

利用转录组分析款冬萜类化合物生物合成关键酶基因及表达特征

贺润丽¹, 王晓英¹, 韩毅丽¹, 刘计权¹, 杜晨晖¹, 王莉花^{2*}

1. 山西中医药大学, 山西 太原 030619

2. 云南省农业科学院生物技术与种质资源研究所, 云南省农业生物技术重点实验室, 农业部西南作物基因资源与种质创制重点实验室, 云南 昆明 650223

摘要: 目的 挖掘与款冬萜类化合物生物合成途径相关的关键酶基因。方法 利用 Illumina HiSeq2500 测序平台对野生款冬花蕾和叶进行转录组测序, 使用 Trinity 软件 *de novo* 从头组装, 并通过基因功能注释、差异表达基因分析等生物信息学方法进行分析, 挖掘款冬萜类化合物生物合成途径相关的酶基因。结果 转录组测序获得 39 912 371 条高质量 reads (SRA 号: SRR9113366, SRR9113367), 组装获得 91 118 条 Unigene, 55 830 条 Unigene 在 NR、Swiss-Prot、GO、COG、KEGG 等数据库得到注释。鉴定出参与款冬萜类化合物生物合成相关酶基因的 129 个 Unigene, 包括 91 个参与三萜骨架生物合成酶基因, 32 个萜类合成酶基因, 6 个细胞色素 P450 基因, 其中差异表达基因 25 个, 涉及上游 MVA 途径的 4 个差异基因在花蕾中表达量高于叶, MEP 途径的 5 个差异基因在叶中表达高于花蕾, 另外还有 10 个基因在叶中高表达, 9 个基因在花蕾中高表达。根据差异基因 HMGR、TPS、AS、CYP450 基因在花蕾中高表达, 推测可能与花蕾中萜类物质含量高有关。结论 从款冬转录组数据库中初步获得了可能参与萜类化合物生物合成途径的候选关键酶基因, 为进一步阐明款冬萜类化合物生物合成的分子机制奠定基础。

关键词: 款冬; 转录组; 萜类化合物; 生物合成; 基因表达

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2020)20-5302-09

DOI: 10.7501/j.issn.0253-2670.2020.20.024

Analysis of key genes and their expression characteristics related to terpenoid biosynthesis in *Tussilago farfara* based on transcriptome sequencing

HE Run-li¹, WANG Xiao-ying¹, HAN Yi-li¹, LIU Ji-quan¹, DU Chen-hui¹, WANG Li-hua²

1. Shanxi University of Chinese Medicine, Taiyuan 030619, China

2. Key Lab of Southwestern Crop Gene Resources and Germplasm Innovation, Ministry of Agriculture, Yunnan Provincial Key Laboratory of Agriculture Biotechnology, Biotechnology and Germplasm Resources Institute, Yunnan Academy of Agriculture Sciences, Kunming 650223, China

Abstract: Objective To screen candidate genes involved in the terpenoid biosynthetic pathway of *Tussilago farfara*. **Methods** The transcriptome of buds and leaves of wild *T. farfara* were respectively sequenced using the Illumina HiSeq 2500 high-throughput sequencing platform. The clean reads were *de novo* assembled by Trinity software, and the assembled sequences then followed by a series of bioinformatics analysis such as gene function annotation and differential expression gene. According to sequence annotation and differentially expressed genes analysis, the key enzyme genes related to the terpenoid biosynthesis were identified. **Results** After high through-put sequencing, a total of 39 912 371 clean reads were obtained (SRA accession: SRR9113366, SRR9113367). The clean reads were then assembled into 91 118 unigenes. A total of 55 830 unigenes were annotated by a similarity search against NR, Swiss-Port, GO, COG, KEGG five public databases. Base on KEGG annotation and differentially expressed genes, totally 129 catalytic enzyme genes referring to the terpenoid biosynthesis were identified, including 91 terpenoid backbone biosynthesis genes, 32 terpene synthases, and 6 cytochrome P450 (CYP450) genes. Among them, 25 genes were differentially expressed. The expression

收稿日期: 2020-02-03

基金项目: 山西省国际科技合作项目 (批准号: 2012081045-1); 山西省现代农业产业技术体系建设专项; 山西省中药现代化关键技术研究振东专项 (2016ZD0108); 山西省重点研发计划 (201803D221012-6); 山西中医药大学科技创新团队项目 (2018-TD-009); 云南省人才培养项目 (2014HB062)

作者简介: 贺润丽, 博士, 副教授, 从事药用植物资源开发利用研究。Tel: (0351) 3179903 E-mail: herunli666@163.com

***通信作者** 王莉花, 硕士, 研究员, 从事植物遗传育种研究。Tel: (0871) 65894713 E-mail: wanglihua70@hotmail.com

of four enzyme genes in MVA pathway in leaves were higher than that in buds, while the five enzyme genes in MEP pathway were lower in leaves than that in buds. In addition, 10 genes were highly expressed in leaves, and nine genes were highly expressed in buds. According to the high expression of differentially expressed HMGR, TPS, AS, CYP450 genes in buds, it was speculated that these genes may be related to the high content of terpenoids in flower buds. **Conclusion** This work obtained candidate key enzyme genes that may be involved in the biosynthesis of terpenoid by transcriptome sequencing. The results laid a foundation for further elucidating the molecular mechanism of terpenoid biosynthetic pathway in *T. farfara*.

Key words: *Tussilago farfara* L.; transcriptome; terpenoid; biosynthesis; gene expression

款冬 *Tussilago farfara* L. 为菊科款冬属植物, 其干燥花蕾是我国传统中药材款冬花, 具有润肺下气、止咳化痰的功效, 临床用于治疗新旧咳嗽、喘咳痰多、劳嗽咳血等症^[1]。在欧洲, 款冬叶也是一种具有悠久历史的治疗咳嗽的草药^[2]。研究发现款冬花含有倍半萜、三萜皂苷、黄酮、苯丙素、生物碱、挥发油、色原酮等多种化合物^[3], 其中萜类化合物是一类主要活性物质, 现已分离出款冬酮 (tussilagone)、款冬花酮 (tussilagonone)、款冬花素内酯 (tussilagolactone) 等 42 个倍半萜成分, 具有升高血压、呼吸兴奋、抗炎、抗过敏、抗肿瘤等作用^[4-6]; 款冬二醇 (faradiol)、山金车二醇 (arnidiol)、巴耳新二醇 (bauer-7-ene-3 β , 16 α -diol)、巴耳三萜醇 (bauenenol) 和异巴耳三萜醇 (isobauerenol) 等三萜类成分^[3-4], 具镇痛、抗炎、抗肿瘤等作用^[3,7-10]。尽管萜类化合物具有重要的药理作用, 但这些化合物大多在款冬中含量非常低, 提高其含量不仅能提高款冬花药材的品质, 也能满足不断增长的市场需求。

随着高通量测序技术的发展, 可以对基因组信息匮乏物种的基因组和转录组功能基因进行全面细致的研究分析, 高通量转录组测序技术已广泛应用于人参、西洋参、凤丹等药用植物次生代谢生物合成途径功能基因的研究^[11-13]。应用转录组测序技术, 聂佳慧等^[14]对不同生长时期款冬叶中苯丙素类成分生物合成相关基因的表达进行了比较分析, 贾岩等^[15]对款冬花不同发育阶段的苯丙素类成分生物合成相关酶基因进行了比较研究, 但作为一种非模式植物, 款冬基因组和功能基因的研究还是比较缺乏, 生物合成途径信息有限。李静等^[16]研究显示款冬花和叶含有相似的化学成分, 款冬酮、款冬花素酯等倍半萜类物质在款冬花蕾中的含量远远高于叶, 所以为了获得款冬萜类物质生物合成途径的相关酶基因, 本研究利用 Illumina HiSeq 2500 平台对款冬花蕾、叶进行比较转录组测序分析, 挖掘款冬中参与萜类化

合物生物合成途径的关键酶基因, 为款冬品质改良与功能基因研究奠定基础。

1 材料

2017 年 3 月 7 日于山西省榆次市榆社县采集未出土野生款冬的带花蕾根茎 (海拔 1 063 m, 37°15' N, 112°59' E), 移栽于花盆中, 室内自然条件下生长, 室内温度 16~20 °C, 3 月 10 日采集新鲜花蕾, 花蕾呈紫红色, 4 月 15 日采集新鲜叶片, 每样重复 3 次, 纯净水冲洗干净, 以铝箔纸包裹液氮速冻后, 于 -80 °C 保存, 用于总 RNA 的提取。样品经山西中医药大学刘计权副教授鉴定为款冬 *Tussilago farfara* L. 的花蕾和叶。

2 方法

2.1 总 RNA 提取

使用 Trizol 试剂盒 (Total RNA Extractor SK1312, 上海生工生物工程股份有限公司) 分别提取款冬花蕾和叶的总 RNA, 提取流程严格参照说明书进行操作。采用 Nanodrop、Qubit 2.0、Agilent 2100 方法检测 RNA 样品的纯度、浓度、完整性等, 检测合格后, 将 3 份 RNA 样品分别等量混合用于转录组测序。

2.2 文库构建、转录组测序和数据组装

用带有 Oligo (dT) 的磁珠从总 RNA 中分离 mRNA, 加入 fragmentation buffer 将 mRNA 随机打断为短片段; 以短片段 mRNA 为模板, 使用六碱基随机引物和逆转录酶合成双链 cDNA; 纯化双链 cDNA 并进行末端修复、加 poly (A) 尾、连接测序接头; 用 AMPure XP beads 进行片段大小选择, 最后通过 PCR 扩增得到 cDNA 文库。分别使用 Qubit2.0、Agilent 2100 对文库的浓度和插入片段大小进行检测, 库检合格后, 用 Illumina HiSeq2500 2×125 bp (Illumina 公司, 美国) 平台进行双末端测序。将测序所得的原始序列进行过滤, 去除接头序列及低质量序列等获得高质量序列 (clean reads), 利用 Trinity 软件通过序列之间的重叠信息进行从头组装获得重叠群 (contigs), 依据序列两端

信息及重叠群间的相似性进行聚类整合和延长,得到转录本(transcripts),从中选取最长的转录本作为非重复序列基因(Unigenes)^[13]。

2.3 功能注释

使用 BLASTx 软件将 Unigene 序列分别与 NR (ftp://ftp.ncbi.nih.gov/blast/db/), Swiss-Prot (http://www.uniprot.org/), GO (http://www.geneontology.org/), COG (http://www.ncbi.nlm.nih.gov/COG/), KEGG (http://www.genome.jp/kegg/) 等数据库比对,进行序列相似性分析与功能注释预测 ($E < 1 \times 10^{-5}$),然后再利用 HMMER 软件与 Pfam 数据库比对 ($E < 1 \times 10^{-10}$),获得 Unigene 的功能注释信息^[13]。

2.4 基因差异表达分析

Unigene 表达量采取 FPKM (fragments per kb per million reads) 法进行计算,对基因长度与测序深度进行归一化处理。根据基因表达量的 FPKM 值计算出该基因在不同样本间的差异表达倍数,采用 DESeq 软件进行样品组间的差异表达分析,在筛选过程中,将错误发现率 $FDR < 0.01$ 、差异倍数 FC (Fold Change) ≥ 2 作为筛选标准。

3 结果与分析

3.1 转录组测序与数据组装

去除测序原始数据中总体质量偏低和含接头序列的读序后,款冬花蕾和叶分别获得 20 135 732、19 776 639 个高质量序列(clean reads),分别包含了 5 069 887 508、4 977 868 324 个碱基,共 10.05 Gb 数据,GC 含量分别为 42.31%、42.52%,碱基错误率(Q30)分别为 91.12%、91.02%,说明样品测序质量较好,可进行后续数据的组装和分析。转录组测序原始数据已经提交 NCBI 的

SRA 序列数据库,登陆号为 SRR9113366, SRR9113367。

利用 Trinity 软件进行从头组装,花蕾和叶分别获得 72 844、50 398 个 Unigene,花蕾的总长度、N50 长度、平均长度分别为 44 509 629、1 081、611 nt;叶的序列总长度、N50 长度、平均长度分别为 36 816 804、1 324、731 nt (表 1)。根据序列之间的重叠部分,对款冬花蕾和叶的序列进行混合组装,共得到 91 118 个 Unigene,平均长度 696 nt, N50 为 1 327 nt (表 1),其中长度在 200~300 nt 的 Unigene 最多,达 39 353 (43.19%) 个,其次是 300~500 nt 的 Unigene 19 396 个 (21.29%), 500~1 000 nt 的 Unigene 有 13 431 个 (14.74%), 1 000~2 000 nt 的 Unigene 有 12 199 个 (13.39%), 长度大于 2 000 nt 的 Unigene 为 6 739 (7.40 %) 条,款冬 Unigene 序列长度分布信息详见图 1。

表 1 款冬转录组序列组装结果统计

Table 1 Statistic of *T. farfar* transcriptome sequencing data

项目	花蕾	叶	所有 Unigene
碱基数	5 069 887 508	4 977 868 324	—
Clean reads	20 135 732	19 776 639	—
GC 含量/%	42.31	42.52	—
Q30/%	91.12	91.02	—
Contig	2 864 496	2 434 576	—
Transcripts	150 076	118 469	—
Unigene 数	72 844	50 398	91 118
总长度/nt	44 509 629	36 816 804	63 459 066
N50 长度/nt	1 081	1 324	1 327
平均长度/nt	611	731	696



图 1 Unigene 长度分布

Fig. 1 Length distribution of Unigenes in *T. farfar*

3.2 Unigene 功能注释

3.2.1 序列功能注释 将拼接所得的 Unigene 通过 BLASTx 分别比对到 NR、Swiss-Prot、GO、COG、KEGG 等数据库, 共有 55 830 (61.27%) 个 Unigene 得到注释, 其中, 在 NR 数据库获得注释的 Unigene 最多 54 365 个 (59.67%), 在 Swiss-Prot、Go、COG、KEGG 等数据库获得注释的 Unigene 数依次为 36 080 (39.60%)、29 950 (32.87%)、17 604 (19.32%)、13 143 个 (14.42%) (表 2), 由于缺少款冬基因组信息, 超过 1/3 的 Unigene 无注释信息。

款冬功能基因序列与目前研究较深入的 9 种植物有较高的同源性, 依次是葡萄 *Vitis vinifera* L. (5 980, 11.00%)、芝麻 *Sesamum indicum* L. (3 402, 6.26%)、咖啡 *Coffea canephora* Pierre ex Froehn. (3 072, 5.65%)、林烟草 *Nicotiana sylvestris* Speng. et Comes (2 423, 4.46%)、绒毛状烟草 *Nicotiana tomentosiformis* Goodsp. (2 061, 3.79%)、可可 *Theobroma cacao* L. (1 675, 3.08%)、马铃薯 *Solanum tuberosum* L. (1 270, 2.34%)、甜橙 *Citrus sinensis* (L.) Osbeck (1 247, 2.29%)、麻风树 *Jatropha curcas* L. (1 181, 2.17%)。

表 2 Unigene 的功能注释

数据库	注释 Unigene 数	百分比/%
NR	54 365	59.67
Swiss-Prot	36 080	39.60
GO	29 950	32.87
COG	17 604	19.32
KEGG	13 143	14.42
总计	55 830	61.27

3.2.2 GO 分类 GO 是国际化的基因功能分类数据库, 用于全面描述生物体基因及其产物的生物学特性。经过比对 GO 数据库, 将款冬 29 950 条 Unigene 归入细胞组分 (56 808)、分子功能 (36 137)、生物学过程 (79 014) 3 大类别, 进一步被细分为 52 个功能类别, 这一分类结果显示了款冬基因表达谱的总体情况。其中, 代谢进程 (20 397)、细胞进程 (17 492)、细胞器 (8 474)、催化活性 (15 763)、结合活性 (14 655)、细胞部分 (13 232) 和细胞 (13 145) 功能组中涉及的 Unigene 较多, 而翻译调节活性 (3)、胞外基质部分 (4) 和金属伴侣蛋白活性 (5) 功能组中涉及的 Unigene 较少 (图 2)。

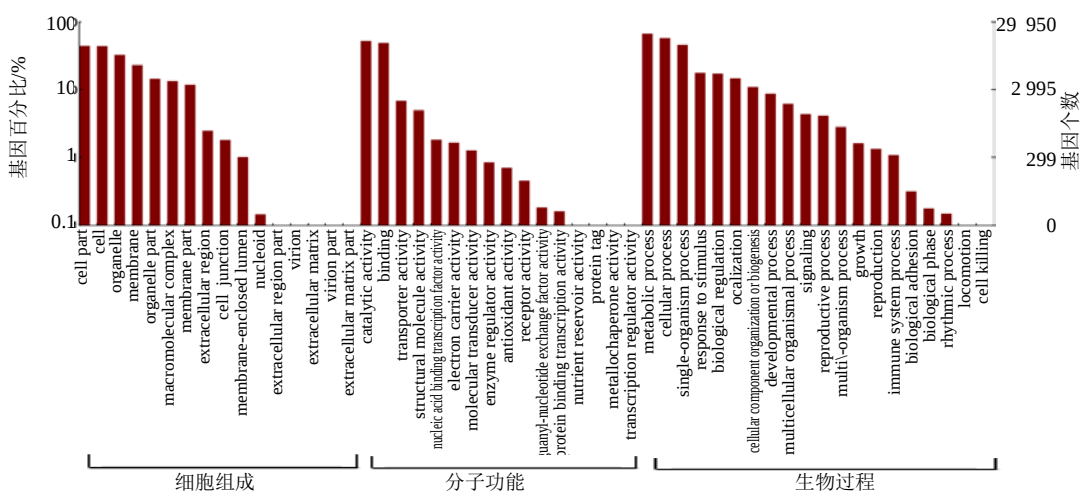


图 2 Unigene 的 GO 注释

Fig. 2 GO annotation distribution of Unigenes

3.2.3 COG 注释 COG 数据库是可以对基因产物进行同源分类的数据库, 经过与 COG 数据库比对, 17 604 个 Unigene 归属于 24 个类别 (图 3), 各个类别的基因表达丰度也不尽相同, 其中, 关于一般功能预测的基因最多 (4 513, 25.64%), 其次是翻译、核糖体结构和生物合成 (2 325, 13.21%)、DNA 复制、重组与损伤修复 (2 205, 12.53%)、转录 (2 152, 12.22%), 信号转导机制

(1 922, 10.92%), 而核结构类基因 (6, 0.03%) 和细胞运动类基因 (24, 0.14%) 较少, 胞外结构没有 Unigene 归入, 还有 579 (3.29%) 个 Unigene 功能未知 (图 3), 这些 Unigene 很可能是款冬的特异基因。此外, 有 655 (3.72%) 个 Unigene 参与了次生代谢物质的生物合成、转运与降解, 对这些基因的功能进行深入研究有助于明确款冬药效成分生物合成的分子机制。

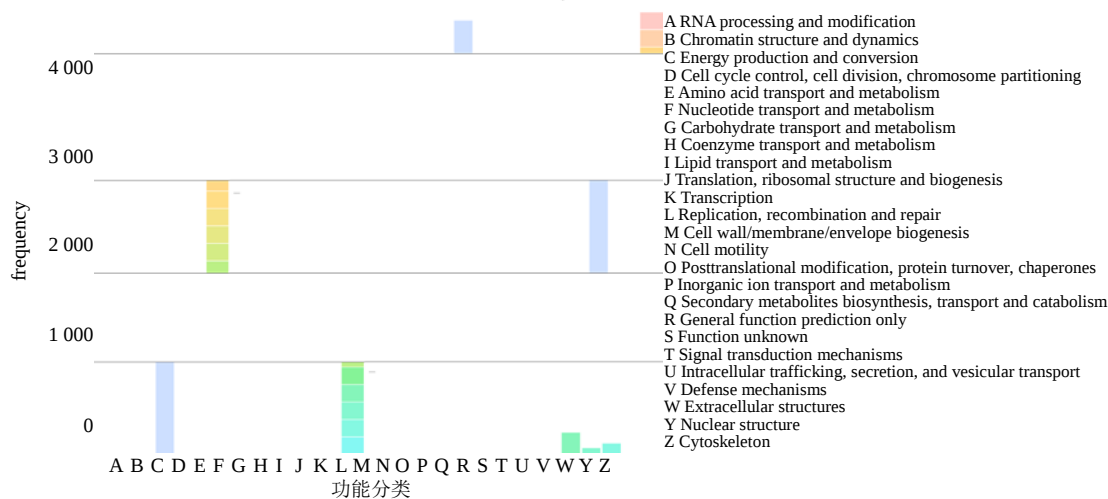


图 3 Unigene 的 COG 分类

Fig. 3 COG classification of Unigenes

3.2.4 KEGG 代谢途径注释 KEGG 是系统分析基因产物的代谢途径及功能的数据库。对款冬测序序列进行 KEGG 代谢途径分析, 共有 13 143 个 Unigene 归类于 118 条代谢途径。其中, 核糖体途径包含 Unigene 最多 1 189 个 (9.05%), 其次是内质网蛋白质加工 548 个 (4.17%), 氧化磷酸化 468 个 (3.56%), 剪接体 446 个 (3.39%), RNA 转运 439 个 (3.34%), 而亚油酸代谢、咖啡因代谢、油菜素内酯的生物合成途径相关基因较少, 分别仅有 1、6、7 条。映射到次生代谢途径的基因有: 萜类骨架合成途径 (ko00900) 的 Unigene 有 91 个, 苯丙素类生物合成途径 (ko00940) 143 个, 泛醌和萜醌类物质生物合成 (ko00130) 57 个, 二萜生物合成 (ko00904) 25 个, 类黄酮生物合成 (ko00941) 39 个, 黄酮和黄酮醇生物合成 (ko00944) 8 个, 类胡萝卜素生物合成 (ko00906) 58 个, 表明款冬次生代谢生物合成途径的多样性及化学成分的复杂性, 为后续阐明其功效物质的生物合成提供了理论依据。

3.3 款冬花蕾和叶差异表达基因分析

以基因表达水平相对值 FPKM 值大于等于 2, 且 $FDR < 0.01$ 作为标准筛选差异基因, 结果在款冬花蕾和叶之间获得差异基因 6 909 个, 占总数的 7.58%。相对于花蕾, 叶中有 4 266 个基因表达上调, 2 643 个基因表达下调, 其中 1 017 个基因归属于 106 条 KEGG 代谢途径, 差异基因数目最多的途径为核糖体途径 (ko03010), 有 147 个基因; 其次是光合作用途径 (ko00195) 差异基因 78 个; 萜骨架

生物合成途径 (ko00900) 差异基因 15 个, 苯丙烷类生物合成途径 (ko00940) 差异基因 41 个。

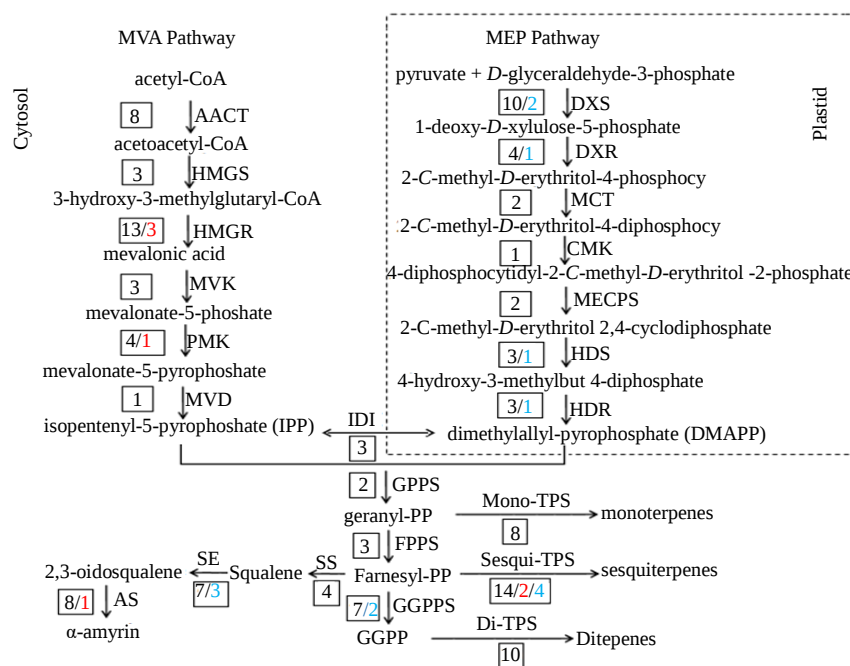
3.4 参与萜类化合物生物合成相关酶基因的筛选

由 KEGG 注释结果可知, 款冬萜类物质合成途径大致可分为萜类前体物质的合成、三萜骨架的合成和后修饰 3 个阶段。上游阶段前体物质生物合成相关的酶包括甲羟戊酸 (MVA) 途径的乙酰辅酶 A 乙酰转移酶 (acetyl-CoA C-acetyltransferase, AACT)、羟甲基戊二酰辅酶 A 合成酶 (hydroxymethylglutaryl-CoA synthase, HMGS)、羟甲基戊二酰辅酶 A 还原酶 (hydroxymethylglutaryl-CoA reductase, HMGR)、甲羟戊酸激酶 (mevalonate kinase, MVK)、磷酸甲羟戊酸激酶 (phosphomevalonate kinase, PMK)、甲羟戊酸焦磷酸脱羧酶 (phosphomevalonate decarboxylase, MVD); 2-甲基-D-赤藻糖醇-4-磷酸 (MEP) 途径的 1-脱氧-D-木酮糖-5-磷酸合酶 (1-deoxy-D-xylulose-5-phosphate synthase, DXS)、1-脱氧-D-木酮糖-5-磷酸还原酶 (1-deoxy-D-xylulose-5-phosphate reductoisomerase, DXR)、2-甲基-D-赤藓醇-4-磷酸胞苷酰转移酶 (2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, MCT)、4-焦磷酸胞苷-2-甲基-D-赤藓醇激酶 (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, CMK)、2-甲基-D-赤藓醇-2,4-环焦磷酸合酶 (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, MECPS)、羟甲基丁烯基-4-焦磷酸合酶 (4-hydroxy-3-methylbut-2-enyl-diphosphate synthase, HDS)、羟甲基丁烯基-4-磷酸

还原酶 (4-hydroxy-3-methylbut-2-enyl- diphosphate reductase, HDR), 此外还有异戊烯基焦磷酸异构酶 (isopentenyl-diphosphate delta-isomerase, IPPI) [17]。根据萜类物质生物合成的分子途径, 在款冬转录组数据中鉴定出上游前体合成阶段 MVA 途径 6 个关键酶的 32 个 Unigene, MEP 途径 7 个关键酶的 25 个 Unigene 以及异戊烯基焦磷酸异构酶 (IPPI) 3 个 Unigene (图 4)。组织差异表达分析表明, MVA 途径的 3 个 HMGR 基因和 1 个 PMK 基因在花蕾中表达量高于叶; 而 MEP 途径的 2 个 DXS 基因, DXR、HDS、HDR 各 1 个基因在叶中表达量高于花蕾, MEP 途径基因在款冬叶中高表达可能与叶中的叶绿体丰富有关 (表 3)。

中游阶段通过香叶基二磷酸合成酶 (geranyl pyrophosphate synthase, GPPS)、法尼基二磷酸合

成酶 (farnesyl diphosphate synthase, FPPS)、香叶基香叶基焦磷酸合成酶 (geranylgeranyl pyrophosphate synthase, GGPPS)、鲨烯合成酶 (squalene synthase, SS)、鲨烯环氧酶 (squalene monooxygenase, SE)、amyrin 合成酶 (amyrin synthase, AS) 等关键酶催化生成 α -香树酯 (α -amyrin) [17]。在款冬的转录组数据中有 31 个 Unigene 注释为中游萜类骨架合成途径的 6 个关键酶 (图 4), 其中 2 个 GGPPS 基因和 3 个 SE 基因在叶中表达量较高, 1 个 AS 基因在花蕾中表达量较高。amyrin 合成酶 (AS) 为乌苏烷型三萜骨架生物合成的关键酶, 本研究鉴定出 8 个 AS 基因, 其中 1 个 AS 基因 T02Unigene8627 在花蕾中的表达量是叶的 14 倍 (图 4、表 3), 可能是花蕾萜类物质合成的关键酶。



方框内黑色数字表示注释的总基因数, 红色数字表示在花蕾中表达上调的基因数, 蓝色数字表示在叶中表达上调的基因数

Black figures in panes indicate the number of unigenes corresponding to the catalytic gene in the pathway, red and blue figures represent up-regulated unigenes in flower buds and leaves respectively

图 4 款冬萜类骨架生物合成途径

Fig. 4 Schematic for putative terpenoid skeleton biosynthetic pathways of *T. farfara*

款冬转录组数据库还鉴定出 32 个萜类合成酶 (terpene synthase, TPS), 其中 8 个 Unigene 被注释为 R-芳樟醇合酶、蒎烯合成酶、香叶醇合酶等单萜合成酶 (Mono-TPS), 14 个 Unigene 注释为大根香叶烯合酶、法尼烯合酶、没药烯合酶、石竹烯合酶等倍半萜合成酶 (Sesqui-TPS),

10 个 Unigene 注释为柯巴基焦磷酸合成酶、贝壳杉烯合酶等二萜合成酶 (Di-TPS)。6 个倍半萜合成酶在花蕾和叶中差异表达, 其中 4 个在叶中高表达, 2 个在花蕾中高表达, 倍半萜合成酶 CL6293Contig1、CL12267Contig1 的表达量远远高于其它倍半萜合成酶, 且在花蕾的表达量分别

表 3 参与萜类化合物生物合成的差异表达基因

Table 3 Differentially expressed Unigenes involved in terpenoid biosynthesis

基因	差异表达 Unigene	长度/nt	FPKM		log ₂ FC	上调/下调↑/↓
			花蕾	叶		
HMGR	CL2652Contig1	1 827	386.25	15.71	-4.12	↓
	T01Unigene11476	1 161	56.44	8.00	-2.31	↓
	T02Unigene19576	1 908	1 346.97	4.47	-7.73	↓
PMK	T01Unigene22580	843	19.92	2.68	-2.35	↓
DXS	CL5015Contig1	2 124	0.91	6.37	3.22	↑
	T01Unigene11442	2 124	0.85	11.28	4.14	↑
DXR	T02Unigene8512	636	22.28	80.64	2.34	↑
HDS	T02Unigene24503	225	7.78	34.25	2.54	↑
HDR	T01Unigene13401	1 362	24.19	263.27	3.93	↑
GGPPS	T01Unigene15875	936	4.83	22.43	2.67	↑
	CL10869Contig1	1 056	14.24	137.11	3.75	↑
SE	T01Unigene13419	1 656	2.21	216.12	7.07	↑
	T01Unigene3292	1 006	0.96	4.96	2.71	↑
	T01Unigene5873	606	0.67	5.04	3.07	↑
AS	T02Unigene8627	2 304	92.17	6.62	-3.30	↓
	T01Unigene5718	1 686	0.17	10.31	5.86	↑
	CL5671Contig1	1 719	0.61	6.44	3.73	↑
Ses-TPS	T01Unigene23838	366	0	3.40	3.67	↑
	T01Unigene11791	1 200	0	50.31	9.73	↑
	CL6293Contig1	405	102.37	1.44	-5.48	↓
CYP716A12	CL12267Contig1	1 686	254.37	0.75	-7.81	↓
	T02Unigene1908	1 483	51.27	3.78	-3.25	↓
	T02Unigene18759	1 149	39.04	2.46	-3.47	↓
CYP72A154	T02Unigene23667	1 602	119.80	7.50	-3.50	↓
	T01Unigene2519	1 575	0	48.09	9.50	↑

↑表示基因在叶中表达上调, ↓表示基因在花蕾中表达上调

↑ represent up-regulated unigene in leaves, ↓ represent down-regulated unigene in flower buds

是叶的 71、339 倍,在一定程度上表明倍半萜合成反应更集中在花蕾中;另外 2 个倍半萜合成酶 T01Unigene23838、T01Unigene11791 只在叶中特异表达(图 4 和表 3)。

乌苏烷型三萜母核 α -amyrin 形成后,细胞色素 P450 氧化酶(CYP450)催化 α -amyrin 上不同位置发生氧化反应,经过不同位置与不同数量的结构修饰,大大提高了三萜结构的多样性。CYP450 是一个庞大且复杂多样的基因家族,仅少数几个参与三萜皂苷的 CYP450 被鉴定出来,为了有效缩小候选基因的范围,更确切地鉴定可能参与三萜皂苷生物合成途径的候选基因,本研究利用已知的参与三萜皂苷修饰的 CYP450 氨基酸序列对款冬转录组数据做 BLASTp,共筛选出 6 个同源性大于 50%,基因长度大于 1 000 bp 的 Unigene^[18]。其中 4 个 Unigene 差异表达,T02Unigene1908 被注释为 β -香树脂醇-28-氧化酶(β -amyrin-28-oxidase),在

花蕾中高表达;T02Unigene18759、T02Unigene23667、T01Unigene 2519 均被注释为 11-羧基- β -香树脂醇-30-氧化酶(11-oxo- β -amyrin 30-oxidase),T02Unigene18759、T02Unigene23667 在花蕾中高表达,T01Unigene2519 在叶中特异表达(表 3)。

4 讨论

RNA-Seq 高通量测序技术已经成为非模式生物基因信息解读、新基因发掘与基因功能研究的一种快捷途径,药用植物连翘 *Forsythia suspense* (Thunb.) Vahl^[19]、人参 *Panax ginseng* C. A. Meyer^[20]、丹参 *Salvia miltiorrhiza* Bunge^[21]、西洋参 *Panax quinquefolius* L.^[12]、蒙古黄芪等^[22]都进行了转录组测序研究,获得了大量有价值的信息和功能基因^[18-22]。聂佳慧等^[14]对不同发育阶段的款冬叶进行了转录组测序分析,获得 46 793 条 Unigene;贾岩等^[15]对不同发育阶段的款冬花进行了转录组测序分析,获得 44 553 712 条 Unigene;而本研究采用

倍半萜物质差异大的花蕾和叶^[15]进行比较转录组分析,经组装得到 91 118 条 Unigene,这种基因信息量的差别可能与测序平台、取材部位、取材时间等不同因素有关。

通过与公共数据库比对,款冬转录组有 55 830 (61.27%) 条 Unigene 获得了功能注释,其中,大量与款冬次生代谢物生物合成相关酶基因被发现,涉及萜类骨架生物合成途径基因 91 个,苯丙素类生物合成途径基因 143 个,泛醌和萜醌类物质生物合成基因 57 个,二萜生物合成基因 25 个,类黄酮生物合成基因 39 个,黄酮和黄酮醇生物合成基因 8 个,类胡萝卜素生物合成基因 58 个等。鉴定出与萜类物质生物合成相关的基因包括 91 个萜类骨架生物合成途径的酶基因,32 个单萜、倍半萜、双萜合成酶基因以及 6 个参与萜类生物合成的 CYP450,这为款冬萜类合成途径关键酶基因的研究提供了丰富的序列信息。但这些基因的功能注释全部来自生物信息学分析,需要进一步进行基因功能的研究。

本研究 Unigene 表达水平用 FPKM 值代表,可定量比较同一代谢路径不同酶基因的表达水平,根据 FPKM 值在款冬花蕾和叶共筛选出 6 909 个差异基因。差异基因表达分析表明:在萜类化合物合成上游途径中,款冬的 MVA 途径的 3 个 HMGR 基因和 1 个 PMK 基因在花蕾中表达量高;MEP 途径的 2 个 DXS 基因,DXR、HDS、HDR 各 1 个基因在叶中表达量高;中游途径的 2 个 GGPPS 基因和 3 个 SE 基因在叶中表达量较高,说明这些基因的表达水平有明显的器官特异性,且款冬上游途径关键酶的表达模式与蒙古黄芪的表达模式相类似^[22]。根据李静的研究表明倍半萜类物质在款冬花蕾中的含量远远高于叶,推测 HMGR、PMK 可能与款冬萜类物质生物合成的有关。研究表明 MVA 途径的 HMGR 被认为是萜类物质生物合成的关键限速酶,也是重要调控点^[17],款冬中注释为 HMGR 基因的 Unigene 数最多,达 13 个,且有 3 个 HMGR 基因在花蕾中表达量高于叶,其中 T02Unigene19576、CL2652Contig1 在花蕾的表达量高达 1 346.97、386.25,分别是叶的 301、25 倍(表 3),推测这 2 个 HMGR 基因在花蕾中的高表达可能与花蕾中倍半萜类成分含量高有关,是款冬萜类物质合成的关键酶,值得进一步深入研究其功能。

在植物细胞中,单萜、二萜化合物以香叶基焦磷酸(GPP)为前体在质体中合成,而倍半萜和三

萜类化合物以法尼基焦磷酸(FPP)为前体在细胞质中合成。TPS 是合成萜类次生代谢物的关键酶,款冬大部分萜类合成酶表达水平较低,FPKM 值在 0.27~20.36,而 2 个倍半萜合成酶 CL6293Contig1、CL12267Contig1 在花蕾中表达量较高,分别为 254.37、102.37,是叶的 71、339 倍,这些基因的高表达可能与款冬花蕾中倍半萜成分含量较高有关^[16],值得进一步深入研究。萜类合成酶也是萜类代谢工程的靶位点,且成功的例子也很多,Besumbes 等将催化 GGPP 生成紫杉醇烯的紫杉醇烯合酶(TXS)在拟南芥中表达,当采用糖皮质激素诱导后,紫杉醇烯在转基因植物叶片的干重达 600 $\mu\text{g/g}$,比组成型表达的转基因植株紫杉醇烯的含量提高 30 倍^[23]。本研究鉴定出的一些萜类合成酶,有望通过基因工程手段提高款冬萜类物质的含量。

本研究还鉴定出可能参与款冬乌苏烷型三萜皂苷合成相关的 8 个 AS 基因、6 个 CYP450 基因。AS 是催化 2,3-环氧角鲨烯环化生成 α -amyrin 或 β -amyrin 的关键酶,T02Unigene8627 在花蕾的表达量是叶的 14 倍;3 个 CYP450 基因 T02Unigene1908、T02Unigene18759、T02Unigene23667 在花蕾中也高表达,这些候选基因可能参与了三萜皂苷的生物合成,还需使用更可信的方法进行筛选和验证。中药资源转录组研究逐渐成为一个新的研究热点,通过转录组数据的分析,可推测有效成分生物合成途径,挖掘有效成分代谢途径的关键酶基因,为中药资源活性成分的生物合成与调控提供新的途径,从而为中药的质量控制、品种选育提供技术支撑。

参考文献

- [1] 中国药典 [S]. 一部. 2015.
- [2] Li Z Y, Zhi H J, Zhang F S, et al. Metabolomic profiling of the antitussive and expectorant plant *Tussilago farfara* L. by nuclear magnetic resonance spectroscopy and multivariate data analysis [J]. *J Pharm Biomed Anal*, 2013, 75: 158-164.
- [3] 吴琪珍, 张朝凤, 许翔鸿, 等. 款冬花化学成分和药理活性研究进展 [J]. *中国野生植物资源*, 2015, 34(2): 33-36.
- [4] 李 静, 秦雪梅, 李震宇. 款冬花中倍半萜类成分的研究进展 [J]. *中草药*, 2017, 48(14): 2964-2971.
- [5] 李一平, 王筠默. 款冬花酮的升压机理 [J]. *中国药理学报*, 1986, 7(4): 333.
- [6] 李艳芳, 陈雪园, 熊卫艳, 等. 款冬酮抗过敏作用研究 [J]. *中国新药杂志*, 2015, 24(13): 1517-1522.

- [7] Ukiya M, Akihisa T, Yasukawa K, *et al.* Constituents of Compositae plants. 2. triterpene diols, triols, and their 3-O-fatty acid esters from edible *Chrysanthemum* flower extract and their anti-inflammatory effects [J]. *J Agric Food Chem*, 2001, 49(7): 3187-3197.
- [8] Yasukawa K, Akihisa T, Oinuma H, *et al.* Inhibitory effect of Di- and trihydroxy triterpenes from the flowers of compositae on 12-O-tetradecanoylphorbol-13-acetate-induced inflammation in mice [J]. *Biol Pharm Bull*, 1996, 19(10): 1329-1331.
- [9] Yasukawa K, Akihisa T, Oinuma H, *et al.* Inhibitory effect of taraxastane-type triterpenes on tumor promotion by 12-O-tetradecanoylphorbol-13-acetate in two-stage carcinogenesis in mouse skin [J]. *Oncology*, 1996, 53(4): 341-344.
- [10] Zitterl-Eglseer K, Sosa S, Jurenitsch J, *et al.* Anti-oedematous activities of the main triterpendiol esters of marigold (*Calendula officinalis* L.) [J]. *J Ethnopharmacol*, 1997, 57(2): 139-144.
- [11] Li C F, Zhu Y J, Guo X, *et al.* Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer [J]. *BMC Genom*, 2013, 14(1): 1-11.
- [12] Sun C, Li Y, Wu Q, *et al.* De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis [J]. *BMC Genom*, 2010, 11(1): 1-12.
- [13] 谢冬梅, 俞年军, 黄璐琦, 等. 基于高通量测序的药用植物“凤丹”根皮的转录组分析 [J]. 中国中药杂志, 2017, 42(15): 2954-2961.
- [14] 聂佳慧, 张福生, 田 栋, 等. 款冬叶不同发育阶段的转录组学分析 [J]. 中草药, 2018, 49(13): 3095-3101.
- [15] 贾 岩, 张福生, 肖淑贤, 等. 款冬花不同发育阶段的代谢组学和比较转录组学分析 [J]. 中国生物化学与分子生物学报, 2017, 33(6): 615-623.
- [16] 李 静, 李 娟, 贾金萍, 等. 基于 UHPLC-Q Extractive 轨道阱高分辨质谱的款冬花、叶的化学比较 [J]. 药学学报, 2018, 53(3): 444-452.
- [17] 许晓双, 张福生, 秦雪梅. 三萜皂苷生物合成途径及关键酶的研究进展 [J]. 世界科学技术—中医药现代化, 2014, 16(11): 2440-2448.
- [18] 郑夏生, 罗秀秀, 徐 晖, 等. 岗梅转录组及其乌索烷型三萜皂苷生物合成相关酶基因的发掘 [J]. 世界科学技术—中医药现代化, 2014, 16(7): 1505-1512.
- [19] 王兴春, 谭河林, 陈 钊, 等. 基于RNA-Seq技术的连翘转录组组装与分析及 SSR 分子标记的开发 [J]. 中国科学: 生命科学, 2015, 45(3): 301-310.
- [20] Chen S, Luo H, Li Y, *et al.* 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng* [J]. *Plant Cell Rep*, 2011, 30(9): 1593-1601.
- [21] 李 滢, 孙 超, 罗红梅, 等. 基于高通量测序 454 GS FLX 的丹参转录组学研究 [J]. 药学学报, 2010, 45(4): 524-529.
- [22] Chen J, Wu X T, Xu Y Q, *et al.* Global transcriptome analysis profiles metabolic pathways in traditional herb *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao [J]. *BMC Genom*, 2015, 16(7): 1-20.
- [23] Besumbes Ó, Sauret-Güeto S, Phillips M A, *et al.* Metabolic engineering of isoprenoid biosynthesis in *Arabidopsis* for the production of taxadiene, the first committed precursor of taxol [J]. *Biotechnol Bioeng*, 2004, 88(2): 168-175.