and tissue-specific expression of the *FsUGTs* family from *F. suspensa*. It also screened *FsUGTs* that may be involved in the glycosylation of phillygenin to phillyrin, laying a foundation for the further identification of phillygenin glycosyltransferases and the understanding of the regulatory mechanisms underlying the synthesis of medicinal components in *F. suspensa*.

**Key words:** *Forsythia suspensa* (Thunb.) Vahl; UGT gene family; phillyrin; tissue-specific differential expression; chromosome localization

连翘 *Forsythia suspensa* (Thunb.) Vahl 始载于《神农本草经》，为木犀科连翘属落叶灌木，又名黄花条、连壳、黄奇丹等<sup>[1]</sup>。连翘的应用始自宋代，多以果实入药，是我国临床常用传统中药之一，连翘根及连翘茎叶亦供药用<sup>[2]</sup>。其性微寒，味苦，归肺、心、胆经，具有清热解毒、消肿散结、疏散风热的功效，临床常用于急性风热感冒、痈肿疮毒、淋巴结核、尿路感染等症，中医素有“疮家圣药”之称<sup>[3]</sup>，为双黄连口服液、双黄连粉针剂、清热解毒口服液、连草解热口服液、银翘解毒冲剂等中药制剂的主要原料<sup>[4]</sup>。我国盛产连翘，主产于我国山东、江苏、山西、河南、湖北和四川等省<sup>[5]</sup>，其资源丰富、成分复杂、生物活性多样，不仅是临床常用的中药材，也是许多中成药的重要原料，还可用于园林绿化、水土保持、食品、化妆品等<sup>[6]</sup>。

连翘苷 (phillyrin, PHN) 是连翘的主要药理活性成分之一，属于双环氧木脂素糖苷类化合物，是苯丙氨酸在一系列酶促催化下经松柏醇合成连翘脂素 (phillygenin, PG)，再经糖基化修饰而成<sup>[7]</sup>，具有多种药理活性和广泛的应用价值<sup>[8]</sup>。现代药理研究表明，连翘苷具有抗氧化、抗炎、抗菌、抗病毒、调血脂、强心利尿、解热等活性<sup>[9-11]</sup>，在医药卫生、保健品行业有良好的应用前景。连翘苷等化合物是连翘属植物的标志性代谢产物，但其生物合成的功能基因仍属未知<sup>[12]</sup>，限制了其合成生物学的发展和开发利用，因此阐明其生物合成及调控机制具有重要意义和应用价值。

糖基转移酶 (glycosyltransferase, GT) 是一类

以活化的糖基供体为底物，通过催化蛋白质、脂质、激素或苯丙烷化合物等多种物质的糖基化，调节其生化特性与亚细胞定位的酶<sup>[13-14]</sup>。糖基转移酶根据氨基酸序列相似性、底物特异性与催化特异性，被划分为 110 个不同的家族 (CAZy, <http://www.cazy.org>)，其中糖基转移酶 1 号家族 (GT family 1, *GT1*) 基因成员数量最多、功能最重要，又称为 UDP-糖基转移酶 (UDP-glycosyltransferase, *UGTs*) 基因家族。UGT 一般以尿苷二磷酸-糖 (UDP-糖) 作为糖基供体，催化糖分子转移到受体分子上，从而调节受体分子生物活性，水溶性和稳定性等。在调节植物激素平衡、内外源物质的解毒以及防御反应和次生代谢产物的修饰方面发挥着重要作用<sup>[15]</sup>。植物 UGT 往往是植物种属特有苷类化合物下游合成的关键酶。到目前为止，由于解析的晶体结构数量较少，对 UGT 家族酶的了解仍然有限。鉴于 UGT 家族糖基转移酶的重要性，对 UGT 家族基因的克隆及功能鉴定仍然具有重要的意义<sup>[16]</sup>。

本研究从连翘基因组和转录组中鉴定到 82 个 *FsUGTs* 基因，并进行生物信息学分析。利用实时荧光定量 PCR 分析了 *FsUGTs* 基因在连翘不同组织 (根、茎、叶、果) 的表达特性，利用 HPLC 分析了连翘苷在不同组织 (根、茎、叶、果) 的含量，筛选得到可能参与到连翘脂素糖基化形成连翘苷的基因，为解析连翘中连翘苷生物合成的分子机制提供研究基础。

## 1 材料

研究所用材料于 8 月上旬采自河南洛阳市宜阳

县连翘种植基地,由河南科技大学侯典云教授鉴定为连翘 *F. suspensa* (Thunb.) Vahl。采集样品包括幼根、茎、叶和果,取样后立即放入液氮中冷冻,再置于 $-80\text{ }^{\circ}\text{C}$ 冰箱保存备用。对照品连翘苷(批号B20725)购自上海源叶生物科技有限公司,质量分数大于98%。

## 2 方法

### 2.1 *FsUGTs* 的全基因组鉴定和系统发育树的构建

连翘的基因组数据从 NCBI (<https://www.ncbi.nlm.nih.gov/>), ID:PRJNA809471 下载<sup>[17]</sup>。连翘叶片与果实转录组数据由课题组于2021年10月在美吉生物公司 (<https://www.majorbio.com/>) 通过 RNA 测序 (RNA-Seq) 技术获得。从连翘转录组序列中筛选获得连翘 *FsUGTs* 基因序列,使用 ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) 在线分析预测基因序列的开放阅读框 (open reading frame, ORF), 获得连翘 *FsUGTs* 基因的序列全长,从 Pfam 数据库 (<http://pfam.xfam.org/>) 下载 UGT 保守结构域 (PF00201)<sup>[18]</sup> 的隐藏马尔可夫模型 (HMM) 文件,并对连翘蛋白数据进行了 HMM 搜索,  $E$  值设置为  $1 \times 10^{-10}$ , 结合 CDD (<https://www.ncbi.nlm.nih.gov/cdd>) 分析 *FsUGTs* 的保守结构域以获取 *FsUGTs*。使用 ExPASy5 的 ProtParam 工具评估了这些蛋白质序列的物理化学性质<sup>[19]</sup>, 并利用 Busca 在线网站 (<http://busca.biocomp.unibo.it/>) 预测蛋白质亚细胞定位<sup>[20]</sup>。为了分析 UGT 蛋白序列的进化关系,本研究通过查阅文献并在 NCBI 上查找,选择了属于 A~N 组的 18 个拟南芥 UGTs, 属于 O、P 和 Q 组的 4 个玉米 UGTs, 以及属于 R 组的 2 个山茶花 UGTs。随后,在 MEGA 11 中使用 Clustal W 方法对 *FsUGTs* 与这些已识别的分组 UGTs 进行序列比对,并使用邻接方法生成进化树, bootstrap 值设置为 1 000, 其他条件不变<sup>[21]</sup>。进化树在 itol 在线 URL (<https://itol.embl.de/>) 上进行了可视化<sup>[22]</sup>。

### 2.2 染色体定位

*FsUGTs* 基因的染色体位置信息和染色体长度来源于 *Forsythia suspensa* 基因组。并使用 TB-tool<sup>[23]</sup> 软件进行染色体定位。

### 2.3 顺式调节元件分析

使用 TBtool<sup>[23]</sup> 从起始密码子获得 *FsUGTs* 上游的 2 000 bp 序列。在 PlantCare 网站 (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 上检测到启动子序列的顺式调控元件<sup>[24]</sup>, 随后使用

TBtoolsII 可视化这些元件。

### 2.4 基序、结构域和基因结构分析

通过在线网站 Meme (<http://meme-suite.org/tools/meme>) 分析 *FsUGTs* 的保守基序, 参数设置保持不变<sup>[25]</sup>。 *FsUGTs* 的保守结构域由 NCBI 的批量 CDD 搜索工具提取 (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)<sup>[26]</sup>, 并使用 TBtools 软件可视化。使用 TBtools 的基因结构可视化程序分析 *FsUGTs* 基因的内含子和外显子区域。

### 2.5 不同组织中连翘苷含量检测

参考《中国药典》2025 年版<sup>[1]</sup> 下连翘苷的测定方法, 每个样品重复 3 次。选择无菌研钵在液氮中将连翘的根、茎、叶、果实 (鲜样) 充分研磨, 然后称取 0.5 g 连翘样品粉末用 15 mL 70% 的甲醇于超声清洗仪 ( $60\text{ }^{\circ}\text{C}$ , 30 min) 进行萃取, 并用 70% 甲醇补足缺失质量, 混匀后,  $4\text{ }^{\circ}\text{C}$ 、4 000 r/min 离心 10 min, 上清液用  $0.22\text{ }\mu\text{m}$  有机滤膜过滤到 1 mL 棕色液相色谱进样小瓶上样检测。按照《中国药典》2025 年版方法, 以连翘苷为对照品, 以质量浓度为横坐标 ( $X$ ), 峰面积为纵坐标 ( $Y$ ), 绘制标准曲线方程为  $Y=0.067\ 3\ X-1.5265$ ,  $R^2=1.000\ 0$ 。根据《中国药典》2025 年版方法测定各部位中连翘苷的含量, 并使用 prism 软件进行可视化分析。

### 2.6 总 RNA 提取、cDNA 合成和 qRT-qPCR

使用多糖多酚植物 RNA 提取试剂盒 (RC411-01, 南京诺唯赞有限公司) 提取连翘 RNA。用 HiScript II 1st Strand cDNA Synthesis Kit (+gDNA wiper) (R212-02, 南京诺唯赞) 进行逆转录反应合成 cDNA。使用 Primer Premier 5.0 软件设计引物。*FsUNK1* 基因用作内参基因 (表 1), ChamQ Blue Universal SYBR qPCR Master Mix (Q312-02, 南京诺唯赞科技股份有限公司)。体系如下:  $2 \times$  ChamQ Blue Universal SYBR qPCR Master Mix  $5\ \mu\text{L}$ , 上下游引物各  $0.5\ \mu\text{L}$ , 模板  $2\ \mu\text{L}$ , RNase-free ddH<sub>2</sub>O  $2\ \mu\text{L}$ , 用于 RT-qPCR 实验。荧光定量 PCR 反应程序为:  $95\text{ }^{\circ}\text{C}$  预变性 5 min;  $95\text{ }^{\circ}\text{C}$  变性 30 s;  $58\text{ }^{\circ}\text{C}$  退火 30 s;  $72\text{ }^{\circ}\text{C}$  延伸 1 min; 循环 35 个循环。 $72\text{ }^{\circ}\text{C}$  终延伸 10 min; 重复每个反应进行 3 次, 使用  $2^{-\Delta\Delta Ct}$  法计算基因相对表达量。使用 TBtools 对 33 个目标基因的原始相对表达量进行最小-最大归一化 (Min-Max Normalization) 处理, 按单个基因独立计算: 以每个基因在所有待分析组织中的最小表达量为基准 (归一化为 0), 最大表达量为峰值 (归一化

表1 *FsUNK1*、*FsUGTs* 基因 RT-qPCR 引物Table 1 RT-qPCR primers for *FsUNK1* and *FsUGTs* genes

基因 ID	引物名称	引物序列 (5'-3')
<i>FsUNK1</i>	<i>FsUNK1</i> -F	CAGACCAGCTTTGAGGAGTATC
	<i>FsUNK1</i> -R	GGCCAGAAACCAGTAGTCAATA
<i>FsUGT1</i>	<i>FsUGT1</i> -F	GGACCAACAATACCATCCTTCT
	<i>FsUGT1</i> -R	ATCCTGATGGCTTCGAGTTAAG
<i>FsUGT2</i>	<i>FsUGT2</i> -F	ACTCGACCCTCGAAGGAATA
	<i>FsUGT2</i> -R	ATCCAGCCTTGAGCACATTTA
<i>FsUGT3</i>	<i>FsUGT3</i> -F	CTCTCTTGTGTTCTGGCATCA
	<i>FsUGT3</i> -R	GCAACTGAGCCTTGGTAAATTC
<i>FsUGT4</i>	<i>FsUGT4</i> -F	GGCGTAGGGATGAGGATTATG
	<i>FsUGT4</i> -R	TCTTCTTCTGCACTGCCTTAC
<i>FsUGT5</i>	<i>FsUGT5</i> -F	TGTCCAAGGGTCACACAATC
	<i>FsUGT5</i> -R	GAAAGAGCGGCAGAGATGAA
<i>FsUGT6</i>	<i>FsUGT6</i> -F	CAGTCTCTCTCCGACGATGATA
	<i>FsUGT6</i> -R	TCTGGCAAAGGAGACGAATAAA
<i>FsUGT7</i>	<i>FsUGT7</i> -F	CGCAAGGACACCATCGAATA
	<i>FsUGT7</i> -R	CCTCGAAAGCCCGGTAAAT
<i>FsUGT8</i>	<i>FsUGT8</i> -F	GTGGATTTGTCCGAAGGTAGAA
	<i>FsUGT8</i> -R	CCACCCACTATGAGTCAAGAAC
<i>FsUGT9</i>	<i>FsUGT9</i> -F	GTTAGACTCGAAGCCGAAGAG
	<i>FsUGT9</i> -R	GGCCTCCCACCTTCTACTAATG
<i>FsUGT10</i>	<i>FsUGT10</i> -F	CCGGCTAAAGTTCTCGTGTAT
	<i>FsUGT10</i> -R	GACAGCGCACACTCCTAAA
<i>FsUGT11</i>	<i>FsUGT11</i> -F	GACTCAACGCCACCTTTCT
	<i>FsUGT11</i> -R	TGTCCTCACCATCTCCTCTATT
<i>FsUGT12</i>	<i>FsUGT12</i> -F	GCAACCCTCCTCATGTGTTA
	<i>FsUGT12</i> -R	TCCGGGAACCCAGTCTATAA
<i>FsUGT13</i>	<i>FsUGT13</i> -F	CAGGGAAGAAGTTGCAGAGAA
	<i>FsUGT13</i> -R	GCGTTCTGCAGTATTGTCCTA
<i>FsUGT14</i>	<i>FsUGT14</i> -F	CACAAGTGGCCGTCTTATCT
	<i>FsUGT14</i> -R	CATTGGCCATGCTGCTATTG
<i>FsUGT15</i>	<i>FsUGT15</i> -F	CAGCCACGGTGTGGATTTA
	<i>FsUGT15</i> -R	GGAAGTCTAGGCAGTTGGATTG
<i>FsUGT16</i>	<i>FsUGT16</i> -F	TGGTGGCTAATGGGCTAATG
	<i>FsUGT16</i> -R	CGATCAACCCTCGATCCTTTAC
<i>FsUGT17</i>	<i>FsUGT17</i> -F	AAGTACTTCTTGTGGGTGGTTAG
	<i>FsUGT17</i> -R	GTGGACACCATGACACGATTA
<i>FsUGT18</i>	<i>FsUGT18</i> -F	CCTGTGATTTGTTGGCCTTTC
	<i>FsUGT18</i> -R	TCCACCTCATCTTGCTTCAC
<i>FsUGT19</i>	<i>FsUGT19</i> -F	CAAGCGCCTCGAAGATGATA
	<i>FsUGT19</i> -R	ACTGACTCGGATTCTCGTTTATT
<i>FsUGT20</i>	<i>FsUGT20</i> -F	CCTGCCATTGGAGGATTCTT
	<i>FsUGT20</i> -R	CCACCACAAGTTCCGATTTG
<i>FsUGT21</i>	<i>FsUGT21</i> -F	CCCGACTTGGTTGTAGACAATA
	<i>FsUGT21</i> -R	TCTTGTGGAGCCCAACTAAC
<i>FsUGT22</i>	<i>FsUGT22</i> -F	TTGGAGGAGACTAAAGGGAGAG
	<i>FsUGT22</i> -R	CACCCACAGTGTGTCAAGAA
<i>FsUGT23</i>	<i>FsUGT23</i> -F	GGACTTGAACCTCACCGATAGAC
	<i>FsUGT23</i> -R	CTCGATGGTGCCTCTTTCT
<i>FsUGT24</i>	<i>FsUGT24</i> -F	TCCGACCGATTTCTCCAATAAC
	<i>FsUGT24</i> -R	GGGCTAGACTGCCAAATGATAC
<i>FsUGT25</i>	<i>FsUGT25</i> -F	GGCCTAGAAGCCTCAAACATATC
	<i>FsUGT25</i> -R	GCCTCTTTCCTTGTCTTTCC
<i>FsUGT26</i>	<i>FsUGT26</i> -F	CAGGTGTTCCCATGCTTACT
	<i>FsUGT26</i> -R	CATCTCGTTTCGCCAAGTTTC
<i>FsUGT27</i>	<i>FsUGT27</i> -F	TGCATCTTGACCAGCCTATG
	<i>FsUGT27</i> -R	TCGTGATCACTTTCGCTATTT
<i>FsUGT28</i>	<i>FsUGT28</i> -F	GTCTGAAAGCTCCTGCCATAA
	<i>FsUGT28</i> -R	GGAGAGGCCCTATTGAGTAAATG
<i>FsUGT29</i>	<i>FsUGT29</i> -F	GAAACCCGCATGTCAAACAG
	<i>FsUGT29</i> -R	GATCCACCACAAGGGCTAAA
<i>FsUGT30</i>	<i>FsUGT30</i> -F	CTGGTGACCTTTTGGCTTAT
	<i>FsUGT30</i> -R	GTCCAACCGTGGCTAAGAAT
<i>FsUGT31</i>	<i>FsUGT31</i> -F	GGGAATCGCGGAAGCTATAA
	<i>FsUGT31</i> -R	TCCCATCAACCTCGCTATTG
<i>FsUGT32</i>	<i>FsUGT32</i> -F	GAGCTGGTACGAAGCAGAAA
	<i>FsUGT32</i> -R	AGGGTCTATGCCATTAATAATC
<i>FsUGT33</i>	<i>FsUGT33</i> -F	CAGGTCAGGGCCACATAAAT
	<i>FsUGT33</i> -R	ACATACGACGGATTGCAGATAG

为1)，将单个基因在各组织中的表达量线性缩放至[0, 1]区间以消除不同基因间原始表达量的量级差异，并对结果可视化处理。用 ChiPlot 对 RT-qPCR 结果与连翘苷含量进行相关性分析并可视化<sup>[27]</sup>。

### 3 结果与分析

#### 3.1 *FsUGTs* 蛋白的序列采集和理化表征

通过对连翘转录组与基因组分析获得了134个 UGT 蛋白序列，从 Pfam 数据库 (<http://pfam.xfam.org/>) 下载了 UGT 蛋白保守结构域 (PF00201)<sup>[27]</sup> 的隐藏马尔可夫模型 (HMM) 文件，并对连翘蛋白数据进行了 HMM 搜索，*E* 值设置为  $1 \times 10^{-10}$ ，结合 CDD (<https://www.ncbi.nlm.nih.gov/cdd>) 分析 *FsUGTs* 的保守结构域以获取 *FsUGTs*。并删除了没有保守结构域的序列以获得最终的82个 *FsUGTs* 序列，命名为 *FsUGT1*~*FsUGT82*。

ProtParam 工具分析结果显示，82个 *FsUGTs* 氨基酸长度为385~558 aa，相对分子质量为43 140~

60 750 (平均52 460)，理论等电点为4.81~8.26 (平均5.74)；其中6个 *FsUGTs* (*FsUGT22*、*FsUGT33*、*FsUGT44*、*FsUGT55*、*FsUGT76*、*FsUGT78*) 的疏水指数 (GRAVY) 为正值，其余76个为负值，表明多数 *FsUGTs* 蛋白呈亲水性。亚细胞定位预测显示，91.4% (75个) 的 *FsUGTs* 定位于细胞膜，2.4% (2个) 定位于细胞质，3.6% (3个) 定位于叶绿体，2.4% (2个) 定位于线粒体内膜。

#### 3.2 *FsUGTs* 蛋白的系统发育树分析

将82个 *FsUGTs* 与18个拟南芥 UGTs、4个玉米 UGTs、2个山茶花的 UGTs 进行联合系统进化分析，结果显示 (图1)，82个 *FsUGTs* 可明确分为17个组，对应拟南芥 A~N 组、玉米 O/P 组及山茶花 R 组，各节点 bootstrap 值均大于95，置信度较高。分组统计表明，L 组成员数量最多 (16条)，其次为 A 组 (10条) 和 E 组 (9条)；Q 组未检测到 *FsUGTs* 分布，提示连翘 *FsUGTs* 家族在进化过程中可能发生了 Q 组的丢失或分化。

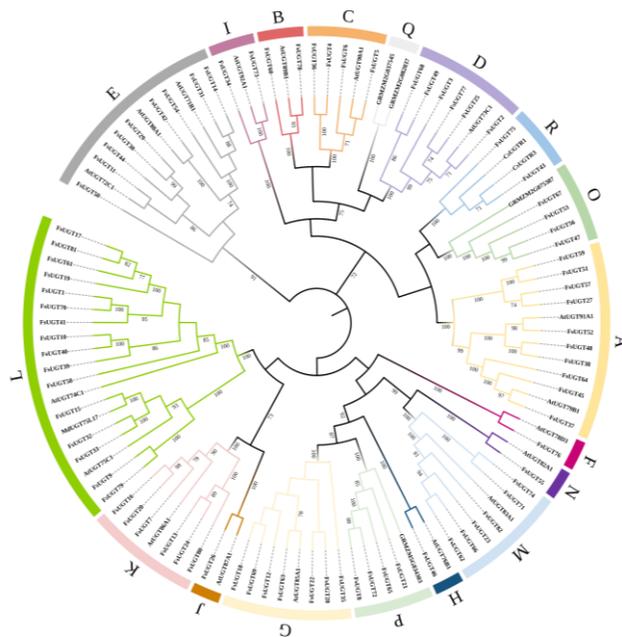


图1 连翘系统发育树分析

Fig. 1 Phylogenetic tree analysis of *F. suspensa*

#### 3.3 *FsUGTs* 的染色体定位

基于连翘基因组染色体信息，通过 TBtools 绘制 *FsUGTs* 染色体定位图 (图2)，结果显示，82个 *FsUGTs* 不均匀分布于14条染色体上，且多数呈簇状分布。其中，Chr10上分布最多 (16个)，Chr11、Chr12、Chr13各分布7个，Chr3上分布最少 (仅2个)；染色体上的簇状分布特征提示，*FsUGTs* 家族

可能通过基因复制事件实现扩张，且复制基因倾向于保留在同一染色体或同源染色体上。

#### 3.4 *FsUGTs* 启动子中的顺式作用元件

对 *FsUGTs* 启动子上游2 000 bp 序列的顺式作用元件预测结果显示 (图3)，共检测到24类顺式作用元件，其中光响应元件分布最多 (73个 *FsUGTs*)，表明 *FsUGTs* 对皂苷代谢的调控可能主

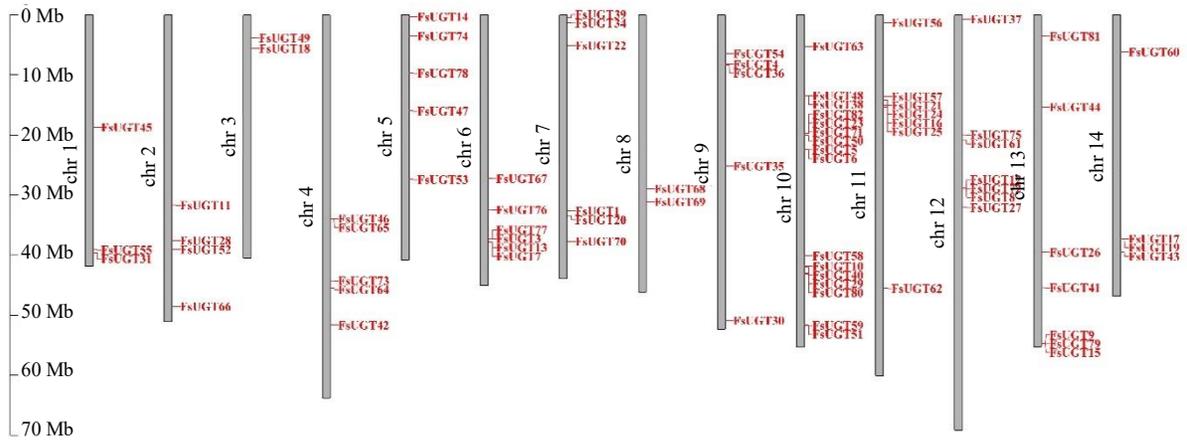


图 2 *FsUGTs* 的染色体定位

Fig. 2 Chromosomal localization of *FsUGTs*



图 3 *FsUGTs* 启动子中的顺式调节元件

Fig. 3 Cis-regulatory elements in promoters of *FsUGTs*

要受光信号诱导；其次为茉莉酸甲酯（MeJA）响应元件（55个 *FsUGTs*），提示 MeJA 可能参与 *FsUGTs* 的表达调控，这与 MeJA 在植物次生代谢中的核心调节作用一致<sup>[38-39]</sup>。此外，脱落酸响应元件、赤霉素响应元件、生长素响应元件及低温响应元件也有较多分布，表明 *FsUGTs* 的表达可能受多种激素及环境胁迫信号的协同调控。

### 3.5 *FsUGTs* 的保守基序、结构域和基因结构

MEME 保守基序分析结果显示（图 4），82 个 *FsUGTs* 共含 10 个保守基序（Motif 1~Motif 10），长度为 15~29 aa；其中 Motif 1 与 Motif 4 共同构成 UGT 家族特征性的 PSPG box 结构域，该结构域是 UGT 结合 UDP-糖供体的关键区域，进一步证实所鉴定的 *FsUGTs* 均属于 UGT 家族。分组对比发

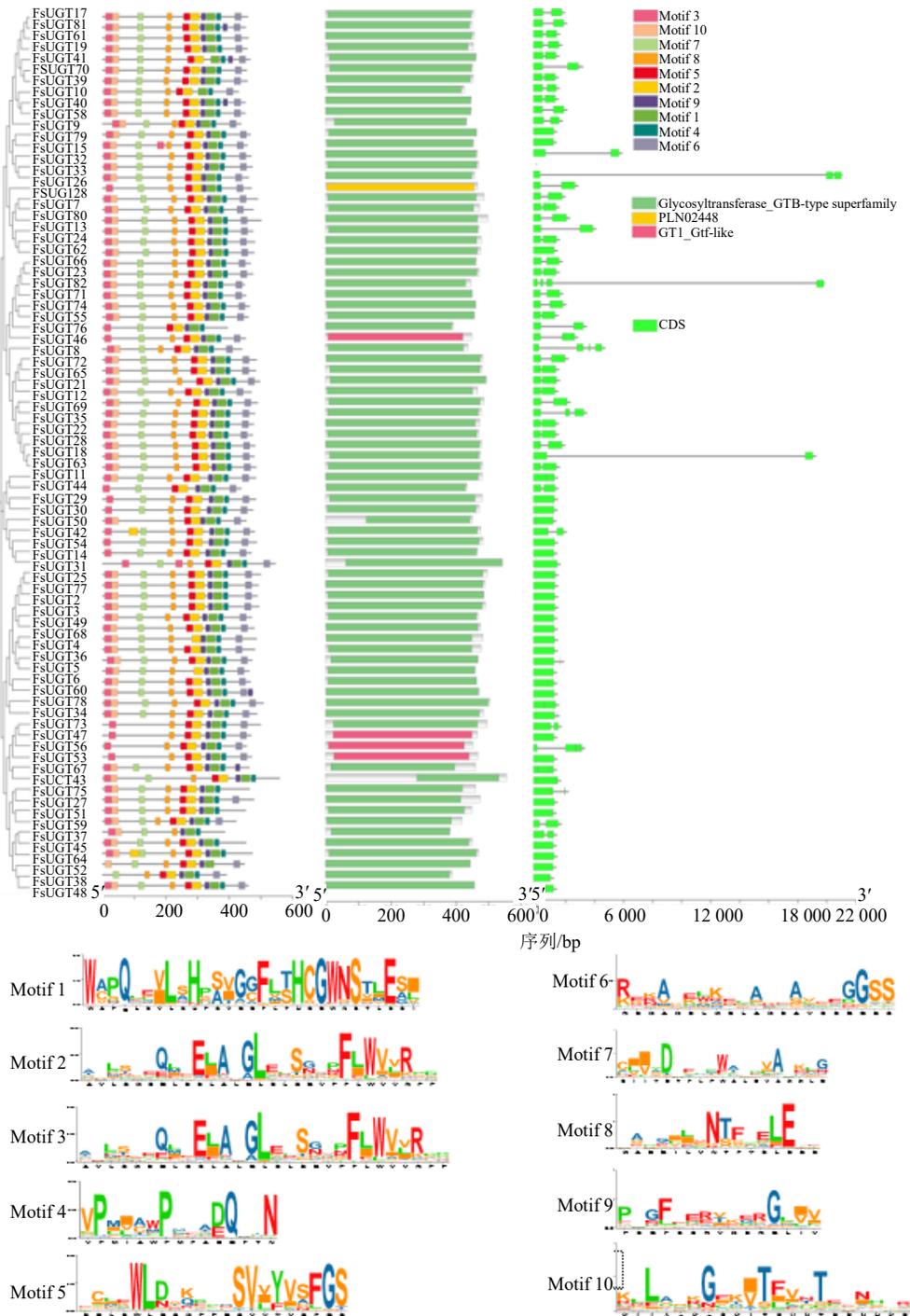


图 4 *FsUGTs* 的保守基序、结构域和基因结构  
Fig. 4 Conserved motifs, domains, and gene structures of *FsUGTs*

现, E、F、J、N 组缺失 Motif 10, O 组缺失 Motif 7 与 Motif 10, A、R 组多数成员缺失 Motif 6; 保守基序的分组特异性缺失可能导致不同组 *FsUGTs* 的底物特异性或催化功能差异。结构域分析显示 (图 4): 所有 82 个 *FsUGTs* 均含 3 个保守结构域 GT1\_Gtf-like、PLN02448 及 Glycosyltransferase\_GTB-type superfamily, 且 3 个结构域均属于 GT-B 折叠类型; 这表明 *FsUGTs* 家族具有典型的 GT-B 拓扑结构, 该结构是核苷酸-糖基依赖性糖基转移酶的核心拓扑类型之一, 为其催化糖基转移反应提供结构基础。

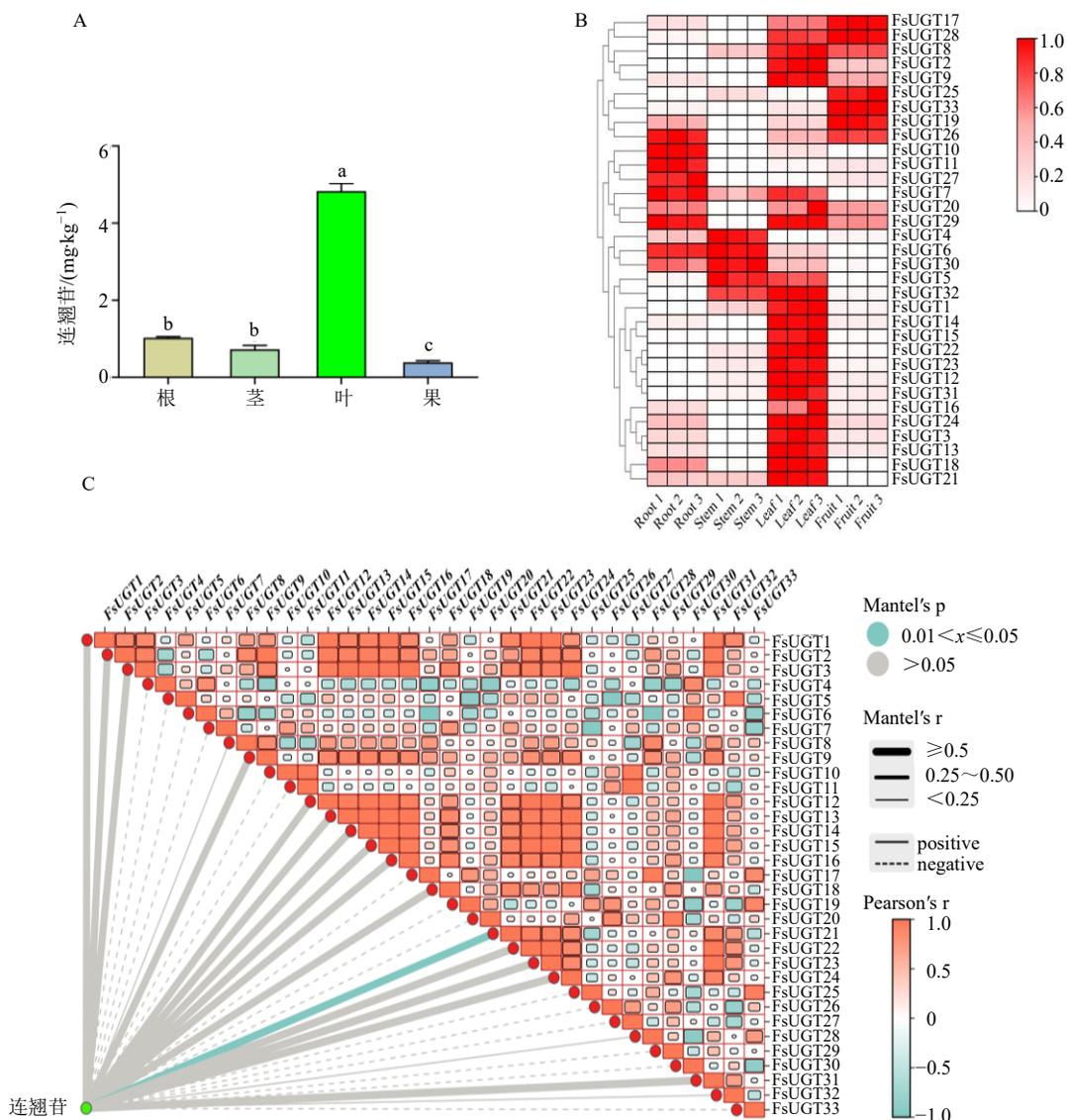
基因结构分析显示 (图 4), *FsUGTs* 的外显子数量为 1~5 个, 其中 83.1% (68 个) 的成员仅含

1~2 个外显子, 属于典型的“少外显子”基因结构; 仅 *FsUGT23* 与 *FsUGT53* 含 5 个外显子, 为家族中 exon 数量最多的成员。外显子-内含子结构的简化特征提示, *FsUGTs* 家族在进化过程中可能通过减少内含子数量提高转录效率, 以快速响应次生代谢需求。

### 3.6 连翘苷合成途径中的 *FsUGTs* 的筛选

为了进一步研究 *FsUGTs* 对连翘脂素的催化作用, 使用比较转录组获得了 33 个在叶片高表达的 *FsUGTs*, 分别对其进行组织差异性表达, 同时测定连翘苷在各组织 (根、茎、叶、果) 中的分布, 最后, 用皮尔逊相关性系数绘制热图。

结果发现 (图 5-A), 在连翘苷的 HPLC 检测



A-不同部位中连翘苷含量, 不同字母表示差异显著,  $P < 0.05$ ; B-qRT-PCR 结果; C-相关性分析。

A- content of forsythin in different parts is significantly different with different letters,  $P < 0.05$ ; B-qRT-PCR results; C-correlation analysis.

图 5 连翘苷合成相关 *FsUGTs* 的筛选

Fig. 5 Screening of *FsUGTs* related to forsythin biosynthesis

中,发现叶中的连翘苷含量最高,显著高于根、茎和果实,其次是根,而果实中连翘苷含量最低。很明显,候选 *FsUGTs* 的基因表达模式应与连翘苷含量的分布具有协同作用,连翘苷含量高的叶片部位候选 *FsUGTs* 应同样高表达。

qRT-PCR 结果显示(图 5-B),33 个 *FsUGT* 中,大多数在叶片中的表达量都显著高于在果实中的表达量,证明转录组数据的合理性。相关性分析结果显示(图 5-C),*FsUGT1*~*FsUGT3*、*FsUGT9*、*FsUGT12*~*FsUGT16*、*FsUGT18*、*FsUGT21*~*FsUGT24* 和 *FsUGT31* 与连翘苷组织分布的皮尔逊相关性系数大于 0.8,为正向强相关,因此,选择了 *FsUGT1*~*FsUGT3*、*FsUGT9*、*FsUGT12*~*FsUGT16*、*FsUGT18*、*FsUGT21*~*FsUGT24* 和 *FsUGT31* 作为催化连翘脂素合成连翘苷的候选基因。其中 *FsUGT1* 与欧洲橄榄变种 *Sylvestris* 的 *UGT74F2* 和 *UGT75L17* 基因具有高度相似性。据报道 *UGT75L17* 是苹果中的根皮素 4'-*O*-糖基转移酶,可以催化根皮素转化为三叶草苷<sup>[29]</sup>。表明 *FsUGT1* 在催化 4'-*O*-糖基化位点的潜力。

#### 4 讨论

糖基化是植物的普遍修饰反应,经修饰获得的糖基化产物具有不同的生物学功能,使得糖基化反应成为影响及调节植物生长代谢平衡的重要机制<sup>[13]</sup>。目前,苯丙烷代谢途径糖基化修饰的研究多集中于黄酮类化合物,如柑橘<sup>[30]</sup>、罗勒<sup>[31]</sup>、欧芹<sup>[32]</sup>等,而关于木脂素类相关的 UGT 的研究较少,仅在菘蓝等物种中有少量鉴定<sup>[33-34]</sup>。本研究围绕连翘木脂素(连翘苷)的糖基化合成,系统鉴定并分析 *FsUGTs* 家族,为木脂素类化合物的糖基化机制研究提供了新视角。

本研究中,本课题组首先从 *F. suspensa* 基因组和转录组中鉴定出 82 个 *FsUGTs*,并研究了它们的理化特性。根据 *FsUGTs* 与其他物种 UGTs 的系统发育关系,*FsUGTs* 被分为 17 组。关于 UGT 家族的分类已有很多研究,Barvkar 等<sup>[35]</sup>对拟南芥 UGTs 和亚麻 UGTs 的蛋白质序列进行聚类分析,并将其分为 14 个主要亚科(A~N)。Yao 等<sup>[36]</sup>将淫羊藿的 UGT 分为 16 个组(A~Q,无 R 组)。Cui 等<sup>[37]</sup>则在茶树中鉴定出 1 个新的 R 类群。本实验通过与已鉴定的 UGT(来自拟南芥的 18 个 UGTs(A~N)、来自玉米的 4 个 UGTs,以及属于山茶花 UGTs 稻的 2 个 UGT)构建了进化树。从而将 82 个 *FsUGTs* 分为 17 组(图 1),Q 类群中没有 *FsUGT*。这表明

不同物种之间的 UGT 存在差异,并且随着研究的深入和扩大,UGT 家族不断增加新的类群。这为后续解析植物 UGT 家族的分化规律、保守性提供了新的物种案例。

*FsUGTs* 的亚细胞定位预测表明大部分 UGT 定位于细胞膜、这或许与糖基化后产物运输与储存有关<sup>[38]</sup>。染色体定位结果表明 82 个 *FsUGTs* 基因成簇分布,不均一地分布在 14 条染色体上,这在其他物种中也是如此<sup>[39]</sup>。顺式作用元件分析表明 *FsUGTs* 主要受到光调节,因此可以推测,光信号可能是影响连翘 UGT 基因转录的主要因素;MeJAs 在 *FsUGTs* 调控中同样起到重要作用,而这也与 MeJAs 在植物次生代谢中的重要作用相符合<sup>[40-41]</sup>。

连翘苷是连翘的主要药理活性成分之一,对其合成通路进行研究对提高连翘中连翘苷的含量、获得高质量连翘药材有重要意义。HPLC 检测发现叶中的连翘苷显著高于根、茎和果实。这与金宗元等<sup>[42]</sup>的结果相一致。为此,通过比较转录组选择在叶片中高表达的 UGTs 进行 qRT-PCR 检测。结果表明,*FsUGT1*-*FsUGT3*、*FsUGT9*、*FsUGT12*-*FsUGT16*、*FsUGT21*-*FsUGT24* 和 *FsUGT31* 的表达量分布与连翘苷在不同组织(根、茎、叶、果)含量的分布高度一致,因此,推测其可能在连翘脂素糖基化形成连翘苷的过程中发挥重要作用。

综上所述,本研究对连翘 *FsUGTs* 基因家族进行了系统鉴定,揭示其家族进化关系、结构多样性、进化保守性与调控潜力,筛选出可能参与连翘苷合成的关键基因,填补了木脂素类化合物糖基化相关 UGT 基因系统研究的空白;这不仅为连翘药用成分的合成机制研究提供了理论依据、也为后续研究提供了明确的探索方向与关键启示,如聚焦光信号与 MeJAs 对连翘苷合成相关 *FsUGTs* 的调控机制、通过体外酶活实验与体内功能验证,明确其在连翘苷糖基化合成中的真实功能等。相信随着科学技术的发展,多组学联合分析(转录组、代谢组、蛋白质组与表观组深度融合)、基因编辑等技术的广泛应用将会进一步解析药用成分合成通路,使通过基因工程手段改良连翘品质、提高连翘苷等药用成分含量、或是通过异源表达系统实现连翘苷的体外规模化发酵生产成为现实。

**利益冲突** 所有作者均声明不存在利益冲突

## 参考文献

- [1] 中国药典 [S]. 一部. 2025: 178.
- [2] 钱心悦, 陈星蕊, 张腾腾, 等. 连翘药用品种及其部位本草考证 [J]. 甘肃中医药大学学报, 2023, 40(6): 108-113.
- [3] 康梦娇, 张伟, 赵林华, 等. 重构本草—连翘 [J]. 吉林中医药, 2023, 43(3): 331-334.
- [4] 白映伟, 胡宏. 连翘的多种利用 [J]. 甘肃科技, 2012, 28(9): 147-148.
- [5] 杜一凡, 谷子玉, 钟岳, 等. 基于 HPLC 指纹图谱和多指标定量的连翘花质量评价研究 [J]. 中草药, 2025, 56(21): 7956-7963.
- [6] 冯治朋, 高秀强, 韩颜超, 等. 连翘的研究进展 [J]. 现代农业科技, 2018(12): 60-64.
- [7] Yuan W J, Zhang S P, He Z Y, *et al.* Comparative transcriptome analyses identify genes involved into the biosynthesis of forsythins and forsythoside A in *Forsythia suspensa* [J]. *Funct Integr Genomics*, 2022, 22(5): 731-741.
- [8] 房辉, 胡倩莲, 李竹君. 连翘中连翘苷提取方法的比较 [J]. 广东化工, 2021, 48(13): 61-62.
- [9] 陈星蕊, 汤瑜晨, 庄家蝶, 等. 连翘不同部位化学成分及药理作用研究进展 [J]. 甘肃中医药大学学报, 2024, 41(06):53-67.
- [10] 孙玲, 孙楠, 徐攀, 等. 连翘化学成分及抗呼吸道病毒作用机制研究进展 [J]. 中医药导报, 2025,31(03):146-154.
- [11] Deng R X, Zhou S N, Yang X, *et al.* Two new furofuran lignan glycosides from *Forsythia suspensa* leaves [J]. *Phytochem Lett*, 2021, 41: 34-37.
- [12] 刘津, 于思礼, 马雅婷, 等. 天然木脂素的代谢工程和合成生物学研究 [J]. 中草药, 2016, 47(14):2556-2562.
- [13] 于安东, 刘琳, 龙瑞才, 等. 植物 UDP-糖基转移酶 (UGT) 的功能及应用前景 [J]. 植物生理学报, 2022, 58(4): 631-642.
- [14] Speeckaert N, Adamou N M, Hassane H A, *et al.* Characterization of the UDP-glycosyltransferase UGT72 family in poplar and identification of genes involved in the glycosylation of monolignols [J]. *Int J Mol Sci*, 2020, 21: 5018.
- [15] 姚英俊. 苦荞糖基转移酶基因 FtUGT163 的克隆及功能分析 [D]. 成都: 四川农业大学, 2021.
- [16] 马迎新, 商金海, 于慧珊, 等. 苹果 GT1 家族 A 组中 8 个糖基转移酶基因克隆及蛋白表达 [J]. 聊城大学学报自然科学版: 2025, 38(2): 307-321.
- [17] Jia Y, Chen S, Chen W, *et al.* A chromosome-level reference genome of Chinese balloon flower (*Platycodon grandiflorus*) [J]. *Front Genet*, 2022, 13: 869784.
- [18] Zhang H, Hua X, Zheng D, *et al.* De novo biosynthesis of oleanane-type ginsenosides in *Saccharomyces cerevisiae* using two types of glycosyltransferases from *Panax ginseng* [J]. *J Agric Food Chem*, 2022, 70(7): 2231-2240.
- [19] Gasteiger E, Gattiker A, Hoogland C, *et al.* ExPASy: The proteomics server for in-depth protein knowledge and analysis [J]. *Nucleic Acids Res*, 2003, 31(13): 3784-3788.
- [20] Savojardo C, Martelli P L, Fariselli P, *et al.* BUSCA: An integrative web server to predict subcellular localization of proteins [J]. *Nucleic Acids Res*, 2018, 46(W1): W459-W466.
- [21] Kumar S, Stecher G, Li M, *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms [J]. *Mol Biol Evol*, 2018, 35(6): 1547-1549.
- [22] Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation [J]. *Nucleic Acids Res*, 2021, 49(W1): W293-W296.
- [23] 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(7): 1194-1202.
- [24] Lescot M, Dehais P, Thijs G, *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences [J]. *Nucleic Acids Res*, 2002, 30(1): 325-327.
- [25] Nystrom S L, McKay D J. Memes: A Motif analysis environment in R using tools from the MEME suite [J]. *PLoS Comput Biol*, 2021, 17(11): e1008991.
- [26] Wang J, Chitsaz F, Derbyshire MK, *et al.* The Conserved Domain Database in 2023 [J]. *Nucleic Acids Res*, 2023, 51(D1): D384-D388.
- [27] Fan J, Rao W, Peng D, *et al.* Genome-wide identification of UGT gene family and functional analysis of PgUGT29 in platycodon grandifloras [J]. *Int J Mol Sci*, 2025, 26(8): 4832.
- [28] Li X H, Li J R, Zhao Q C, *et al.* Physiological, biochemical, and genomic elucidation of the ensifer adhaerens M8 strain with simultaneous arsenic oxidation and chromium reduction [J]. *J Hazard Mater*, 2023, 441: 129862.
- [29] Pandey R P, Li T F, Kim E H, *et al.* Enzymatic synthesis of novel phloretin glucosides [J]. *Appl Environ Microbiol*, 2013, 79(11): 3516-3521.
- [30] Yuan Z Y, Li G, Zhang H X, *et al.* Four novel Cit7GlcTs functional in flavonoid 7-O-glucoside biosynthesis are vital to flavonoid biosynthesis shunting in citrus [J]. *Hortic Res*, 2024, 11(6): uhae098.
- [31] Shah S, Rastogi S, Akhtar MQ, *et al.* Mining and functional characterization of a flavonoid glucosyltransferase (OtUGT) involved in 7-O-glucosylation of apigenin in

- Ocimum tenuiflorum* [J]. *Plant J*, 2025, 121(6): e70111.
- [32] An S, Yamashita M, Iguchi S, *et al.* Biochemical characterization of parsley glycosyltransferases involved in the biosynthesis of a flavonoid glycoside, apiin [J]. *Int J Mol Sci*, 2023, 24(23): 17118.
- [33] 陈泉. 菘蓝木脂素糖基转移酶 UGT71B5 的功能研究 [D]. 厦门: 华侨大学, 2020.
- [34] 姜银银. 菘蓝新木脂素糖基转移酶的挖掘和体外酶学研究 [D]. 北京: 中国中医科学院, 2022.
- [35] Barvkar V T, Pardeshi V C, Kale S M, *et al.* Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varied expression patterns [J]. *BMC Genomics*, 2012, 13: 175.
- [36] Yao Y Y, Gu J J, Luo Y J, *et al.* Genome-wide analysis of UGT gene family identified key gene for the biosynthesis of bioactive flavonol glycosides in *Epimedium pubescens* Maxim [J]. *Synth Syst Biotechnol*, 2022, 7(4): 1095-1107.
- [37] Cui L, Yao S, Dai X, *et al.* Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*) [J]. *J Exp Bot*, 2016, 67(7): 2285-2297.
- [38] Asada K, Salim V, Masada-Atsumi S. A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in *Madagascar Periwinkle* [J]. *Plant Cell*, 2013, 25(10): 4123-4134.
- [39] Zhao X, Feng Y, Ke D, *et al.* Molecular identification and characterization of UDP-glycosyltransferase (UGT) multigene family in pomegranate [J]. *Horticulturae*, 2023, 9(5): 540-557.
- [40] De Geyter N, Gholami A, Goormachtig S. Transcriptional machineries in jasmonate-elicited plant secondary metabolism [J]. *Trends Plant Sci*, 2012, 17(6): 349-359.
- [41] Ali B. Practical Applications of jasmonates in the biosynthesis and accumulation of secondary metabolites in plants [J]. *Biocatal Agric Biotechnol*, 2021, 38: 102205.
- [42] 金宗元, 赵娟娟, 张全玲, 等. 连翘 *FsUGT83A1* 基因的克隆、生物信息学和表达分析 [J]. *中药材*, 2022, 45(10): 2340-2345.

[责任编辑 时圣明]