

• 药材与资源 •

茉莉酸甲酯诱导下远志幼苗转录组分析及三萜类生物合成途径关键酶基因挖掘

彭亮, 颜永刚, 陈莹, 沈霞, 高静, 李依民, 张岗, 张明英, 胡本祥, 杨冰月*, 孙静*
陕西中医药大学药学院/陕西省秦岭中草药应用开发工程技术研究中心, 陕西 西安 712046

摘要: 目的 获得远志 *Polygala tenuifolia* 响应茉莉酸甲酯 (methyl jasmonate, MeJA) 处理的转录组学信息, 挖掘远志三萜类化合物骨架生物合成的关键酶基因。方法 以生长 30 d 的远志组培苗为材料, 分别用无菌水 (CK)、50 $\mu\text{mol/L}$ MeJA、100 $\mu\text{mol/L}$ MeJA 处理 24 h, 采用 Illumina HiSeqTM 2000 150PE 进行转录组测序, 使用 Trinity 软件完成 Unigene 的 *de novo* 拼接, 基于 BLAST 实现 Unigene 的分类、功能注释、代谢通路分析、蛋白功能注释、差异基因分析和筛选等。结果 最终获得 52.19 Gb 数据, 通过 *de novo* 拼接注释得到 Unigene 54 426 条, 平均长度为 1 604 bp, 注释成功率 100%。通过对 MeJA 处理前后基因进行差异分析, 共筛选出差异基因 3 390 个, 其中有 1 287 个上调, 2 103 个下调, 且以 100 $\mu\text{mol/L}$ MeJA 处理的差异基因总数及上调数最高。KEGG 富集分析表明, 差异基因主要富集于苯丙烷类生物合成、半胱氨酸与蛋氨酸代谢、淀粉蔗糖代谢、光合生物的固碳作用、萜类骨架生物合成等 KEGG 通路中。找到与远志三萜类骨架生物合成相关的基因 59 条, 其中 AACT、HMGS、HMGR、MK、PMK、MPD、DXS、IDI、FPPS、SQS、SE 和 β -AS 受 MeJA 诱导后表达量上调。结论 对 MeJA 处理后的远志幼苗转录组进行分析, 获得远志三萜类骨架生物合成相关的候选基因, MeJA 可以诱导其三萜类骨架合成相关基因的表达, 为远志的分子生物学研究提供了丰富的数据资源, 也为后期开展远志三萜皂苷类化合物次生代谢途径解析奠定了基础。

关键词: 远志; 转录组; 茉莉酸甲酯; 三萜类; Unigene

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2020)09-2517-13

DOI: 10.7501/j.issn.0253-2670.2020.09.029

Transcriptome analysis of *Polygala tenuifolia* seedlings induced by methyl jasmonate and key genes mining for triterpenoid biosynthetic pathway

PENG Liang, YAN Yong-gang, CHEN Ying, SHEN Xia, GAO Jing, LI Yi-min, ZHANG Gang, ZHANG Ming-ying, HU Ben-xiang, YANG Bing-yue, SUN Jing

Shaanxi Qinling Application Development and Engineering Center of Chinese Herbal Medicine, College of Pharmacy, Shaanxi University of Chinese Medicine, Xi'an 712046, China

Abstract: Objective To obtain the transcriptome sequence database induced by methyl jasmonate (MeJA) and identify the genes related to the biosynthesis of triterpenoid saponin in *Polygala tenuifolia*. **Methods** The seedlings grown for 30 d were respectively treated with sterile water, 50 $\mu\text{mol/L}$ MeJA and 100 $\mu\text{mol/L}$ MeJA for 24 h. The transcriptome data of seedlings of *P. tenuifolia* were obtained by Illumina HiSeqTM 2000 150PE sequencing and *de novo* splicing of Unigene was realized by Trinity software. The GO classification, KOG functional annotation, metabolism of KEGG metabolic pathway, protein function annotation analysis, differential gene analysis and screening were completed based on BLAST. **Results** A total of 52.19 Gb clean data were obtained after the transcriptome of *P. tenuifolia* being assembled by Trinity software, and 54 426 Unigenes were assembled with an average length of 1 604 bp. All Unigenes were annotated in the public databases NR, NT, KEGG, Swissprot, GO, and Pfam. Through differential analysis of genes responding to MeJA, a total of 3 390 differentially expressed genes (DEGs) were found, of which 1 287 were up-regulated

收稿日期: 2020-01-06

基金项目: 公益性行业(中医药)科研专项经费项目(201507002-1-08); 陕西中医药大学“秦药”品质评价及资源开发学科创新团队项目(2019-QN01); 2019年医疗服务与保障能力提升补助资金(中医药事业传承与发展部分)“全国中药资源普查项目”(财社[2019]39号)

作者简介: 彭亮, 男, 副教授, 研究方向为中药资源评价与开发利用、分子药理学研究。E-mail: ppengliang@126.com

*通信作者 杨冰月, 女, 副教授, 研究方向为中药品种、质量及资源开发研究。E-mail: bingyiyang@126.com

孙静, 女, 教授, 研究方向为中药制剂过程的关键技术及适应性研究。E-mail: ph.175@163.com

and 2 103 were down-regulated. The response of DEGs showed that the total number and up-regulated number of *P. tenuifolia* seedlings treated by 100 $\mu\text{mol/L}$ MeJA was the highest. KEGG enrichment analysis showed that differentially expressed genes were significantly enriched in metabolic pathways including phenylpropanoid biosynthesis, cysteine and methionine metabolism, starch and sucrose metabolism, carbon fixation in photosynthetic organisms and terpenoid backbone biosynthesis. Furthermore, a total of 59 Unigenes involved in anthraquinones biosynthesis were found according to the assignment of KEGG pathway. Expression analysis showed that AACT, HMGS, HMGR, MK, PMK, MPD, DXS, IDI, FPPS, SQS, SE and β -AS were up-regulated after being induced by MeJA. **Conclusion** In this study, the transcriptome of *P. tenuifolia* seedlings treated with methyl jasmonate was analyzed, and candidate genes related to triterpenoid skeleton biosynthesis of *P. tenuifolia* were obtained. MeJA can induce the expression of genes related to triterpenoid skeleton synthesis, which provided a wealth of data resources for the molecular biology research and also laid the foundation for the analysis of the secondary metabolic pathways of triterpenoid saponins in *P. tenuifolia*.

Key words: *Polygala tenuifolia* Willd.; transcriptome; methyl jasmonate; triterpenoid; Unigene

远志 *Polygala tenuifolia* Willd. 为远志科 (Polygalaceae) 多年生草本植物, 其干燥根是我国传统中药材, 被《中国药典》2015 年版收录为远志药材来源^[1]。远志药材味苦、辛, 性温, 具有抗衰老、益智安神、祛痰镇咳、抗惊厥、镇静、保肝利胆和抗肿瘤等功效, 用于治疗心肾不交引起的失眠多梦、健忘惊悸等症^[2], 在临床益智药处方中的使用频率名列前 3 位^[3]。目前, 已从远志中分离得到远志皂苷类、吡啶类和糖酯类等多种化学成分, 其中远志皂苷是远志的特征性化学成分, 也是其主要的药效物质基础, 多达 50 余种, 且全部为齐墩果烷型五环三萜类化合物^[4]。现代药理研究表明, 远志皂苷具有极佳的抗痴呆、改善记忆障碍及脑保护等作用, 作用机制主要有抗神经细胞凋亡及抗氧化活性、胆碱能效应、阻遏炎症信号通路并减轻脾切除术认知障碍、改善海马区齿状回的长效增强效应、抗 β -淀粉样蛋白 (A β) 神经毒性等有关^[5-10]。

在高等植物中, 齐墩果烷型三萜类化合物的生物合成历经前体供应 (precursor supply)、骨架合成 (skeleton synthesis)、萜类合成 (saponin synthesis) 3 个阶段 (图 1)^[11]。第 1 阶段为基本前体物质异戊烯基焦磷酸 (isopentenyl diphosphate, IPP) 的生成 (C5), 涉及到的关键酶有甲羟戊酸途径 (MVA 途径) 的 (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, MK)、磷酸甲羟戊酸激酶 (phosphomevalonate kinase, PMK)、甲羟戊酸 5-焦磷酸脱羧酶 (mevalonate 5-diphosphate

decarboxylase, MPD) 和甲基赤藓醇磷酸途径 (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-磷酸还原异构酶 (1-deoxy-D-xylulose-5-phosphate reductoisomerase, DXR)、2-C-甲基-D-赤藓糖醇 4-磷酸胞苷酰转移酶 (2-C-methyl-D-erythritol 4-phosphate cytidylyl transferase, MCT)、4-(胞苷-5-二磷酸)-2-C-甲基赤藓糖醇激酶 [4-(cytidine-5-diphospho)-2-C-methylerythritol kinase, CMK]、2-C-甲基-D-赤藓糖醇 2,4-环二磷酸合酶 (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, MCS)、4-羟基-3-甲基-2-烯基二磷酸合酶 (4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, HDS)、4-羟基-3-甲基-2-烯基二磷酸还原酶 (4-hydroxy-3-methylbut-2-enyl diphosphate reductase, HDR)。第 2 阶段为形成共同的底物 2,3-氧化鲨烯 (2,3-oxidosqualene, OS) (C30), 涉及到的关键酶有异戊烯二磷酸 δ 异构酶 (isopentenyl-diphosphate delta-isomerase, IDI)、牻牛儿基焦磷酸合成酶 (GPPS)、法尼基焦磷酸合酶 (FPPS)、鲨烯合酶 (squalene synthetase, SQS)、鲨烯环氧酶 (squalene epoxidase, SE)。第 3 阶段 β -香树脂醇合成酶 (β -amyrin synthase, β -AS) 会催化 OS 形成 β -香树脂醇 (C30), 然后 β -香树脂醇在细胞色素 P450 (cytochrome P450, CYP450) 作用下使 C-28 氧化生成齐墩果酸, 最终由 CYP450 和糖基转移酶 (uridine diphosphate glycosyltransferase, UGT) 等催化产生官能化反应形成齐墩果烷型五环三萜皂苷^[4]。

迄今, 在模式植物中与三萜类骨架生物合成相关的功能基因均已被克隆并进行了广泛地研究, 人参、三七、西洋参等药用植物中也获得了大量与三

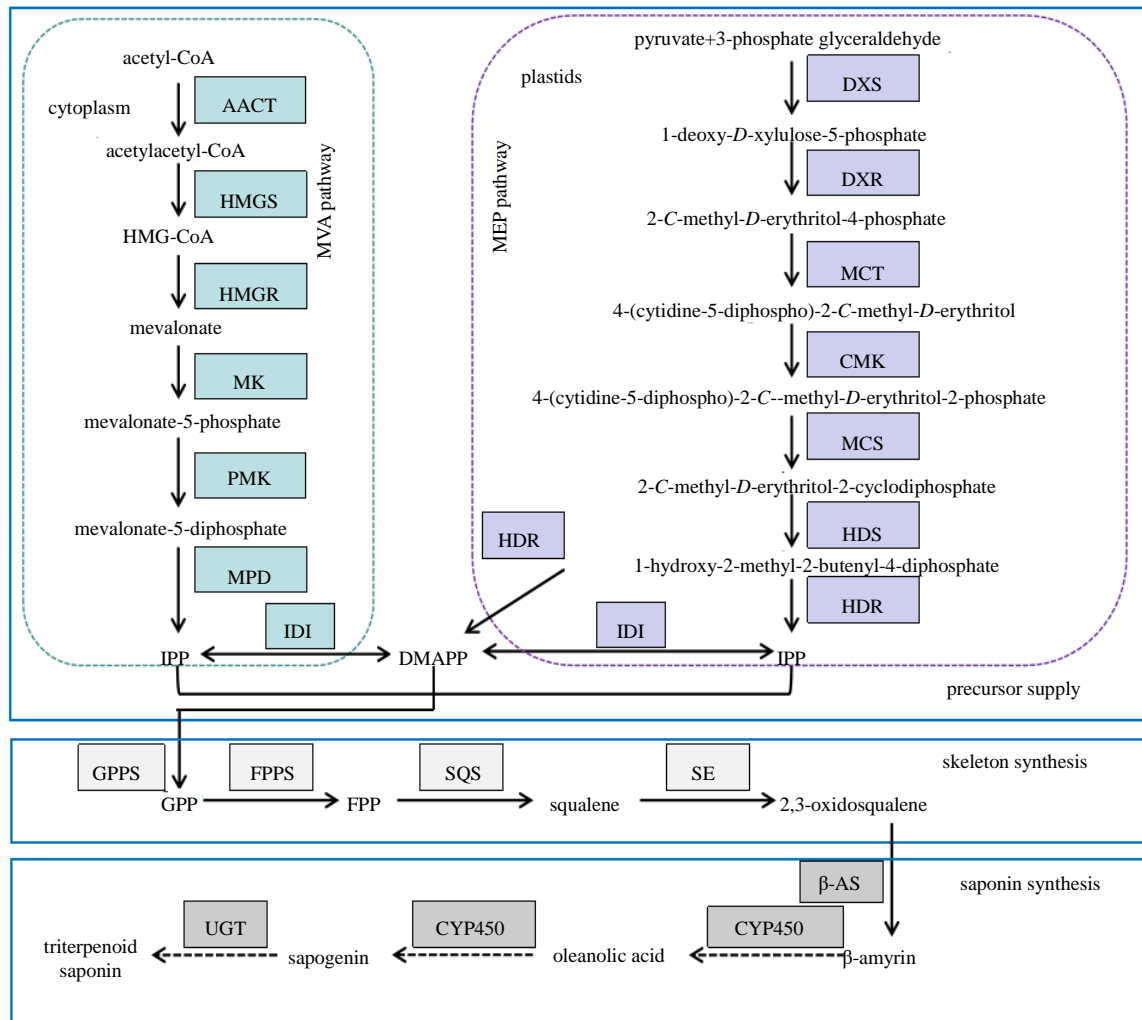


图 1 远志三萜皂苷类生物合成途径

Fig. 1 Biosynthesis framework of triterpene saponins of *P. tenuifolia*

萜类化合物合成相关的关键酶基因并验证了它们的功能^[12]。

然而，远志中三萜皂苷类仍主要以提取工艺、药用价值、药理作用研究居多，生物合成分子机制研究相对较少，相关报道有基于远志根转录组的相关分析^[13-14]、预测基因的组织特异性表达^[15]、2个氧化鲨烯环化酶（OSCs）的克隆及功能研究^[16]以及利用转录组测序揭示 MeJA 诱导的远志幼苗 OSC 基因分析^[17]等，对其生物合成途径和代谢机制依然知之甚少。MeJA 是植物次生代谢中重要的信号转导分子，通过信号转导可以激活或抑制相应转录因子的活性，并调控与植物次生代谢相关的关键酶基因的表达，影响关键酶的活性，从而调控次生代谢产物的合成^[18-19]。Rahimi 等^[20]研究表明在人参中，MeJA 诱导增强了 MEP 和 MVA 途径相关基因的表达以调控人参皂苷的合成；阳春砂幼苗喷施 MeJA 后其

AvHMGR、AvDXR 的表达显著上调^[21]。课题组前期采用不同浓度 MeJA 处理远志愈伤组织、幼苗、1 年生植株后，发现 MeJA 可以促进远志总黄酮、总酚、总皂苷、细叶远志皂苷、远志皂苷元、3,6'-二芥子酰基蔗糖等成分的积累（部分数据未发表）^[22]。因此，推测施加 MeJA 可能会激发远志中三萜类次生代谢产物合成相关基因的表达。本研究分别使用无菌水（CK）、50 μmol/L MeJA（MJ50）、100 μmol/L MeJA（MJ100）处理培养 30 d 的远志幼苗 24 h，采用 Illumina HiSeq™ 2000 150PE 进行转录组测序，构建 MeJA 处理前后远志幼苗的转录组文库，分析差异基因转录表达谱，挖掘与远志三萜类骨架生物合成相关的关键基因，以期揭示 MeJA 处理后远志幼苗的转录组整体特征及三萜类骨架合成相关基因的差异表达，为阐明远志三萜皂苷类成分生物合成途径及品质提高等奠定基础。

1 材料与amp;方法

1.1 材料

样品由陕西省王益区中药资源普查小组采集，经陕西中医药大学胡本祥教授鉴定为远志 *Polygala tenuifolia* Wild. 组培苗。培养温度为 20 ℃，光周期为 12 h 光照/12 h 黑暗，培养 30 d，取长势一致的远志幼苗，用无菌水及 50、100 μmol/L MeJA 处理 24 h 后取样，蒸馏水冲洗干净后，用滤纸吸干表面液体，液氮速冻后置于 -80 ℃ 冰箱备用。

1.2 RNA 提取及 cDNA 文库构建

采用 Qiagen RNeasy Mini Kit 提取样本总 RNA，去除 gDNA，检测完整性后，用带有 Oligo (dT) 的磁珠富集 mRNA；加入打断试剂后将 mRNA 打断成短片段，以打断后的 mRNA 为模版合成一链 cDNA 以及二链 cDNA，纯化回收、黏性末端修复、cDNA 的 3' 末端加上碱基 “A” 并连接接头，然后进行片段大小选择，最后进行 PCR 扩增，构建好的文库用 Agilent 2100 Bioanalyzer 质检。

1.3 测序及数据组装

使用 Illumina HiSeq™ 2000 150PE 进行测序。Trinity 对测序数据集进行 *de novo* 组装。

1.4 功能注释

所有组装的 Unigene 使用 DIAMOND 软件与 Nr 数据库 (<http://www.ncbi.nlm.nih.gov/>)、SwissProt 数据库 (<http://www.expasy.ch/sprot/>)、京都基因与基因组百科全书 (KEGG) (<http://www.genome.jp/kegg/>)、Pfam 数据库 (<http://pfam.xfam.org/>)、GO 数据库 (<http://www.geneontology.org>) 和 eggNOG 数据库 (<http://eggnogdb.embl.de/>) 进行对比，阈值小于 0.000 01。

1.5 差异表达基因水平分析

RSEM 对 bowtie 的比对结果进行统计，进一步得到了每个样品比对到每个基因上的 readcount 数目，并对其进行 FPKM 转换，进而分析 Unigene 的表达水平。采用 TMM 对 read count 数据进行标准

化处理，再用 DEGseq 进行差异分析，筛选阈值 (q value) < 0.005 且 $|\log_2(\text{fold change})| > 1$ 。GO 功能显著性富集分析首先把所有差异表达基因向 Gene Ontology 数据库 (<http://www.geneontology.org/>) 的各个 term 映射，计算每个 term 的基因数目，然后找出与整个基因组背景相比，在差异表达基因中显著富集。Pathway 显著性富集分析以 KEGG Pathway 为单位，应用超几何检验，找出差异基因相对于全部有注释的基因显著富集的 pathway。

1.6 远志三萜类骨架合成相关基因的挖掘与差异性分析

依据文献提出的高等植物三萜类化合物骨架合成途径，结合 KEGG 注释结果和数据库中已知的基因信息，利用本地 Blast 进行检索比对，确定本转录组数据库中与远志三萜类化合物骨架合成相关的基因。提取转录组数据库中关键基因的 FPKM 值进行差异性分析。

1.7 实时荧光定量 PCR (qRT-PCR) 检测

采用 qRT-PCR 对远志转录组数据 (CK、MJ50、MJ100) 中三萜类骨架合成的代表性差异基因进行表达量检测。使用管家基因 β -actin 作为内参基因，并根据各核苷酸序列片段设计 qRT-PCR 引物 (表 1)。PCR 反应体系为 20 μL: cDNA 约 150 ng, SYBR® Premix EX Taq TM II (2×) 10 μL, 正反引物各 1.0 μL, 加 ddH₂O 至 20 μL。PCR 程序为 95 ℃、30 s; 95 ℃、5 s, 55 ℃、30 s, 72 ℃、1 min, 40 个循环，熔解曲线条件设置为 65~95 ℃ 5 s, 每升高 0.5 ℃ 时收集一次荧光，结果采用 2^{-ΔΔCt} 法进行计算，3 次重复。

2 结果与分析

2.1 RNA 提取与质量控制

用 Qiagen RNeasy Kit 提取远志幼苗总 RNA，琼脂糖凝胶电泳检测 RNA 质量，Nanodrop 检测 RNA 的纯度 (A_{260}/A_{280} 值)，Qubit 进行 RNA 浓度的精确定量，Agilent 2100 精确检测 RNA 的完整性，

表 1 荧光定量 PCR 引物序列

Table 1 Real-time PCR primers

基因	正向引物 (5'-3')	反向引物 (3'-5')
HMGR	CGTCGTTTATCTACCTTATTTGGC	GCTTTTATCACCTCCTCATCCT
MK	CGGTTGTTTCATGGATCTACTGC	TCAAGTCCCATATCCTTCAGC
DXS	GCTATGAATAATGCTGGCTACCT	GCTCCTTGACCACTCACTACTC
SE	TATGAAGCGACCAATTGTGAAAG	GCTCAACGGATCTATGGGTAT
β -actin	GCTGCCCTGAGGTCCTCTT	TGGAACCACCGATCCAGACA

结果显示 RNA 质量和完整性较高 (RIN>7.5), 将同一处理 3 个重复 RNA 混匀后进行高通量测序, 构建 CK、MJ50 和 MJ100 3 个转录组文库。

2.2 测序与组装结果

用 Illumina HiSeq™ 2000 150PE 平台进行远志幼苗转录组的高通量测序, 共获得 357 326 062 条 raw reads, 经过测序质量控制, 得到 347 899 116 条 clean reads 和 52.19 G clean bases, Q_{20} 平均值为 97.57%, Q_{30} 达 92.85%, GC 含量为 43.71%, 测序质量非常高, 对原始数据进行过滤后, 剔除了低质量的数据, 能够满足后续分析研究的需要。

采用 Trinity 软件对序列进行拼接, 共获 215 567 条 transcripts, 在 transcripts 数据的基础上, 进一步对序列进行组装, 共获得 54 426 条 Unigene, 具体情况见表 2。

表 2 远志转录组组装结果

Table 2 Summary of transcript assembly generated from *P. tenuifolia*

长度/bp	转录本数量/条	独立基因数量/条
200~500	35 330	8 119
500~1 000	42 415	17 236
1 000~2 000	56 630	13 200
>2 000	81 192	15 871
总条量	215 567	54 426
总长度	397 432 127	87 296 697
N ₅₀ 长度	2 689	2 445
N ₉₀ 长度	926	712
平均长	1 844	1 604

2.3 Unigene 功能注释

通过选择 BLAST 参数 E 值 $<1 \times 10^{-5}$ 和 HMMER 参数 E 值 $<1 \times 10^{-10}$, 最终所有 Unigene 都被注释到各数据库中, 见表 3。其中 KEGG、KOG 数据库中注释的 Unigene 数量在 15 000 以下, 其余数据库的注释数量则都超过 20 000 条, 在 NR 数据库中的注释数最多, 为 31 147 条。

Unigene 与 NR 数据库进行 BLAST 比对, 发现本研究转录组测序获得的 Unigene 与其他物种的匹配度并不高, 均低于 10%, 其中欧洲栓皮栎 *Quercus suber* L. 所占比例最高, 为 8.81%, 其后依次是胡桃 *Juglans regia* L. (6.36%)、大豆 *Glycine max* (Linn.) Merr. (4.95%)、枣 *Cajanus cajan* L. (4.94%) 和绒毛烟草 *Ziziphus jujuba* Mill. (3.91%), 其他物种为 71.03% (图 2)。

表 3 远志 Unigenes 注释结果

Table 3 Summary statistics of functional annotation for *P. tenuifolia* sequences in public databases

注释数据库	数目	百分比/%
NR	31 147	57.22
NT	21 490	39.48
KEGG	12 197	22.41
SwissProt	23 721	43.58
PFAM	23 800	43.72
GO	23 800	43.72
KOG	6 993	12.84
Annotated in all databases	4 398	8.08
Annotated in at least one database	35 353	64.95
总 Unigene 数	54 426	100.00

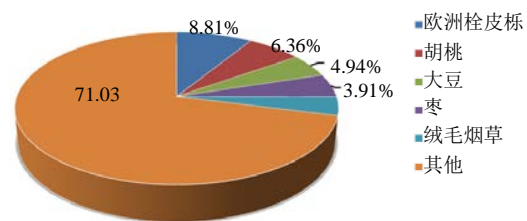


图 2 NR 数据库比对结果的主要物种分布

Fig. 2 Summary statistics of species classification in NR databases

2.4 GO 分类

共有 37 127 条 Unigene 被注释到 GO 分类 (图 3), 其中, 样本基因数量在 10 000 条以上且功能在参与的生物过程 (biological process) 分类中主要聚集于细胞过程 (cellular process, 13 907 条), 代谢过程 (metabolic process, 12 839 条) 和单组织过程 (single-organism process, 10 104 条); 在分子功能 (molecular function) 分类中主要聚集于连接 (binding, 13 415 条) 和催化活性 (catalytic activity, 10 572 条); 而在细胞组分 (cellular component) 中主要聚集于细胞 (cell, 7 763 条) 和细胞部位 (cell part, 7 763 条)。

2.5 KOG 功能分类

由图 4 可知, 共有 15 015 条 Unigene 被注释到 25 种 KOG 分类中, 涉及了大多数的生命活动。其中, 翻译后修饰 (posttranslational modification)、蛋白质周转 (protein turnover) 和分子伴侣 (chaperones) 是最大类别, 包含 1 000 条 Unigene, 占被注释到 Unigene 总数的 14.35%; 其次是功能预

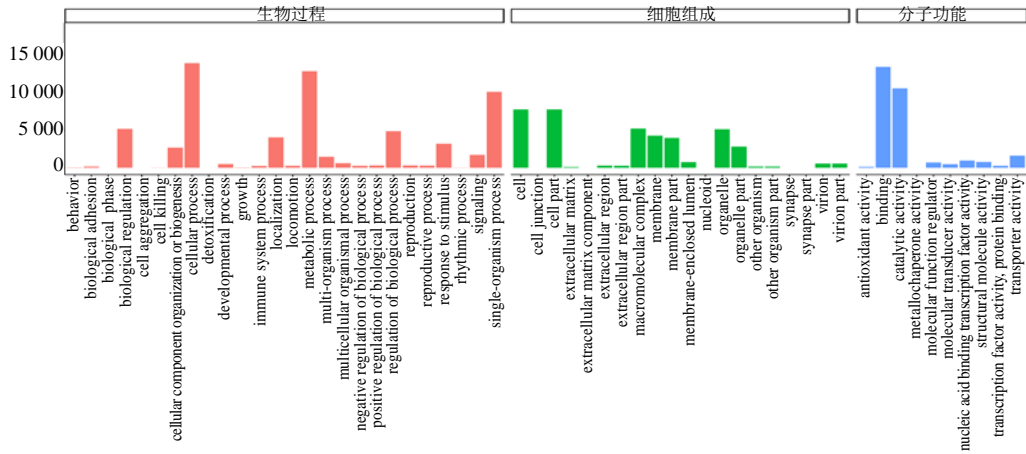


图 3 Unigene 的 GO 分类结果

Fig. 3 GO classification of assembled Unigene

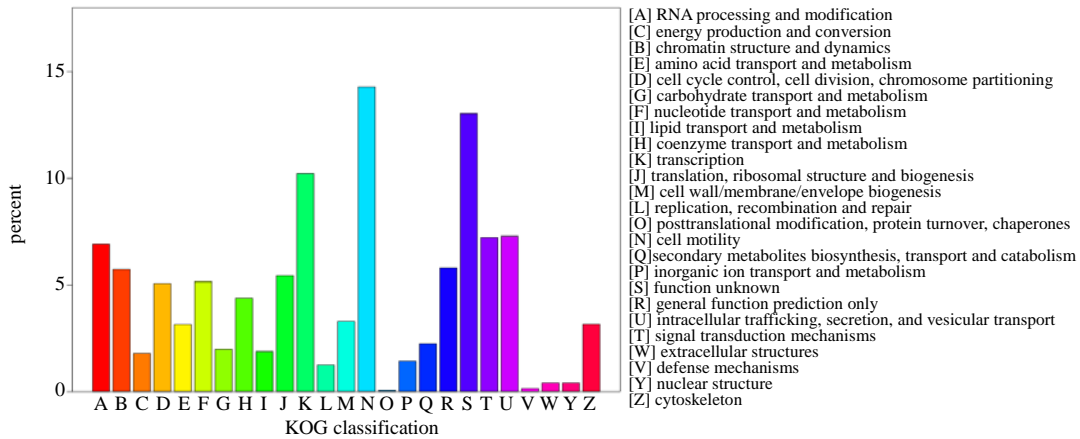


图 4 Unigene 的 KOG 分类结果

Fig. 4 KOG classification of assembled Unigene

测类基因 (general function prediction only), 包含 912 条 Unigene; 而细胞外结构 (extracellular structures) 和细胞运动 (cell motility) 基因较少, 分别为 11 条和 4 条; 其他类别的基因表达丰度都各不相同。

2.6 KEGG 代谢通路分析

通过比对, 在 KEGG 数据库中共有 12 197 条 Unigene 得到注释, 涉及到 5 条主通路: 细胞过程 (A, cellular processes), 环境信息处理 (B, environmental information processing), 遗传信息处理 (C, genetic information processing), 代谢 (D, metabolism), 有机系统 (E, organismal systems); 19 条子通路。结果表明 (图 5), 以碳水化合物代谢 (Carbohydrate metabolism) 的最多, 为 1 051 条; 其次为翻译 (translation) 和折叠 (folding)、分类 (sorting)、降解 (degradation), 分别有 904 和 722 条; 最少的是膜运输 (membrane transport), 为 59 条。

2.7 差异表达基因的分析与鉴定

以同条件下无菌水处理的远志幼苗作对照, 分析了 50、100 $\mu\text{mol/L}$ MeJA 处理 24 h 后的差异表达基因。用 FPKM 值反映对应基因的表达量, 根据 $\text{FDR} < 0.05$ 且 $|\log_2\text{FC}| > 1$ 筛选差异表达基因, 共筛选出 3 390 个差异基因 (图 6), 其中 MJ50 对 CK 的总变异 Unigene 1 321 条, 上调 Unigene 416 条, 下调 905 条; MJ100 对 CK 的总变异 Unigene 1 960 条, 上调 Unigene 831 条, 下调 1 129 条; MJ100 对 MJ50 的总变异 Unigene 109 条, 上调 Unigene 40 条, 下调 69 条, 说明本研究中 MeJA 处理下远志幼苗中有更多的基因表达受到抑制。

随后, 为了鉴定不同 MJ 处理下远志幼苗中共同和特异的差异表达基因, 用不同组间的 DEGs 生成韦恩图 (图 6), 可以看出, 在 CK vs MJ50 与 CK vs MJ100 中, DEGs 的重叠部分达到最大 (1 199 条), 在 MJ100 处理时, DEGs 的特异性表达达到

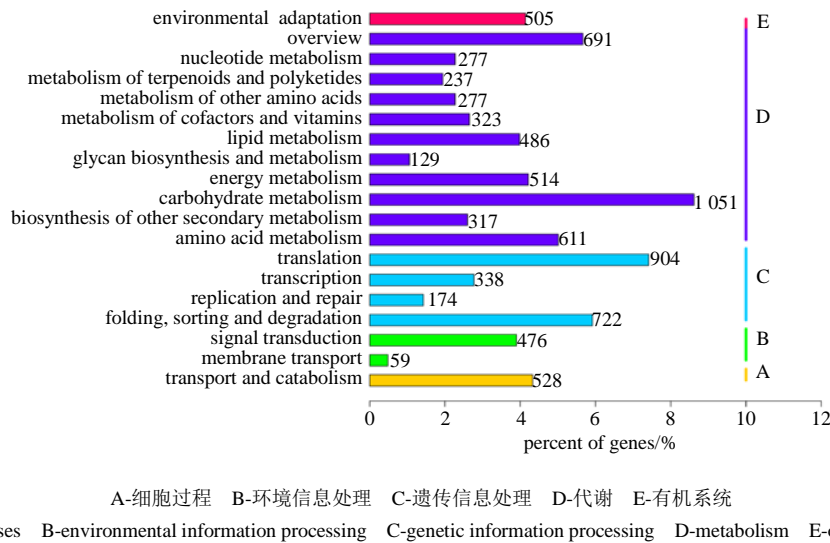


图 5 Unigene 的 KEGG 分类结果

Fig. 5 KEGG classification of assembled Unigene

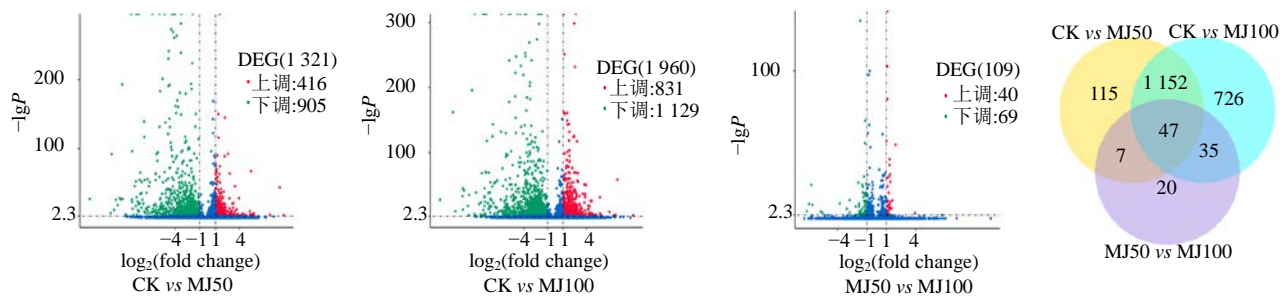


图 6 差异表达基因火山及韦恩图

Fig. 6 Volcano plot and venn diagram of differentially expressed genes

最高 (726 条), 而 MJ50 vs MJ100 中, DEGs 的特异性表达最低 (20 条)。同时, 由图可知, 随着 MeJA 浓度的增加, MJ50 vs MJ100 组中 DEGs 的重叠部分显著降低, 说明远志幼苗中大部分基因在 MeJA 处理后出现了差异, 但和浓度之间的关系较小。

2.8 差异表达基因的 GO 分类与富集分析

Gene ontology (Go) 是用来全面描述生物体中基因和基因产物的功能分类体系。本研究中 CK vs MJ50、CK vs MJ100、MJ50 vs MJ100 分别有 982、1 487、81 条差异基因在 GO 数据库中得到注释信息, 在 CK vs MJ50 组差异表达基因中, 与生物过程相关的主要富集于代谢过程 (metabolic process, 641 条)、细胞过程 (cellular process, 574 条), 与细胞组分相关的主要富集于膜 (membrane, 228 条)、细胞部位 (cell part, 221 条) 和细胞内 (intracellular, 221 条); 与分子功能相关的主要富集于催化活性 (catalytic activity, 674 条)、连接

(binding, 496 条) 等; CK vs MJ100 组差异表达基因中, 与生物过程 (Molecular Function) 相关的主要富集于代谢过程 (metabolic process, 974 条)、细胞过程 (cellular process, 862 条), 与细胞组分相关的主要富集于膜 (membrane, 410 条) 和细胞部位 (cell part, 410 条), 与分子功能相关的主要富集于细胞质 (cytoplasm, 957 条)、连接 (binding, 786 条) 等; MJ50 vs MJ100 组差异表达基因中, 与生物过程 (molecular function) 相关的主要富集于代谢过程 (metabolic process, 54 条)、单组织过程 (single-organism process, 47 条), 与细胞组分相关的主要富集于膜 (membrane, 19 条)、细胞部位 (cell part, 16 条) 和细胞内 (intracellular, 16 条); 与分子功能相关的主要富集于催化活性 (catalytic activity, 59 条)、连接 (binding, 34 条) 等 (图 7), 说明 MeJA 处理后明显影响了远志幼苗的代谢过程, 以 MJ100 处理更为显著。

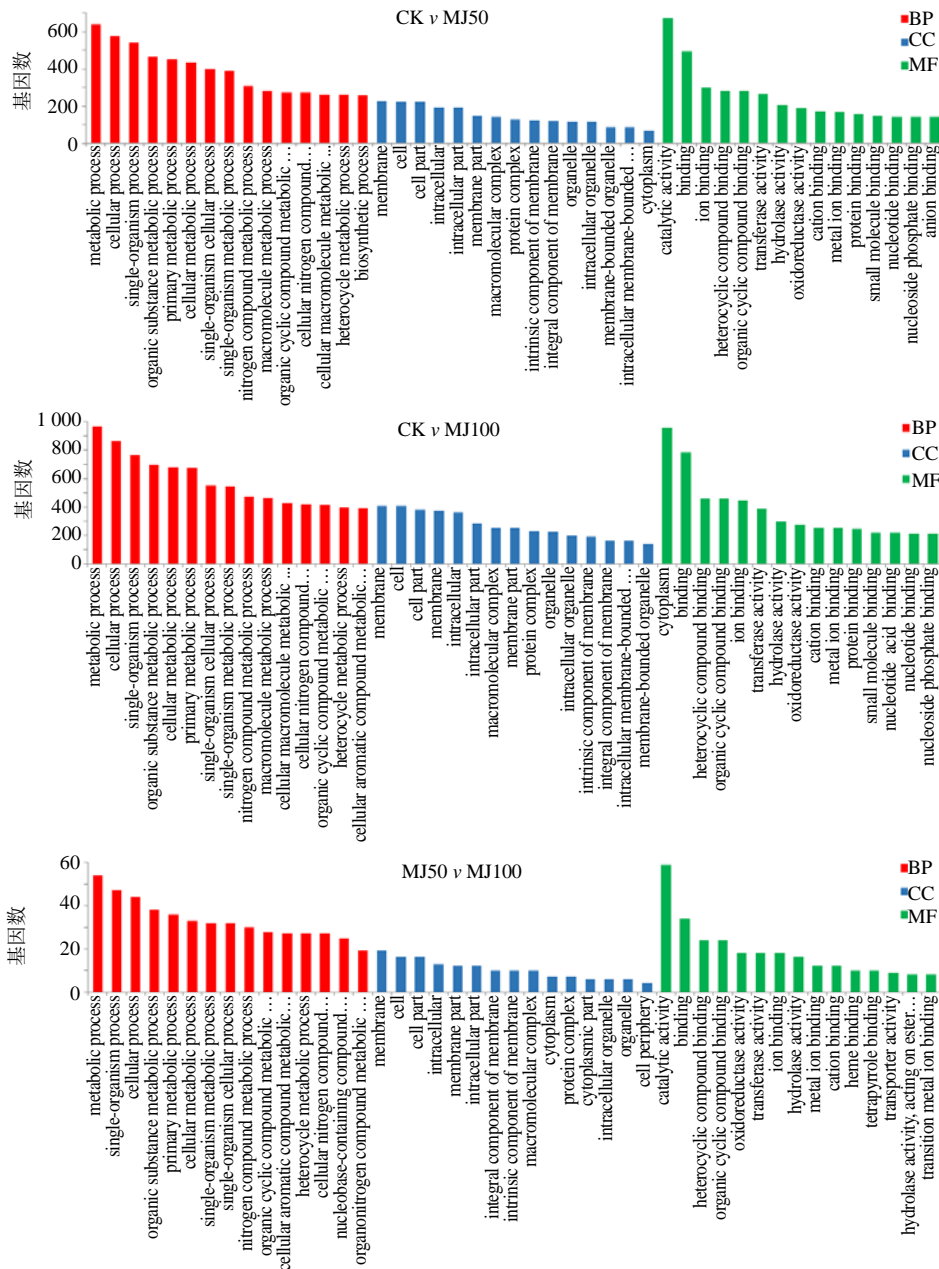


图 7 差异表达基因 GO 注释分类

Fig. 7 GO annotation of differently expressed gene

2.9 差异表达基因的 KEGG 代谢途径分析

通过与 KEGG 数据库进行比对来探究远志幼苗受 MeJA 诱导影响的代谢途径, CK vs MJ50、CK vs MJ100、MJ50 vs MJ100 各有 655、916、56 条差异基因映射到 KEGG Pathway 数据库中, 分别涉及了 96、99、27 个代谢通路。差异表达基因参与最多的 6 个通路是: CK vs MJ50 组中为苯丙烷类生物合成 (phenylpropanoid biosynthesis, 45 条)、半胱氨酸与蛋氨酸代谢 (cysteine and methionine metabolism, 39 条)、淀粉蔗糖代谢 (starch and sucrose metabolism,

27 条)、光合生物的固碳作用 (carbon fixation in photosynthetic organisms, 25 条)、萜类骨架生物合成 (terpenoid backbone biosynthesis, 20 条)、氨基糖和核苷酸糖代谢 (amino sugar and nucleotide sugar metabolism, 20 条); CK vs MJ100 组中为苯丙烷类生物合成 (57 条)、半胱氨酸与蛋氨酸代谢 (47 条)、淀粉蔗糖代谢 (34 条)、糖酵解/糖异生 (glycolysis/gluconeogenesis, 32 条)、二羧酸代谢 (glyoxylate and dicarboxylate metabolism, 30 条)、植物激素信号转导 (plant hormone signal transduction,

30 条), 另外萜类骨架生物合成为 20 条; MJ50 vs MJ100 组中为苯丙氨酸代谢 (8 条)、糖酵解/糖异生 (5 条)、戊糖磷酸途径 (pentose phosphate pathway, 4 条)、果糖和甘露糖代谢 (fructose and mannose metabolism, 4 条)、光合生物的固碳作用 (4 条)、植物激素信号转导 (4 条)、植物病原

相互作用 (plant-pathogen interaction, 4 条)、未见萜类骨架生物合成 (图 8)。说明萜类骨架生物合成通路与远志幼苗响应 MeJA 处理的应答相关, 可以根据差异基因的功能注释进一步挖掘具有远志三萜皂苷类生物合成功能的基因, 用于基因工程育种。

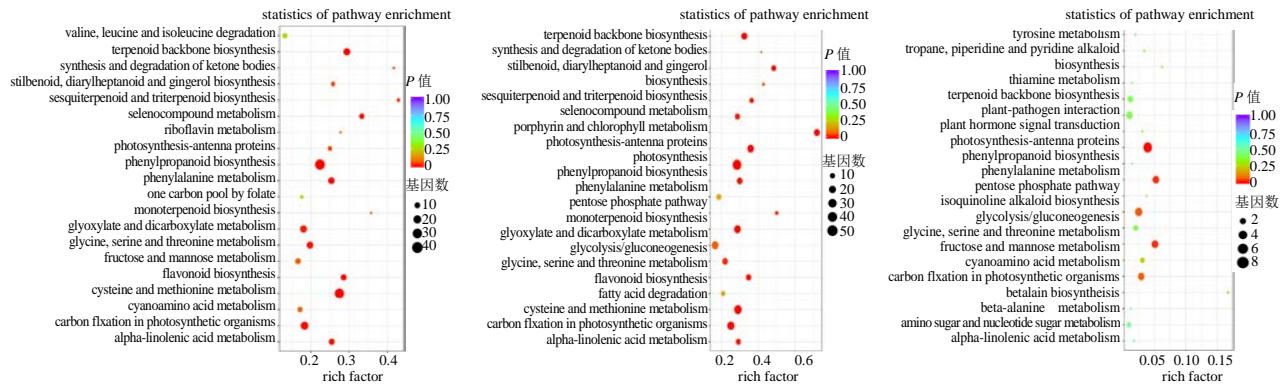


图 8 差异表达基因在 KEGG 中的分类统计富集

Fig. 8 KEGG pathway annotation of differentially expressed genes

2.10 远志三萜类化合物骨架生物合成的关键基因筛选

通过对远志根转录组注释所得的 Unigene 进行分析, 涉及到远志三萜类化合物骨架生物合成的 18 个关键酶和催化 2,3-氧化鲨烯环化形成齐墩果烷型三萜的前体 β -香树脂醇的 β -AS 全部被注释到, 共有 59 条 Unigene, 具体信息见表 4。

2.11 远志三萜类化合物骨架生物合成途径中关键基因的表达分析

由于注释到的 19 基因大部分都有多个 Unigene, 通过分析发现这些 Unigene 的表达量虽相差很大但其表达模式是一致的, 故以所有 Unigene 的表达量之和作为注释基因表达丰度的衡量。从图 9 可以看出, MeJA 处理后远志三萜类骨架生物合成途径第 1 阶段 MVA 途径的 AACT、HMGS、HMGR、MK、PMK、MPD 的表达量上调, 以 MJ100 处理后值最大; MEP 途径 DXS 的表达量上调, DXR、MCS、HDR 的表达量下调, MCT、CMK、HDS 的表达量则变化不大。第 2 阶段除 GPPS 的表达量下调外, IDI、FPPS、SQS、SE 的表达量都上调, β -AS 的表达量也随着处理浓度的增加出现上调。差异性分析表明, 远志三萜类皂苷途径中, 除 MEP 途径中的 3 个基因: MCT、CMK、HDS 外, 均在 MeJA 诱导后出现了显著差异, 推测 MeJA 通过诱导 MVA 途径中的关键酶基因表达以合成远志

三萜类产物的通量较大, 而此时 MEP 途径则可能参与远志中其他支路次生代谢产物的合成。

采用荧光定量 PCR 技术对 CK、MJ50、MJ100 中远志三萜类化合物骨架生物合成途径的 4 个代表性差异基因进行表达量检测, 结果显示 HMGR、MK、DXS、SE 的相对表达量与转录组数据库中相应的 FPKM 变化趋势基本一致, 其中 SE 转录表达量最高, 也是 CK 和 MJ50、MJ100 中转录活性差异最大的基因。同时, 对不同处理间差异基因的表达量进行了差异性分析, 可知 MeJA 处理极显著的诱导了相关基因表达量的上调, 而且不同浓度 MeJA 处理间的基因表达量存在显著或极显著的差异 (图 10)。

3 讨论

远志作为我国传统中药, 远志皂苷 (齐墩果烷型五环三萜) 是其主要的药效成分, 关于其化学成分和药理作用方面的研究已有良好的基础, 但是远志的基因组信息未知, 严重限制了远志皂苷生物合成途径及相关功能基因的研究。本实验利用 Illumina HiSeq™ 2000 150PE 高通量测序技术对 MeJA 处理前后的远志幼苗进行转录组测序, 得到 347 899 116 条 clean reads 和 52.19 G clean bases, 组装后获得 54 426 条 Unigene, N50 为 2 445 bp, 平均长度 1 604 bp, 所有 Unigene (100%) 被注释到 NR、NT、COG、KEGG、SwissProt、PFAM、GO

表 4 远志三萜类化合物骨架生物合成相关候选基因

Table 4 Candidate genes involved in triterpenoid skeleton biosynthesis identified in *P. tenuifolia*

基因名称	Ko 号	数目	功能
acetyl-CoA C-acetyltransferase (AACT)	K00626	2	transferase activity, transferring acyl groups other than amino-acyl groups, 3-oxoacyl-[acyl-carrier-protein] synthase activity
hydroxymethylglutaryl-CoA synthase (HMGS)	K01641	4	hydroxymethylglutaryl-CoA synthase activity, microtubule binding, extracellular ATP-gated cation channel activity, DNA-directed DNA polymerase activity, purinergic nucleotide receptor activity
hydroxymethylglutaryl-CoA reductase (HMGR)	K00021	7	hydroxymethylglutaryl-CoA reductase (NADPH) activity, coenzyme binding
mevalonate kinase (MK)	K00869	3	oxidoreductase activity, metal ion binding, ATP binding
phosphomevalonate kinase (PMK)	K00938	2	DNA binding, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors, phosphomevalonate kinase activity, methyltransferase activity
mevalonate 5-diphosphosphate decarboxylase (MPD)	K01597	1	ATP binding, actin binding
1-deoxy-D-xylulose-5-phosphate synthase (DXS)	K01662	7	1-deoxy-D-xylulose-5-phosphate synthase activity, catalytic activity
1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)	K00099	2	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor, protein binding, NADPH binding
2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (MCT)	K00991	2	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase activity, nucleotidyltransferase activity, uridylyltransferase activity
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS)	Ko1770	2	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase activity
4-(cytidine-5-diphospho)-2-C-methylerythritol kinase (CMK)	K00919	2	nucleic acid binding, zinc ion binding
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS)	K03526	1	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity, oxidoreductase activity, antioxidant activity
4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR)	K03527	3	4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase activity, metal ion binding
isopentenyl-diphosphate delta-isomerase (IDI)	K01823	2	hydrolase activity
geranyl diphosphate synthase (GPPS)	K13789	6	transferase activity, transferring alkyl or aryl (other than methyl) groups
farnesyl diphosphate synthase (FPPS)	K00787	1	transferase activity
squalene synthase (SQS)	K00801	2	transferase activity
squalene epoxidase (SE)	K00511	6	squalene monooxygenase activity, FAD binding, flavin adenine dinucleotide binding, oxidoreductase activity
β -amyrin synthase (β -AS)	K15813	2	1-deoxy-D-xylulose-5-phosphate synthase activity, ATP binding, kinase activity, protein binding, catalytic activity

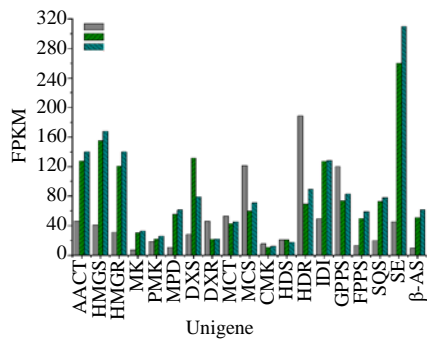
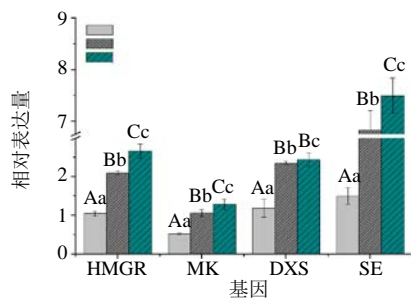


图 9 关键基因的表达模式分析

Fig. 9 Analysis of expression patterns of key genes



不同小写字母示差异显著, $P < 0.05$; 不同大写字母示差异极显著, $P < 0.01$
different lowercase letters show significant differences, $P < 0.05$;
different capital letters show extremely significant differences, $P < 0.01$

图 10 远志三萜皂苷类骨架合成关键基因的定量 PCR 检测
Fig. 10 Quantitative PCR detection of key genes for triterpenoid saponin skeleton synthesis in *P. tenuifolia*

和 COG 数据库,说明本次远志转录组测序结果所得数据量大,质量较好,可满足后续数据分析的要求, Illumina HiSeq 高通量测序技术可作为批量挖掘远志皂苷生物合成途径功能基因的有效工具。同时,本实验构建的远志幼苗转录组数据库中 Unigene 数少于已有报道的远志根转录组数据^[13-14]、MJ100 处理培养基培养 50 d 后的远志幼苗转录组数据^[17],但 N₅₀、平均长度等占优,推测 Unigene 数较少与本实验所用材料为 30 d 远志幼苗相关。

植物体中三萜类化合物的整个代谢过程涉及到的酶数目众多,因此关键酶的确定至关重要^[19]。基于高通量测序的转录组文库可以从合成途径入手,克服单个酶的表达与产物关系的局限性,而从整体水平上了解各个酶与产物的关系^[23]。如 Sun 等^[24]通过对 MeJA 处理前后的红豆杉培养细胞进行转录组测序,获得了 40 348 条 Unigene,

注释后发现大多数与转录调控、DNA 和组蛋白修饰以及 MeJA 信号网络相关,为解析 MeJA 调控红豆杉中紫杉醇的生物合成途径提供了丰富的数据信息。三萜皂苷根据皂元结构又分为五环三萜和四环三萜,相应催化形成三萜类物质的三萜合酶分为合成四环三萜的达玛烷合成酶 (dammarendiol synthase, DS) 和合成五环三萜的香树素合成酶 (amyrin synthase, AS),本研究转录组数据匹配到的三萜合酶为催化 OS 形成齐墩果烷型三萜的 β -AS,并没有序列匹配到四环三萜相关的合成酶 DS,与远志皂苷类型为齐墩果烷型三萜皂苷一致。在此基础上,参考其他植物中三萜类化合物的生物合成途径,对远志根转录组注释所得的 Unigene 进行分析,涉及到远志皂苷骨架生物合成的 19 个关键酶基因都被注释到,共获得相关 Unigene 59 条 (表 3)。

MeJA 具有广谱的生理效应,除可以调节植物的生长和发育外,还参与植物对机械伤害、病害等逆境的防御响应以及诱导次生代谢产物的合成^[18]。研究表明,MeJA 可以通过诱导药用植物三萜类物质的生物合成途径关键酶基因的表达从而促进其成分的积累,如灵芝中,MeJA 处理可以显著提高 HMGR、SQS 和 OSC 基因的表达水平,从而促进灵芝三萜的生物合成^[25];京大戟愈伤组织中,MeJA 处理可以诱导 HMGR、SQS、FPPS 基因的表达,使大戟三萜类物质含量明显提高^[26]。MeJA 处理远志幼苗后 (数据未列出),相比 CK,远志总皂苷含量分别提升 3.34 倍 (MJ50)、6.27 倍 (MJ100);远志皂苷元含量分别提高 1.61 倍 (MJ50)、2.54 倍 (MJ100);细叶远志皂苷分别提高 1.45 倍 (MJ50)、2.82 倍 (MJ100),与腾红梅等^[27]的研究结果一致,证实 MeJA 可显著提高远志皂苷类成分的含量。转录组学差异分析结果则表明,MeJA 处理后远志转录组共有 2 550 条 DEG 在 GO 数据库中得到注释,主要富集的 GO 功能有催化活性 (GO: 0003824)、代谢过程 (GO: 0008152)、细胞过程 (GO: 0009987)、单组织过程 (GO: 0044699)、连接 (GO: 0005488)、膜 (GO: 0016020)、细胞部位 (GO: 0044464) 等。同时,MeJA 处理后共有 1 627 条差异基因映射到 KEGG Pathway 数据库中,涉及 99 个代谢通路,差异基因参与较多的通路有苯丙烷类生物合成、半胱氨酸与蛋

氨酸代谢、淀粉蔗糖代谢、糖酵解/糖异生、二羧酸代谢、植物激素信号转导、萜类骨架生物合成等；关键基因的差异表达分析进一步表明 MeJA 上调了远志三萜类骨架生物合成途径相关酶基因：AACT、HMGS、HMGR、MK、PMK、MPD、DXS、IDI、FPPS、SQS、SE 和 β -AS 的表达量，下调了 DXR、MCS、HDR 的表达量，推测 MeJA 可通过诱导第一阶段 MVA 途径和第二阶段中的关键酶基因表达而提高远志三萜皂苷类成分的积累，而 MEP 途径中的部分相关酶基因则可能参与其他支路次生代谢产物的合成。

本研究利用转录组测序技术构建了 MeJA 诱导前后的远志幼苗的转录组学文库，并获得了远志三萜皂苷类化合物骨架生物合成的关键酶基因，对这些酶基因进一步的差异表达分析和荧光定量 PCR 验证能有效阐明 MeJA 调控远志皂苷生物合成的分子机理。可以为远志三萜皂苷类生物合成途径的进一步完善，关键酶基因的克隆与功能验证，以及通过外源激素和基因工程调控目标产物提供一定的数据支持。

参考文献

- [1] 中国药典 [S]. 一部. 2015.
- [2] 彭亮, 杨冰月, 张琳, 等. 野生与栽培远志 HPLC 指纹图谱及化学模式识别研究 [J]. 中草药, 2018, 49(21): 4998-5003.
- [3] Cao Q, Jiang Y, Cui S Y, *et al.* Tenuifolin, a saponin derived from *Radix Polygalae*, exhibits sleep-enhancing effects in mice [J]. *Phytomedicine*, 2016, 23: 1797-1805.
- [4] 张福生, 孔冉冉, 陈彤垚, 等. P450s 介导远志皂苷等齐墩果烷型植物三萜生物合成的研究进展 [J]. 药学报, 2019, 54(6): 1000-1009.
- [5] 莫微, 林志宏. 中药远志改善记忆障碍研究进展 [J]. 世界科学技术—中医药现代化, 2012, 14(4): 1913-1916.
- [6] Chen Y J, Huang X B, Li Z X, *et al.* Tenuigenin protects cultured hippocampal neurons against methylglyoxal-induced neurotoxicity [J]. *Eur J Pharmacol*, 2010, 645: 1-8.
- [7] 陈庆林, 陈勤, 金蓓蓓. 远志皂苷对 AD 小鼠学习记忆能力及中枢胆碱能系统标志酶活性的影响 [J]. 中药药理与临床, 2011, 27(3): 33-36.
- [8] Yu L Y, Sun L, Chen S L. Protective effect of senegenin on splenectomy-induced postoperative cognitive dysfunction in elderly rats [J]. *Exp Ther Med*, 2014, 7: 821-826.
- [9] Sun F, Sun J D, Han N, *et al.* Polygala saponin F induces long-term potentiation in adult rat hippocampus via NMDA receptor activation [J]. *Acta Pharmacol Sin*, 2012, 33: 431-437.
- [10] Li X, Cui J, Yu Y, *et al.* Traditional Chinese nootropic medicine *Radix Polygalae* and its active constituent onjisaponin B reduce β -amyloid production and improve cognitive impairments [J]. *PLoS One*, 2016, 11(3): e0151147.
- [11] 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.
- [12] 张艳敬, 候志芳, 梁韶, 等. 人参属药用植物三萜皂苷合成途径关键酶的研究进展 [J]. 特产研究, 2016, 38(2): 53-57.
- [13] 李晓伟. 基于高通量转录组测序与植物代谢组学技术研究远志皂苷生物合成途径 [D]. 太原: 山西大学, 2014.
- [14] Kikuchi A, Yamamura Y, Murakami Y, *et al.* Discovery of genes involved in onjisaponin biosynthesis from *Polygala tenuifolia* [J]. *Planta Med*, 2016, 81(S1): S1-S381.
- [15] 李娟, 孔冉冉, 梅月菊, 等. 栽培远志 qRT-PCR 内参基因筛选与 P450s 基因的时空表达分析 [J]. 中草药, 2019, 50(24): 6084-6090.
- [16] Jin M L, Lee D Y, Um Y, *et al.* Isolation and characterization of an oxidosqualene cyclase gene encoding a β -amyrin synthase involved in *Polygala tenuifolia* Willd. saponin biosynthesis [J]. *Plant cell reports*, 2014, 33(3): 511-519.
- [17] Jin M L, Lee W M, Kim O T. Two Cycloartenol synthases for phytosterol biosynthesis in *Polygala tenuifolia* Willd [J]. *Int J Mol Sci*, 2017, 18(11): 2426.
- [18] Ho T T, Murthy H N, Park S Y. Methyl jasmonate induced oxidative stress and accumulation of secondary metabolites in plant cell and organ cultures [J]. *Int J Mol Sci*, 2020, 21(3): 716-720.
- [19] Sadeghnezhad E, Sharifi M, Zare-Maivan H, *et al.* Time-dependent behavior of phenylpropanoid pathway in response to methyl jasmonate in *Scrophularia striata* cell cultures [J]. *Plant Cell Rep*, 2020, 39(2): 227-243.
- [20] Rahimi S, Devi B S R, Khorolraghaa A, *et al.* Effect of salicylic acid and yeast extract on the accumulation of jasmonic acid and sesquiterpenoids in *Panax ginseng* adventitious roots [J]. *Russian J Plant Physiol*, 2014, 61(6): 811-817.
- [21] 王焕, 杨锦芬, 邓可, 等. 茉莉酸甲酯影响阳春砂

- 挥发性萜类代谢和基因转录 [J]. 世界科学技术—中医药现代化, 2014, 16(7): 1528-1536.
- [22] 李 洁, 胡本祥, 彭 亮, 等. 水杨酸和茉莉酸甲酯对远志愈伤组织生长和相关酶活性及化学成分的影响 [J]. 中草药, 2019, 50(12): 2976-2982.
- [23] 慧 芳, 刘秀岩, 李宗谕, 等. 转录组测序技术在药用植物研究中的应用 [J]. 中草药, 2019, 50(24): 6149-6155.
- [24] Sun G, Yang Y, Xie F, *et al.* Deep sequencing reveals transcriptome reprogramming of *Taxus × media* cells to the elicitation with methyl jasmonate [J]. *PLoS One*, 2013, 8(4): 1-13.
- [25] 任 昂. 茉莉酸甲酯诱导灵芝三萜生物合成的机制研究 [A] // 中国菌物学会. 第十届全国食用菌学术研讨会论文汇编 [C]. 北京: 中国菌物学会, 2014.
- [26] 张文娟, 曹小迎, 蒋继宏. 茉莉酸甲酯诱导大戟三萜类代谢的研究 [J]. 广西植物, 2015, 35(4): 590-596.
- [27] 滕红梅, 谢旭蓉, 王玲丽, 等. 外源物质对远志根中皂苷含量的影响 [J]. 南方农业学报, 2014, 45(12): 2143-2147.