麦冬 UDP-糖基转移酶 (UGT) 基因家族鉴定及表达模式分析

郑文芳[1], 张文颖[1], 秦众护[1], 李　铮[2], 傅欣彤[2], 闫海霞[2*], 罗红梅[1*]

(1. 中国医学科学院 北京协和医学院, 药用植物研究所, 道地药材品质保障与资源持续利用全国重点实验室, 北京 100193;　2. 北京市药品检验研究院 (北京市疫苗检验中心), 国家药品监督管理局监管科学创新研究基地, 中药成分分析与生物评价北京市重点实验室, 北京 102206)

摘要: 甾体皂苷是麦冬 *Ophiopogon japonicus* (L. f.) Ker Gawl. 的重要活性成分, 其结构与药效受糖基化修饰影响, 但麦冬糖基转移酶研究较少。本研究完成了麦冬根、茎、叶的 *de novo* 转录组 (无参转录组) 测序及组装、注释, 并结合结构域筛选, 对麦冬尿苷二磷酸依赖型糖基转移酶 (UDP-dependent glycosyltransferase, UGT) 家族开展鉴定与系统发育分析。共获得 109 条 UGT 候选编码序列 (coding sequence, CDS) 并鉴定出 34 个全长 UGT 候选 CDS 序列 (*OjUGT1*～*OjUGT34*)。系统发育分析结果将麦冬与拟南芥 *UGT* 基因家族划分为 15 个亚家族, 其中 D 组 (UGT73 亚家族) 与 G 组 (UGT85 亚家族) 成员相对数量较多。保守基序分析与多序列比对分析显示 31 个 OjUGT 蛋白具有典型的植物次生代谢产物糖基转移酶保守基序 (plant secondary product glycosyltransferase, PSPG)。表达谱分析揭示一些 *OjUGT* 具有明显的组织表达差异, *OjUGT32* (UGT73 亚家族) 在块根中特异高表达, *OjUGT34* (UGT73 亚家族) 在叶片中特异高表达。跨物种比较提示天门冬科内 *UGT73*/*UGT85* 基因存在谱系特异性扩张, 可能与甾体皂苷合成有关。综合系统发育特点、PSPG 保守基序分析、三维结构预测与表达特异性分析结果, 推测 *OjUGT32* 与 *OjUGT34* 可能参与麦冬甾体皂苷糖基化过程。本研究构建了麦冬 *UGT* 基因家族的表达谱, 预测分析了 *OjUGT* 的理化性质与序列特征, 为阐明麦冬皂苷糖基化合成提供候选基因。

关键词: 麦冬; 生物信息学; 糖基转移酶; 皂苷; 基因表达

中图分类号: R931　　**文献标识码**: A　　**文章编号**: 0513-4870(2026)06-2329-10

Identification and expression pattern analysis of UGT family in *Ophiopogon japonicus*

ZHENG Wen-fang[1], ZHANG Wen-ying[1], QIN Zhong-hu[1], LI Zheng[2], FU Xin-tong[2], YAN Hai-xia[2*], LUO Hong-mei[1*]

(*1. State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China;　2. Beijing Institute for Drug Control (Beijing Center for Vaccine Control), NMPA Center for Innovation and Research in Regulatory Science, Beijing Key Laboratory of Analysis and Evaluation on Chinese Medicine, Beijing 102206, China*)

Abstract: Steroidal saponins are major active constituents of Maidong (*Ophiopogon japonicus* (L. f.) Ker Gawl.), and their structures and pharmacological activities are strongly influenced by glycosylation, whereas glycosyltransferases from Maidong remain poorly studied. In this study, we performed *de novo* transcriptome sequencing, assembly and annotation of roots, stems and leaves of *O. japonicus*, and combined domain-based screening to identify and conduct phylogenetic analysis of the UDP-dependent glycosyltransferase (UGT) gene

收稿日期: 2025-11-07;　修回日期: 2026-01-21.
基金项目: 中国医学科学院医学与健康科技创新工程探索研究专项 (2024-I2M-TS-018).
*通讯作者 E-mail: yanhaixia7993@163.com; hmluo@implad.ac.cn
DOI: 10.16438/j.0513-4870.2025-1320

family. In total, we obtained 109 UGT candidate coding sequences (CDSs) and identified 34 full-length UGT CDSs (*OjUGT1* − *OjUGT34*). Phylogenetic analysis grouped OjUGTs together with *Arabidopsis UGTs* into 15 subfamilies, among which clade D (UGT73 subfamily) and clade G (UGT85 subfamily) contained relatively more members. Conserved motif analysis and multiple sequence alignment showed that 31 OjUGT proteins possess the canonical plant secondary product glycosyltransferase (PSPG) motif. Expression profiling revealed pronounced tissue-specific differences in OjUGT transcripts, with OjUGT32 (UGT73 subfamily) highly expressed in tuberous roots and OjUGT34 (UGT73 subfamily) highly expressed in leaves. Comparative analyses across species suggested lineage-specific expansion of *UGT73* and *UGT85* genes within Asparagaceae, which may be associated with steroidal saponin biosynthesis. By integrating phylogenetic relationships, PSPG motif conservation, three-dimensional structure prediction and tissue-specific expression patterns, we propose that OjUGT32 and OjUGT34 are likely involved in the glycosylation of steroidal saponins in *O. japonicus*. Overall, this study establishes the expression profile of the *UGT* gene family in Maidong and predicts the physicochemical properties and sequence features of *OjUGTs*, providing candidate genes for elucidating the glycosylation steps in steroidal saponin biosynthesis.

Key words: *Ophiopogon japonicus*; bioinformatics; glycosyltransferase; saponin; gene expression

麦冬 *Ophiopogon japonicus* (L. f.) Ker Gawl. 为天门冬科沿阶草属多年生草本植物, 其块根入药历史悠久, 首载于《神农本草经》, 被归为上品药材[1], 其干燥块根为中国药典收载的常用中药材, 具有养阴生津、润肺清心的功效, 临床上用于改善肺燥干咳、阴虚痨嗽、喉痹咽痛、津伤口渴、内热消渴、心烦失眠和肠燥便秘等症[2]。化学分析揭示麦冬富含多种生物活性物质, 尤以甾体皂苷及多糖类成分为主[3,4]。其中, 麦冬所含甾体皂苷类成分被中国药典列为质量指标成分[2], 在抗炎、抗癌、抗肿瘤、抗突变、抗肺纤维化、保护心脑血管等方面表现出显著的药理活性[5-8], 相关机制研究为开发新型药物提供了重要依据。近年来, 随着药食同源政策的推进, 2024 年, 麦冬被正式列入中国药食同源的目录, 这将推动其在功能性食品和保健品领域的应用。当前国内流通的麦冬药材主要来源于四川和浙江两大产区[9]。其中, 川产麦冬凭借其生长年限较短、单位面积产出量高以及相对低廉的成本优势而成为市场上的主流品种, 在医药市场中占据主导地位[10]。考虑到麦冬与山麦冬在商品名与形态特征上的相近性, 临床与市场流通中存在潜在误标与混用风险[11], 本研究将山麦冬 *Liriope spicata* (Thunb.) Lour. 纳入对照开展并进行比较分析。

甾体皂苷是麦冬的代表性成分之一, 其典型的皂苷元主要是鲁斯可皂苷元 (ruscogenin) 和薯蓣皂苷元 (diosgenin)[9]。甾体皂苷元由于水溶性差, 在生物体内利用度低, 因此常常与糖结合形成皂苷而存在, 例如麦冬皂苷 D 为鲁斯可皂苷元的糖苷[12,13]。麦冬皂苷的生物合成涉及多酶协同作用, 其中氧化鲨烯环化酶 (OSC) 负责基本三萜骨架形成, 细胞色素 P450 (CYP450) 介导羟基化修饰, 随后由尿苷二磷酸依赖型糖基转移酶 (UDP-dependent glycosyltransferase, UGT) 组装糖链并塑造终产物皂苷分子的最终结构特征[14,15]。在植物次生代谢途径中, UGTs 作为关键修饰酶, 识别特定底物, 将核苷酸活化糖分子中的糖基单元转移至皂苷元等受体分子的羟基位点, 决定最终产物的结构多样性和化学稳定性, 影响其药理活性, 其 C 端高度保守的植物次生代谢产物糖基转移酶保守基序 (plant secondary product glycosyltransferase, PSPG), 能够与 UDP-糖供体特异性结合, 确保糖基化反应的精确进行[15-18]。在多种模式与药用植物中, *UGT* 基因家族已完成全基因组水平的鉴定与功能挖掘, 并被证实参与萜/皂苷底物的糖链起始、延伸与末端修饰; 有研究表明 UGT 表达水平与皂苷含量呈正相关, 提示其在药材品质形成中的关键作用[19-21]。例如, 喀西茄 (*Solanum aculeatissimum* Jacq.) 中的 SaGT4A 酶是甾体皂苷合成途径中首个经过功能验证的可介导薯蓣皂苷元、纽替皂苷元以及替告皂苷元等多种甾体皂苷元进行糖基化反应的糖基化酶, 并对部分甾体类生物碱的糖基化也具有催化作用[21]。在人参 (*Panax ginseng* C. A. Mey.) 基因组中共鉴定出约 225 个 *UGT* 基因[22], 其中部分成员已被证实直接参与人参皂苷的糖基化修饰。例如, PgUGT74AE2 和 PgUGT94Q2 分别催化原人参二醇型 (protopanaxadiol, PPD)/化合物 K (compound K, CK) 衍生物在 C-3 位的糖链逐步延长[23]。在甘草 (*Glycyrrhiza uralensis* Fisch.) 中共注释到 90 条 *UGT* 基因, 并进一步筛选出 12 条可能参与甘草甜素生物合成的候选 *UGT* 基因, 其中 UGT73P12 被证实催化甘草酸生物合成末端的第二次葡萄糖醛酸化[24]。

目前, 麦冬皂苷糖基化研究仍鲜有报道, 尤其是 *UGT* 基因家族的系统鉴定与分类, 组织表达谱及与麦

冬皂苷生物合成的相关性研究较为有限。本研究基于课题组完成的麦冬高通量转录组数据对麦冬 *UGT* 基因家族进行注释与鉴定, 结合生物信息学方法预测其系统发育关系、蛋白理化特性与保守结构域/基序特征; 并整合不同组织的表达谱数据, 筛选潜在参与皂苷生物合成的候选基因, 为阐明麦冬甾体皂苷糖基化的分子机制提供研究基础。

材料与方法

材料 本研究所用的麦冬材料采自麦冬主产区四川省绵阳市三台县花园镇, 经中国医学科学院药用植物研究所林余霖老师鉴定为天门冬科沿阶草属麦冬 (*O. japonicus*)。分别采集根、茎、叶组织样本 (图 1), 采集后立即用无菌水冲洗干净, 经液氮快速冷冻 20 s 后, 迅速转移至−80 ℃超低温冰箱保存。所有样本最终采用干冰冷冻运输至北京百迈客生物科技有限公司 (Biomarker Technologies, Beijing, China) 进行转录组测序分析。



Figure 1 Morphology of *Ophiopogon japonicus* (L. f.) plants and different tissues. A, B: Whole-plant morphology of *O. japonicus*; C: Tuberous root; D: Stem; E: Leaf

转录组测序 本研究的真核 *de novo* 转录组 (无参转录组) 测序由北京百迈客生物科技有限公司完成。分别从麦冬根、茎、叶 3 个组织的 9 份样本 (每个组织 3 个生物学重复) 中提取总核糖核酸 (RNA), 并进行严格质量控制; 质检合格的 RNA 用于构建互补脱氧核糖核酸 (complementary DNA, cDNA) 文库。文库经质检合格后, 采用 Illumina NovaSeq X Plus 测序平台 (Illumina Inc., San Diego, CA, USA) 进行 PE150 (paired-end 150 bp) 双端测序。

转录组数据质控、组装和注释分析 将麦冬原始转录组测序数据经 FastQC 软件进行质量评估, 采用 fastp、soap 软件去除接头、低质量的测序读长 (reads) 以及过滤核糖体数据, 使 Q30 (Phred quality score Q30) 大于 85%。随后采用 Trinity (v2.14.0) 进行 *de novo* 组装, 使用参数为--min_contig_length 200, --group_pairs_distance 500, 选取最长转录本后, 使用 MMseqs2 (v13.45111) 软件去除冗余序列, 并利用 TGICL (v2.1) 软件进行聚类分析, 选取最优转录本。最后, 使用 TransDecoder (v5.0.0) 进行基因结构注释, 预测编码序列并获得候选编码区 (coding sequence, CDS)。并使用 BUSCO (Benchmarking Universal Single-Copy Orthologs) 软件对麦冬的转录组注释数据进行质量评估, 数据库版本为 embryophyta_odb10。

麦冬 *OjUGT* 基因的筛选与鉴定 从 UDP-糖基转移酶超家族数据库 (http://www.p450.kvl.dk/.shtml) 中下载 122 条拟南芥 (*Arabidopsis thaliana*) UDP-糖基转移酶蛋白序列作为参考序列。利用 InterProScan (v5.52-86.0) 软件对麦冬转录组数据进行结构域预测, 所用数据库包括保守结构域数据库 (conserved domain database, CDD) 和蛋白家族数据库 (protein families database, Pfam)。根据 CDD 编号 cd03784 和 Pfam 编号 PF00201 进行合并筛选, 获得麦冬候选基因序列, 用于后续多物种基因比较分析。随后, 采用 Transdecoder 软件中的 *get_FL_accs.pl* 脚本提取包含起始密码子和终止密码子的完整编码序列, 并通过人工校验去除长度小于 300 的氨基酸序列, 获得麦冬全长 *UGT* 候选 CDS 序列, 并将其依次命名为 *OjUGT1-34*, 采用 MAFFT (v7.525) 软件将其与拟南芥基因序列进行多序列比对, 参数设置为--auto--maxiterate 1000--localpair; 并使用 RAxML (v8.2.12) 基于最大似然 (maximum likelihood, ML) 构建系统发育树, 模型为 PROTCATWAG, 自举检验 (bootstrap) 重复 100 次; 系统发育树用 chiplot 在线平台 (https://www.chiplot.online/) 进行可视化。

麦冬 OjUGT 蛋白质理化性质及亚细胞定位预测 通过 ExPASy 在线工具 (https://web.expasy.org/protparam/) 对包含完整开放阅读框 (open reading frame, ORF; 含起始密码子与终止密码子) 的麦冬全长 UGT 蛋白 (OjUGT) 进行理化性质预测, 获得氨基酸长度、理论分子量 (molecular weight, MW)、理论等电点 (isoelectric point, pI)、不稳定指数 (instability index)、脂肪族指数 (aliphatic index) 及亲疏水性指数 (grand average of hydropathy, GRAVY) 等参数。蛋白质亚细胞定位采用 Plant-mPLoc 在线工具 (Plant-mPLoc, http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) 进行预测。

麦冬OjUGT蛋白保守基序及蛋白结构可视化分析 将麦冬全长UGT蛋白 (OjUGT) 序列上传至MEME在线平台 (Multiple EM for Motif Elicitation, MEME; https://meme-suite.org/meme/) 进行保守基序 (motif) 预测, 设定最大基序数为15, 显著性阈值为*E* value < 1×10^{-30}, 并以XML格式导出结果; 通过chiplot网站对基序分布与结构特征进行可视化展示。对植物次生代谢产物糖基转移酶保守基序PSPG motif进行多序列比对, 以展示其保守性特征。针对序列相似但PSPG motif存在差异的OjUGT3与OjUGT4, 进行蛋白质三维结构预测; 所得结构采用PyMOL软件对整体构象及PSPG motif区域进行三维可视化与结构比对分析。

***OjUGT*基因在麦冬不同组织部位的表达特异性分析** 通过Bowtie2 (v2.3.5.1) 软件将麦冬根、茎和叶不同组织部位的转录组reads数据比对到转录组组装得到的参考转录本序列, 并结合RSEM (RNA-Seq by expectation-maximization) 进行表达量水平的估计, 以FPKM (fragments per kilobase of transcript per million mapped reads) 值表示基因的表达丰度。将*OjUGT*基因的表达量及拷贝数信息导入chiplot在线平台进行可视化展示与比较分析。

麦冬*UGT*基因家族的系统发育与多物种比较分析 为进一步探究基因家族的进化关系, 本研究下载了山麦冬 (*Liriope spicata*) 的转录组数据以及玉米 (*Zea mays* L.)、大麦 (*Hordeum vulgare* L.)、水稻 (*Oryza sativa* L.)、葱 (*Allium fistulosum* L.)、石刁柏 (*Asparagus officinalis* L.) 和食用芦笋 (*Asparagus kiusianus* Makino) 6个物种的基因组序列, 其中山麦冬转录组数据来源于NCBI数据库 (National Center for Biotechnology Information, NCBI), SRA编号为SRR15803258, 组装注释方法同麦冬; 玉米、大麦、水稻的基因组注释数据来源于Ensembl Plants数据库, 葱的基因组注释数据来源于中国国家基因库数据库 (China National GeneBank Sequence Archive, CNSA), 石刁柏的基因组注释数据来源于NCBI, 食用芦笋的基因组注释数据来源于Plant GARDEN数据库。各物种蛋白编码基因集的注释完整性采用BUSCO进行完整性评估。随后使用InterProScan (v5.52-86.0) 软件对各物种蛋白序列进行保守结构域注释, 并据UDP-糖基转移酶特征结构域 (Pfam PF00201和CDD cd03784) 筛选UGT候选序列, 以拟南芥 (*Arabidopsis thaliana*) 已鉴定UGT蛋白序列作为参考, 进行多序列比对并构建系统发育树, 基于系统发育拓扑结构以及拟南芥UGT亚家族命名体系将所有基因进行亚家族分类与比较分析。

结果与分析

1 转录组测序、组装与注释分析

9个样本经去接头、去低质量序列后, 得到clean reads 57.86 Gb; 各样本Q30为95.76%～96.37%, GC含量47.92%～49.01%, 均满足Q30 ≥ 85%的质控阈值。*De novo*组装后获得transcripts 229 156条, transcript的N50为1 245; unigenes 100 000条, unigene的N50为1 245, 二者N50相同, 组装完整性较高; 使用TransDecoder预测编码区, 共得到候选CDS与对应蛋白序列各22 532条。BUSCO评估显示完整度C61.4% (单拷贝S59.7%、重复D1.7%)、分段F17.7%、缺失M20.9%, 可满足本研究基因家族鉴定与表达分析的基本需求, 但仍存在一定不完全性。

2 麦冬*UGT*基因家族鉴定与分类

基于目前测序的麦冬根、茎、叶转录组数据注释结果, 本研究首次对麦冬*UGT*基因家族进行系统鉴定。获得109条*OjUGT*候选CDS, 以及34个麦冬全长UGT候选CDS, 并命名为*OjUGT1*～*34*。基于与拟南芥UGT序列联合构建的系统发育树, 所有*UGT*基因被划分为15个系统发育类群 (A～OG组), 麦冬全长UGT候选CDS主要聚集在A、B、C、D、E、G、I、J、L、M和OG组。其中麦冬全长UGT候选CDS在D组 (UGT73亚家族)中的成员数量最多, 有10个; 其次是G组 (UGT85亚家族), 有5个; 在A组 (UGT79/UGT91亚家族)、E组 (UGT71/72/88亚家族)、OG组 (UGT80/81亚家族)、J组 (UGT87亚家族) 中分别存在3个成员; 在B、C中各存在2个成员; I、L、M组中各包含1个成员; 而在F、H、K和N组中没有分布 (图2)。

3 麦冬UGT蛋白质理化性质预测分析

对34个麦冬全长UGT候选CDS所编码蛋白质的基本理化性质进行了预测分析, 结果表明: 这些蛋白的氨基酸长度介于355～588 aa之间, 平均长度为470 aa; 理论分子量范围为39.36～65.01 kDa。等电点 (pI) 分布在4.76～7.26之间, 其中33个OjUGT蛋白的pI < 7, 说明该家族整体以酸性蛋白为主。疏水性分析结果显示, 大多数OjUGT蛋白的GRAVY为负值, 提示整体以亲水性蛋白为主; 少数成员 (如OjUGT9、OjUGT21等) GRAVY为正值, 提示相对更疏水 (表1)。亚细胞定位预测结果显示, 麦冬UGT蛋白主要分布于细胞膜和叶绿体中, 少数定位于细胞质或细胞核 (表1)。

4 麦冬UGT蛋白质保守基序分析

通过MEME在线工具对麦冬OjUGT蛋白序列进行保守基序分析, 共鉴定出15个保守基序 (图3A)。这些基序在长度和氨基酸序列上存在一定差异, 但整体分布具有一定的规律性。每个OjUGT蛋白质序列中

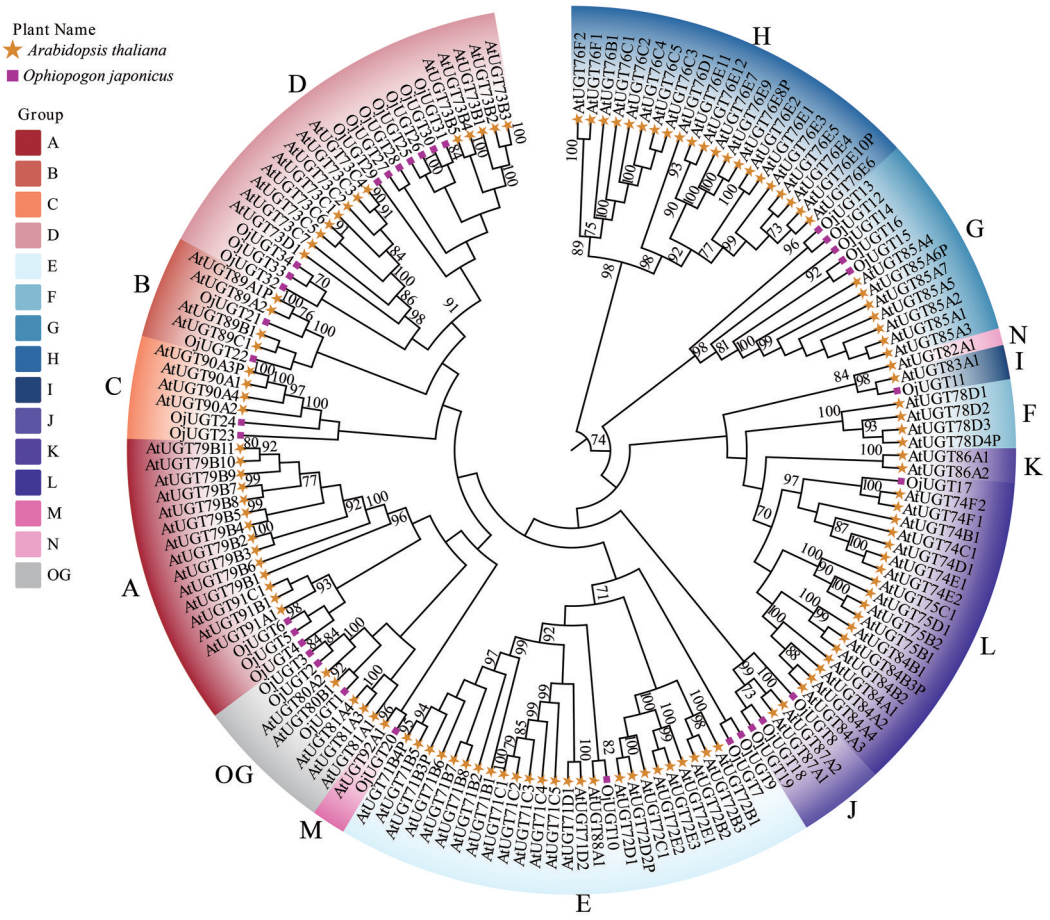


Figure 2 Phylogenetic relationships of the full-length UGT CDS in *O. japonicus*. UGT: UDP-dependent glycosyltransferase; CDS: Coding sequence

所含保守基序的数量在6～14个之间 (图3A)。其中, 除OjUGT1～3蛋白外, 其余31个OjUGT蛋白均包含基序1 (motif_1), 即典型的PSPG motif (图3A)。该基序由44个氨基酸残基组成, 在31条序列中表现出较高的保守性 (图3B)。进一步对OjUGT3与OjUGT4的蛋白三维结构进行比对 (图3C), 发现OjUGT4蛋白在PSPG motif区域呈现更完整的结构特征 (红色标注), 该结构通常作为糖基结合位点。

5 麦冬*UGT*基因在不同组织中的表达模式

为探究麦冬*UGT*基因家族在组织水平的表达特征, 利用麦冬块根、茎、叶的转录组数据, 对34个OjUGT的组织表达模式进行聚类分析, 结果显示, *OjUGTs*在不同组织中存在明显的差异表达模式 (图4), 有5个基因在3种组织中均表现为高丰度表达, 如*OjUGT3*、*OjUGT8*、*OjUGT10*、*OjUGT17*、*OjUGT34*; 另有10个成员表达量较低 (5 < FPKM < 10), 分别为*OjUGT6*、*OjUGT9*、*OjUGT12*、*OjUGT20*、*OjUGT21*、*OjUGT25*、*OjUGT26*、*OjUGT29*、*OjUGT31*、*OjUGT33*; 其中5个为极低表达 (FPKM < 5), 分别为*OjUGT2*、*OjUGT5*、*OjUGT7*、*OjUGT13*、*OjUGT27*。在组织特异性表达方面, *OjUGT32* (D组/UGT73亚家族) 在块根中特异高表达, 而*OjUGT34*在叶片中达到最高表达水平 (图4)。

6 多物种基因数量比较分析

为比较规模差异并推断进化变化, 本研究收集了山麦冬 (*L. spicata*)、玉米 (*Z. mays*)、大麦 (*H. vulgare*)、水稻 (*O. sativa*)、葱 (*A. fistulosum*)、石刁柏 (*Asparagus officinalis*) 和食用芦笋 (*A. kiusianus*) 等7个物种的转录组或基因组数据, 并以BUSCO评估组装完整性。各数据集的BUSCO完整度 (C) 在60%～93% (表2), 满足后续基因筛选与系统发育分析的质量要求。进一步统计分析显示, 不同物种的*UGT*基因家族成员数量和占比 [*UGT*基因数量/预测蛋白编码基因数 (%)] 存在差异: 天门冬科中, 山麦冬UGT候选序列数量最多 (180个, 占比0.28%), 麦冬最少 (109个, 占比0.18%), 而食用芦笋UGT序列数为123个, 占比0.46%; 禾本科物种中, 玉米UGT序列数量最多 (212个), 但占比仅0.15%, 大麦和水稻占比分别为0.53%和0.45% (表2)。

Table 1 Physicochemical properties and subcellular localization of UGT family proteins in *O. japonicus*. GRAVY: Grand average of hydropathy

Gene name	Amino acids length	Molecular mass/Da	pI	Instability index	Aliphatic index	GRAVY	Subcellular localization
OjUGT1	508	56 982.41	5.79	53.87	91.36	−0.091	Cytoplasm
OjUGT2	588	65 013.13	6.77	44.4	90.03	−0.142	Chloroplast
OjUGT3	468	51 820.34	5.81	42.93	88.74	−0.316	Chloroplast
OjUGT4	465	51 902.42	4.76	45.82	96.86	−0.037	Cell membrane, chloroplast
OjUGT5	476	53 202.34	5.8	48.72	88.13	−0.134	Cell membrane
OjUGT6	466	51 758.79	5.4	44.48	102.17	0.028	Cell membrane
OjUGT7	456	50 369.77	6.44	41.91	77.87	−0.296	Cell membrane, chloroplast
OjUGT8	479	51 625.78	5.98	50.44	85.49	−0.126	Cell membrane, chloroplast
OjUGT9	469	50 486.11	5.81	51.04	98.76	0.114	Cell membrane
OjUGT10	461	50 564.09	5.55	52.99	88.46	0.002	Cell membrane, chloroplast
OjUGT11	379	42 262.8	5.7	40.8	88.28	−0.065	Chloroplast, nucleus
OjUGT12	355	39 357.67	5.55	57.8	88.99	−0.208	Chloroplast
OjUGT13	460	51 823.44	6	49.82	85.98	−0.212	Cell membrane, chloroplast
OjUGT14	480	53 909.64	5.38	34.12	81.62	−0.223	Chloroplast
OjUGT15	482	53 605.59	4.97	40.72	88.01	−0.087	Chloroplast
OjUGT16	485	54 052.47	4.92	42.82	85.01	−0.218	Cell membrane, chloroplast
OjUGT17	460	51 280.9	5.01	42.82	84.39	−0.261	Chloroplast
OjUGT18	465	50 771.28	5.69	41.05	90.54	0.045	Cell membrane, chloroplast
OjUGT19	475	51 759.23	4.91	49	98.74	0.012	Cell membrane, chloroplast
OjUGT20	484	53 846.68	7.26	51.99	88.64	−0.169	Cell membrane, chloroplast
OjUGT21	455	48 947.36	5.68	41.39	97.78	0.231	Chloroplast
OjUGT22	491	53 563.47	6.1	48.69	95.68	0.059	Cell membrane, chloroplast
OjUGT23	481	52 491.16	5.8	43.5	90.02	−0.079	Cell membrane, chloroplast
OjUGT24	465	51 317.99	6.2	49.28	91.66	−0.015	Cell membrane, chloroplast
OjUGT25	484	52 907.85	6.25	37.00	83.24	−0.090	Chloroplast
OjUGT26	487	53 652.71	6	44.32	86.88	−0.074	Cell membrane, chloroplast
OjUGT27	368	41 087.11	6.02	46.75	83.99	−0.086	Chloroplast
OjUGT28	486	53 151.24	5.6	45.17	83.85	−0.177	Cell membrane
OjUGT29	466	51 383.75	5.8	46.03	88.71	−0.064	Cell membrane
OjUGT30	478	53 866.4	5.37	55.03	84.06	−0.217	Cell membrane
OjUGT31	488	53 278.94	5.7	42.23	88.34	−0.071	Cell membrane
OjUGT32	496	55 381.9	5.74	37.89	94.52	−0.058	Cell membrane, chloroplast
OjUGT33	495	55 050.53	5.5	36.37	92.77	−0.125	Cell membrane
OjUGT34	487	55 140.38	5.49	44.46	85.01	−0.168	Cell membrane, chloroplast

总体来看, 不同物种*UGT*基因在数量和占比上存在科内和科间差异, 提示*UGT*基因家族在不同物种中可能经历了不同程度的扩张或收缩。

7 多物种*UGT*基因家族进化分析

通过将麦冬、山麦冬及其他比较物种的UGT蛋白质与拟南芥已鉴定的UGT蛋白质进行多序列比对, 采用ML方法构建无根系统发育树 (图5)。基于进化树的拓扑结构, 并结合拟南芥亚家族的命名规则, 对所有的*UGT*基因进行亚家族划分 (图5)。分析结果表明, 所有物种的UGT共归入15个亚家族, 其中D组 (UGT73亚家族)、E组 (UGT71/72/88亚家族)、G组 (UGT85亚家族) 和L组 (UGT74/75/84亚家族) 在多数

Table 2 Statistical summary of *UGT* in different species. BUSCO: Benchmarking universal single-copy orthologs

Family	Species name	Data type	BUSCO completeness (C, %)	Protein-coding genes	Predicted number of *UGT*	*UGT* proportion /%
Asparagaceae	*O. japonicus*	Transcriptome	61.40%	59 855	109	0.18%
Asparagaceae	*L. spicata*	Transcriptome	77.40%	64 349	180	0.28%
Asparagaceae	*A. kiusianus*	Genome	85.60%	59 208	123	0.21%
Asparagaceae	*A. officinalis*	Genome	91.60%	26 459	90	0.34%
Asparagaceae	*A. fistulosum*	Genome	93.60%	62 242	144	0.23%
Poaceae	*H. vulgare*	Genome	93.90%	37 961	201	0.53%
Poaceae	*O. sativa*	Genome	81.30%	42 582	192	0.45%
Poaceae	*Z. mays*	Genome	94.60%	143 679	212	0.15%

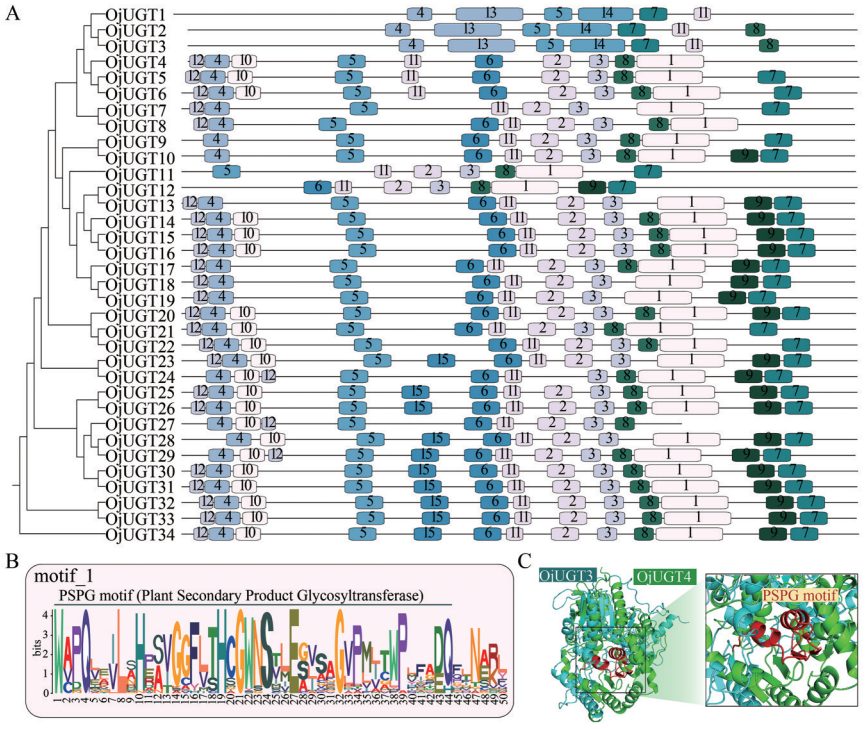


Figure 3 Conserved motifs, PSPG motif, and structural features of *O. japonicus* UDP-glycosyltransferases (OjUGTs). A: Conserved motif distribution of OjUGTs; B: Consensus sequence defining the plant secondary product glycosyltransferase (PSPG) motif; C: Structural comparison of OjUGT3 and OjUGT4. The red-marked regions indicate the PSPG motif. PSPG: Plant secondary product glycosyltransferase

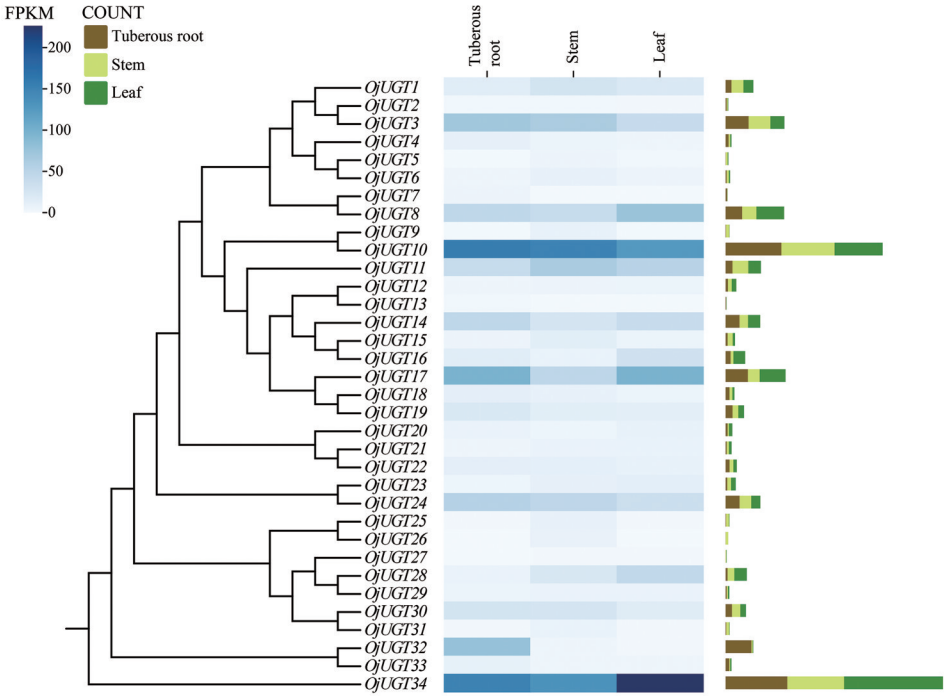


Figure 4 Comparison of gene expression levels and counts of *OjUGTs* in tuberous root, stems, and leaves of *O. japonicus*. FPKM: Fragments per kilobase of transcript per million mapped reads

物种中成员数较多, 提示相对保守的分布特征。麦冬的*UGT*基因分布于13个亚家族, 并在D组 (UGT73亚家族) 与G组 (UGT85亚家族) 显著富集; 相对地, F组 (UGT78亚家族) 和H组 (UGT76亚家族) 未检测到麦冬*UGT*基因 (表3), 各物种在不同系统发育类群中的*UGT*基因数目统计见表3。

Table 3 Distribution of the *UGT* in the different phylogenetic groups

Species name	A	B	C	D	E	F	G	H	I	J	K	L	M	N	OG	Isolated cluster
	79/91	89	90	73	71/72/88	78	85	76	83	87	86	74/75/84	92	82	80/81	
O. japonicus	7	3	2	30	12	0	24	0	2	2	1	15	2	1	3	5
L. spicata	13	7	2	41	18	3	38	0	2	2	4	25	7	1	13	4
A. kiusianus	10	1	5	31	13	3	18	0	1	3	1	18	6	1	6	6
A. officinalis	11	0	1	28	8	3	13	0	1	3	1	11	3	1	4	2
A. fistulosum	6	3	5	44	16	8	32	0	4	1	5	6	4	1	4	6
H. vulgare	15	3	6	26	37	3	36	9	12	6	4	29	4	2	3	6
O. sativa	19	4	8	25	42	2	29	7	8	3	1	23	9	2	4	6
Z. mays	9	3	10	16	39	2	21	16	11	6	2	33	3	4	32	5
A. thaliana	14	4	4	13	25	4	7	21	1	2	2	18	1	1	5	0

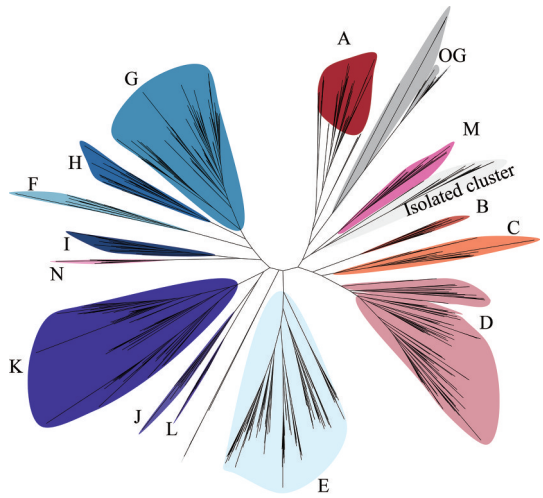


Figure 5 Phylogenetic relationships and subfamily distribution of UGT across multiple species

讨论

皂苷类化合物是麦冬的重要活性成分, 其生物学活性以及结构多样性通常与皂苷的糖基化修饰密切相关, 因此, 系统鉴定麦冬中的 *UGT* 基因家族并解析其进化特征以及表达模式, 对阐明麦冬皂苷类化合物的生物合成以及相关调控机制具有重要意义。本研究从麦冬转录组中鉴定出 34 个包含完整起始与终止密码子的 *OjUGT* 成员, 为后续功能验证提供了明确靶标。谱系划分显示, 麦冬全长成员分布于 11 个组, 主要富集于 D 组 (UGT73 亚家族) 和 G 组 (UGT85 亚家族), 而 F 组和 H 组却未检出, 提示可能存在结构偏倚或谱系特异性保留或缺失。已有研究表明, UGT73 亚家族在多种植物中参与三萜/甾体及黄酮底物的糖基化反应[24-26], 而 UGT85 亚家族则与细胞分裂素等信号分子的糖基修饰相关[27], 结合本研究中 UGT73 与 UGT85 的相对扩张特征, 可推测其扩增可能与麦冬甾体皂苷结构多样化的代谢需求相关, 但该推断仍需与代谢物谱及酶学证据相互印证。基序分析显示, 除 OjUGT1～3 蛋白外, 其他 31 个蛋白均含有典型的 PSPG 基序, 该基序与 UDP-糖供体识别密切相关, 是支持糖基转移活性的关键结构单元, 而 OjUGT1～3 蛋白缺乏完整 PSPG 基序, 并与 UGT80A2/UGT80B1 (OG 组/UGT80 亚家族) 聚类, 已有研究显示 *UGT80A2*/*UGT80B1* 基因与主要甾基糖苷 (SG) 和酰基甾基糖苷 (ASG) 化合物的积累相关[28-31]。需要注意的是, PSPG 的存在与否并非功能的唯一决定因素, “非典型”成员的底物谱与催化位点仍需通过体外酶学与体内代谢证据加以确认, 这一结果与其他药用植物中 *UGT* 基因家族研究一致[32,33]。在表达层面, *OjUGTs* 在根、茎、叶中呈现显著差异表达, 提示其在不同组织中承担不同的代谢功能。其中 *OjUGT32* 在块根中特异高表达, *OjUGT34* 在叶中特异高表达, *OjUGT32* 和 *OjUGT34* 基因均属于 UGT73 亚家族, 而 UGT73 在多种植物中被证实参与三萜/甾体及黄酮底物的糖基化, *OjUGT32* 和 *OjUGT34* 所编码的蛋白可能参与不同部位特有甾体皂苷的糖基化修饰, 推测为参与皂苷糖基化的候选基因; 而 *OjUGT3*、*OjUGT8*、*OjUGT10*、*OjUGT17* 和 *OjUGT34* 所编码的蛋白在多个组织中均有高表达, 提示其可能参与相关基础代谢或广谱次生代谢通路[34]。

多物种比较显示, 部分亚家族 (如 UGT79 亚家族、UGT85 亚家族) 在单/双子叶植物中呈现保守分布, 另一些亚家族 (如 UGT76 亚家族) 则在天门冬目类物种呈现明显的特异性丢失, 反映了天门冬目植物共有的进化模式, 提示这类基因在早期分化并保留, 而非近期的独立进化事件。基于基因数目与亚家族构成, 麦冬和山麦冬这对近缘物种在基因家族分布上表现出高度相似的特征, 但山麦冬的总基因数与个别支系更为丰富, 两者均表现为 UGT73 与 UGT85 等亚家族的显著扩张以及 UGT76 亚家族的完全丢失, 其中, UGT73 亚家族在十字花科物种中包含了最多的串联重复基因拷贝, 在多个串联簇中高度富集, 提示 UGT73 亚家族相比其他 UGT 亚家族更易通过串联复制扩张[31], 这类扩张可能与适应性进化及化学防御/信号调控需求相关。

山麦冬的基因总数及部分亚家族 (如UGT79、UGT89亚家族) 的基因数量明显多于麦冬, 提示山麦冬可能经历了更频繁的基因复制事件或具有更复杂的次生代谢调控网络。

麦冬甾体皂苷含量高、品质优异, 是解析皂苷生物合成途径的重要药用植物[35]。已有研究表明 *UGT* 基因家族在甾体皂苷的末端修饰中发挥关键作用, 尤其是UGT73亚家族在甾醇及黄酮中糖基化的功能已在模式植物中得到证实[36,37]。然而, 麦冬中催化皂苷糖基化的基因仍未明确。结合本研究的系统发育分析与表达模式, 将D组中的 *OjUGT32* 和 *OjUGT34* 基因视为与甾体皂苷糖基化潜在相关的候选UGT。本研究基于转录组开展家族鉴定, 主要用于筛选候选UGT及描述表达和进化的初步特征, 其推测结果仍需参考高质量基因组进一步验证, 后续将通过体外酶学进一步确认其功能。

作者贡献: 郑文芳负责课题方案设计、生物信息学分析以及初稿撰写; 张文颖负责课题方案设计以及初稿撰写; 秦众护负责论文审阅修改; 李铮负责论文审阅修改; 傅欣彤负责论文审阅修改; 闫海霞负责课题设计和论文指导; 罗红梅整体负责课题设计、论文指导及审阅修改。

利益冲突: 本文所有作者声明不存在利益冲突关系。

References

[1] Gu GG. Divine Farmer's Materia Medica (神农本草经) [M]. Vol 1. Beijing: Xueyuan Press, 2007: 44-46.

[2] Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China (中华人民共和国药典) [S]. Vol 1. 2025 Ed. Beijing: China Medical Science Press, 2025: 166-167.

[3] Ding L, Shangguan H, Wang X, et al. Extraction and purification of *Ophiopogon japonicus* polysaccharides, structural characterization, biological activity, safety evaluation, chemical modification and potential applications: A review [J]. Int J Biol Macromol, 2025, 300: 140282.

[4] Ge Y, Chen X, Gođevac D, et al. Metabolic profiling of saponin-rich *Ophiopogon japonicus* roots based on ^{1}H NMR and HPTLC platforms [J]. Planta Med, 2019, 85: 917-924.

[5] Liu Q, Lu JJ, Hong HJ, et al. *Ophiopogon japonicus* and its active compounds: A review of potential anticancer effects and underlying mechanisms [J]. Phytomedicine, 2023, 113: 154718.

[6] Chen MH, Chen XJ, Wang M, et al. *Ophiopogon japonicus* − a phytochemical, ethnomedicinal and pharmacological review [J]. J Ethnopharmacol, 2016, 181: 193-213.

[7] Pu HY, Cao Y, Jiang XL, et al. Steroidal saponins and homoisoflavonoids from the fibrous roots of *Ophiopogon japonicus* and their anti-pulmonary fibrosis activities [J]. Nat Prod Res, 2025, 39: 4055-4065.

[8] Lin Y, Yang CQ, Lian WY, et al. Ophiopogonin D interferes with ferroptosis to reduce the damage of cardiomyocytes induced by ophiopogonin D' [J]. Acta Pharm Sin (药学学报), 2021, 56: 2241-2247.

[9] Tan M, Chen J, Wang C, et al. Quality evaluation of *Ophiopogonis* Radix from two different producing areas [J]. Molecules, 2019, 24: 3220.

[10] Lyu Y, Du WF, Wu HS, et al. Research progress on difference of Zhe Maidong and Chuan Maidong [J]. China J Tradit Chin Med Pharm (中华中医药杂志), 2023, 38: 289-294.

[11] Lei F, Heinrich M, Reich E, et al. Quality variation of Maidong (*Ophiopogon japonicus* and *Liriope spicata*) − a HPTLC-based approach [J]. J Pharm Biomed Anal, 2024, 241: 115990.

[12] Ma Y, Zhao Y, Luo M, et al. Advancements and challenges in pharmacokinetic and pharmacodynamic research on the traditional Chinese medicine saponins: A comprehensive review [J]. Front Pharmacol, 2024, 15: 1393409.

[13] Sobolewska D, Galanty A, Grabowska K, et al. Saponins as cytotoxic agents: An update (2010 − 2018). Part I − steroidal saponins [J]. Phytochem Rev, 2020, 19: 139-189.

[14] Sawai S, Saito K. Triterpenoid biosynthesis and engineering in plants [J]. Front Plant Sci, 2011, 2: 25.

[15] Seki H, Tamura K, Muranaka T. P450s and UGTs: Key players in the structural diversity of triterpenoid saponins [J]. Plant Cell Physiol, 2015, 56: 1463-1471.

[16] Wang M, Ji Q, Lai B, et al. Structure-function and engineering of plant UDP-glycosyltransferase [J]. Comput Struct Biotechnol J, 2023, 21: 5358-5371.

[17] Masada S, Terasaka K, Mizukami H. A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (curcumin glucosyltransferase), a group D family 1 glucosyltransferase from *Catharanthus roseus* [J]. FEBS Lett, 2007, 581: 2605-2610.

[18] Hu S, Wang B, Pei L, et al. Advances and challenges in biomanufacturing of glycosylation of natural products [J]. Fermentation, 2024, 10: 349.

[19] Wang J, Li J, Li J, et al. Transcriptome profiling shows gene regulation patterns in ginsenoside pathway in response to methyl jasmonate in *Panax quinquefolium* adventitious root [J]. Sci Rep, 2016, 6: 37263.

[20] Kang KB, Jayakodi M, Lee YS, et al. Identification of candidate UDP-glycosyltransferases involved in protopanaxadiol-type ginsenoside biosynthesis in *Panax ginseng* [J]. Sci Rep, 2018, 8: 11744.

[21] Kohara A, Nakajima C, Hashimoto K, et al. A novel glucosyltransferase involved in steroid saponin biosynthesis in *Solanum aculeatissimum* [J]. Plant Mol Biol, 2005, 57: 225-239.

[22] Xu J, Chu Y, Liao B, et al. *Panax ginseng* genome examination for ginsenoside biosynthesis [J]. Gigascience, 2017, 6: gix093.

[23] Jung SC, Kim W, Park SC, et al. Two ginseng UDP-glycosyl-

transferases synthesize ginsenoside Rg3 and Rd [J]. Plant Cell Physiol, 2014, 55: 2177-2188.

[24] Nomura Y, Seki H, Suzuki T, et al. Functional specialization of UDP - glycosyltransferase 73P12 in licorice to produce a sweet triterpenoid saponin, glycyrrhizin [J]. Plant J, 2019, 99: 1127-1143.

[25] Zhou M, Fan J, Gao Y, et al. Identification and analysis of *UGT* genes associated with triterpenoid saponin in soapberry (*Sapindus mukorossi* Gaertn.) [J]. BMC Plant Biol, 2024, 24: 588.

[26] Augustin JM, Drok S, Shinoda T, et al. UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-*O*-glucosylation in saponin-mediated insect resistance [J]. Plant Physiol, 2012, 160: 1881-1895.

[27] Hou B, Lim EK, Higgins GS, et al. *N*-Glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana* [J]. J Biol Chem, 2004, 279: 47822-47832.

[28] Rojas RF, Rodriguez TO, Murai Y, et al. Linkage mapping, molecular cloning and functional analysis of soybean gene *Fg2* encoding flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase [J]. Plant Mol Biol, 2014, 84: 287-300.

[29] Stucky DF, Arpin JC, Schrick K. Functional diversification of two UGT80 enzymes required for steryl glucoside synthesis in *Arabidopsis* [J]. J Exp Bot, 2015, 66: 189-201.

[30] Yonekura - Sakakibara K, Hanada K. An evolutionary view of functional diversity in family 1 glycosyltransferases [J]. Plant J, 2011, 66: 182-193.

[31] Yu J, Hu F, Dossa K, et al. Genome-wide analysis of UDP-glycosyltransferase super family in *Brassica rapa* and *Brassica oleracea* reveals its evolutionary history and functional characterization [J]. BMC Genomics, 2017, 18: 474.

[32] Tian M, Yin YP, Wang SY, et al. Genome-wide characterization and prokaryotic expression of *UGT* gene family in *Dipsacus asper* Wall. ex Henry [J]. Mod Tradit Chin Med Mater Med-World Sci Technol (世界科学技术－中医药现代化), 2025, 27: 2035-2049.

[33] Tan ZW, Lu DD, Li L, et al. Identification and characterization of flavonoid 3-*O*-glycosyltransferase gene *CtUF3GT* from safflower (*Carthamus tinctorius* L.) [J]. Acta Pharm Sin (药学学报), 2022, 57: 2543-2551.

[34] Wilson AE, Tian L. Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism [J]. Plant J, 2019, 100: 1273-1288.

[35] Wu FM, Zhang SD, Zeng J, et al. Determination of the 4 representative components in Maidong from different producing areas by HPLC-ELSD method [J]. Chin J Pharm Anal (药物分析杂志), 2016, 36: 1370-1376.

[36] Erthmann PØ, Agerbirk N, Bak S. A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate sapogenins, forming a spectrum of mono-and bisdesmosidic saponins [J]. Plant Mol Biol, 2018, 97: 37-55.

[37] Poppenberger B, Fujioka S, Soeno K, et al. The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids [J]. Proc Natl Acad Sci U S A, 2005, 102: 15253-15258.