

不同生长年限蒙古黄芪转录组学分析及三萜皂苷合成关键基因挖掘

王 祥¹, 王育朋², 安 佳¹, 王七龙¹, 井亚江¹, 黄建萍¹, 张 岗¹, 彭 亮¹, 颜永刚^{1*}

1. 陕西中医药大学 陕西省秦岭中草药应用开发工程技术研究中心/“秦药”研发重点实验室, 陕西 咸阳 712046

2. 陕西海天制药有限公司, 陕西 咸阳 712000

摘要:目的 探究蒙古黄芪 *Astragalus membranaceus* 三萜皂苷的合成基因及参与后加工修饰的 CYP、UGT 基因。方法 对两、三年生蒙古黄芪进行转录组测序和生物信息学分析。结果 Illumina HiSeq 平台测序共得到 42.22 Gb 数据, Trinity 软件拼接得到 369 790 条 Transcripts 和 336 068 条 Unigene, 与非冗余蛋白质序列 (non-redundant protein sequence database, Nr)、基因本体论 (gene ontology consortium, GO)、同源蛋白质簇 (cluster of orthologous groups of proteins, eggNOG/COG)、京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG)、Swiss-Prot、Pfam 等数据库比对注释, 结果显示蒙古黄芪与植物鹰嘴豆 *Cicer arietinum* 同源序列最多。KEGG 分析表明蒙古黄芪差异基因主要富集在萜类骨架生物合成、二萜类生物合成、氨基酰基-tRNA 生物合通路中, 进一步从萜类骨架生物合成通路中鉴别出 114 条差异表达 Unigene, 包含 15 种关键酶基因, 其中三年生蒙古黄芪 *hexPS* 表达均显著上调, *ACAT*、*HMGCS*、*HMGCR*、*MVK*、*MVAK2*、*MVD*、*IDI*、*FDPS*、*GGPS1*、*FNTB*、*STE24*、*STE14*、*FCLY*、*DHDDS* 有上调也有下调表达的 Unigene。进一步挖掘蒙古黄芪三萜皂苷生物合成后修饰基因, 发现 8 条 CYP 差异表达关键酶基因, 分别为 *acyP*、*PPID/CYPD*、*PPIH/CYPH*、*CYP26A*、*CYP51*、*CYP84A*、*CYP61A*、*cypD_E/CYP102A/CYP505*; 2 条 UGT 差异表达关键酶基因 *HUGT*、*SUGT1*。结论 两、三年生蒙古黄芪的皂苷合成基因和 CYP、UGT 基因表达存在显著差异, 这些基因共同参与黄芪皂苷的合成、加工与修饰。

关键词: 蒙古黄芪; 转录组; 差异表达基因; 黄芪皂苷; 生物合成途径

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2023)03-0915-11

DOI: 10.7501/j.issn.0253-2670.2023.03.026

Transcriptomic analysis of *Astragalus mongolicum* from different years and mining of key genes for synthesis of active components

WANG Xiang¹, WANG Yu-peng², AN Jia¹, WANG Qi-long¹, JIN Ya-jiang¹, HUANG Jian-ping¹, ZHANG Gang¹, PENG Liang¹, YAN Yong-gang¹

1. Shaanxi Engineering Research Center for Application and Development of Chinese Medicine in Qinling Mountains/Key Laboratory of “Qin Medicine” Research and Development, Shaanxi University of Chinese Medicine, Xianyang 712046, China

2. Shaanxi Haitian Pharmaceutical Co., Ltd., Xianyang 712000, China

Abstract: Objective To investigate the synthesis genes of triterpenoid saponins from *Astragalus membranaceus* var. *mongholicus* and the CYP and UGT genes involved in post-processing modification. **Methods** Transcriptome sequencing and bioinformatics analysis were performed on 2 and 3 years old Mongolian Astragalus. **Results** Illumina HiSeq platform was used for sequencing, and 42.22 Gb of data were obtained. After Trinity software stitching, 369 790 Transcripts and 336 068 Unigene Transcripts were obtained. Compared with non-redundant protein sequence database (NR), Gene Ontology Consortium (GO), Cluster of Orthologous Groups of proteins (eggNOG/COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot and PFAM databases, the results showed that *Astragalus mongholicus* and chickpea (*Cicer arietinum* Linn.) had the most homologous sequences. KEGG analysis showed that astragalus differential genes were mainly enriched in the terpene skeleton biosynthesis pathway, diterpene biosynthesis pathway and aminoacyl-tRNA biosynthesis pathway. Further, 114 differentially expressed Unigene including 15 key

收稿日期: 2022-08-09

基金项目: 内蒙古自治区科技重大专项 (2019ZD005); 国家中医药管理局全国名老中医药专家传承工作室建设项目 (国中医药人函[2019]41号); 陕西中医药大学“秦药”品质评价与资源开发学科创新团队项目 (2019-QN01)

作者简介: 王 祥, 男, 硕士研究生, 研究方向为中药鉴定与资源。E-mail: wangxiang1166@163.com

*通信作者: 颜永刚, 博士, 教授, 硕士研究生导师, 主要从事中药品种品质与资源开发及中药质量标准研究。E-mail: yunfeng828@163.com

enzyme genes were identified from the terpene skeleton biosynthesis pathway. The expression of *hexPS* was significantly up-regulated in three-year-old Mongolian astragalus. *ACAT*, *HMGCS*, *HMGCR*, *MVK*, *MVAK2*, *MVD*, *IDI*, *FDPS*, *GGPS1*, *FNTB*, *STE24*, *STE14*, *FCLY* and *DHDDS* were up-regulated and down-regulated in Unigene. After further mining of astragalus triterpenoid saponin biosynthetic modifier genes, eight key CYP genes were found to be differentially expressed, which were *acyP*, *PPID/ CYPD*, *PIIH/ CYPH*, *CYP26A*, *CYP51*, *CYP84A*, *CYP61A*, *cypD_E/ CYP102A/ CYP505*. Two UGT differentially expressed key enzyme genes *HUGT* and *SUGT1*. **Conclusion** There are significant differences in the expressions of saponin synthesis genes, CYP and UGT genes in 2- and 3-year-old astragalus. These genes are involved in the synthesis, processing and modification of astragalus.

Key words: *Astragalus membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao; transcriptome; differentially expressed genes; astragalus saponin; biosynthetic pathway

蒙古黄芪 *Astragalus membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao 的干燥根作为中药黄芪应用^[1], 最早见于《神农本草经》, 药用历史悠久, 为“补气之要药”具有补气升阳、固表止汗等功效^[2]。临床用于治疗脑缺血性疾病, 免疫系统疾病, 肾病及血糖、血压等疾病^[3-4]。黄芪为临床常用大宗药材, 且为药食两用药材, 市场需求量大, 野生资源供不应求, 故黄芪药材市场供给以人工种植为主^[5-6]。经市场调研及实地考察, 黄芪药材人工种植采收多以两年或三年生为主, 故本研究以两年生、三年生蒙古黄芪为研究对象。黄芪药材主要有有效成分之一为三萜皂苷类^[7-8], 目前已从黄芪植物中获得三萜皂苷类化合物 50 余种, 主要为黄芪皂苷 I~VIII, 异黄芪皂苷 I、II、IV, 乙酰基黄芪皂苷, 大豆皂苷等^[9-10]。黄芪生物体内化学成分的积累呈动态变化, 不同采收时间有效成分含量差异巨大, 即使在相同年限相同采收时间, 不同个体有效成分含量仍然存在较大差异^[11-12]。这些差异不仅受种质遗传的影响, 也受外界环境因子的影响^[13], 环境因子通过影响生物个体的基因表达调控, 进而影响植物次生代谢产物的合成与积累^[14-15]。故研究黄芪三萜皂苷的合成基因及其表达调控机制尤为重要, 有助于进一步研究如何提高黄芪药材的质量^[16]。

转录组测序技术可测定不同时期个体的功能基因及表达情况, 可将生物基因组遗传信息与功能代谢物建立连接, 有助于分析功能基因^[17]。目前转录组测序技术已广泛应用于细胞^[18]、组织^[19]、动物^[20]、植物^[21]等领域, 极大地拓宽了生物代谢产物的研究方法, 有利于研究药用植物有效成分的合成与积累机制, 尤其适用于像黄芪这样的非模式植物^[22]。目前少量关于黄芪转录组的研究, 主要集中在膜荚黄芪且研究部位为茎、叶、花和种子^[23], 缺少以蒙古黄芪研究对象, 且研究部位为药用部位(根)的转录组分析^[24]。故本研究以蒙古黄芪的根为研究对

象, 对不同生长年限蒙古黄芪转录组数据进行生物信息学分析^[25], 挖掘其有效成分合成基因及相关酶的表达情况, 为后续深入研究蒙古黄芪的植物育种、遗传变异、生长发育、有效成分合成提供理论依据。

1 材料与仪器

1.1 材料

蒙古黄芪样品来源于内蒙古通辽市科尔沁区庆和镇黄芪实验基地(121°60'21"E, 43°35'46"N; 海拔 141 m)。于 2022 年 5 月 1 日分别采集两年生(A)和三年生(B)蒙古黄芪新鲜植物根各 3 份(每份 6 株), 进行分组 A1、A2、A3 和 B1、B2、B3, 液氮速冻, 每组样品分别进行转录组测序和含量测定。植株经陕西中医药大学王继涛高级实验师鉴定为豆科植物蒙古黄芪 *A. membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao。

1.2 仪器

S3000 高效液相仪(华谱科仪(北京)科技有限公司); UM5800plus 蒸发光检测器(上海通微分析技术有限公司); K5500®分光光度计(凯奥公司, 北京); Bio-RAD CFX 96 荧光定量 PCR 仪(美国伯乐公司); Bioanalyzer 2100 系统(美国加利福尼亚州安捷伦科技)的 RNA Nano 6000 检测试剂盒。

2 方法

2.1 不同生长年限蒙古黄芪中三萜皂苷定量测定

每组蒙古黄芪样品中各取 50 g, 按《中国药典》2020 年版项下方法对蒙古黄芪样品分别进行制备和含量测定^[1], 流动相为 0.3% 甲酸水溶液(A)-乙腈(B)。柱温 30 °C; 进样体积 10 μL; 梯度洗脱(0~13 min, 82% A; 13~43 min, 82%~70% A; 43~60 min, 70%~45% A; 60~70 min, 45%~30% A)。ELSD 检测器条件: 漂移管温度: 65 °C, 气体体积流量 2.5 L/min, 信号增益 5。测定黄芪中 6 种主要皂苷(黄芪甲苷、黄芪皂苷 I、黄芪皂苷 II、黄芪皂苷 III、异黄芪皂苷 I、异黄芪皂苷 II)成分的

含量^[26]。

2.2 RNA提取、文库构建

对蒙古黄芪各组样品总RNA分别进行提取后,用Nanodrop检测RNA的纯度、浓度、核酸吸收峰是否正常;用Agilent 2100检测样品中RNA的完整性。样品检测合格后用带Oligo(dT)的磁珠富集mRNA,然后加入Fragmentation Buffer将mRNA随机打断,以之为模板用六碱基随机引物(random hexamers)合成第一条cDNA链,再加入缓冲液、dNTPs、RNase H和DNA polymerase I合成第2条cDNA链,利用AMPure XP beads纯化cDNA,并进行末端修复、加A尾并连接接头,并进行片段大小选择,再经PCR富集后得到cDNA文库。经Qubit 2.0进行初步定量后,用Agilent 2100对文库的insert size进行检测,最后用Q-PCR方法对文库的有效浓度进行准确定量(文库有效浓度>2 nmol)。

2.3 转录组数据的测序与拼接组装

库检合格后,不同文库按照目标下机数据量进行蓄集。用Illumina HiSeq进行测序,对上机测序得到的原始序列(raw reads)进行过滤,去掉含接头的reads和低质量的reads,得到高质量的序列(clean reads)。采用Trinity软件对clean reads进行拼接得到转录本序列,筛选其中必要的非冗余序列(Unigene)用于后续分析。

2.4 Unigene功能注释和分类

用转录本蛋白编码潜能预测工具(CPC2)查找含编码潜能的Unigene,然后用BLAST软件将发掘可编码的Unigene与非冗余蛋白质序列数据库(non-redundant protein sequence database, Nr)、基因本体论(gene ontology consortium, GO)、同源蛋白质簇数据库(cluster of orthologous groups of proteins, eggNOG/COG)、真核生物蛋白相邻类的聚簇(clusters of orthologous group for eukaryotic complete, KOG)、京都基因与基因组百科全书数据库(Kyoto encyclopedia of genes and genomes, KEGG)、Swiss-Prot、Pfam等数据库进行序列相似性对比,获得基因功能注释信息。用FPKM(fragments per kilobase of transcript per million fragments mapped)算法对样品中的mapped reads数目和转录本长度进行归一化,获得各个样本的表达量值。用DESeq(R包)对各组样品进行差异表达分析。利用Benjamini-Hochberg校正方法对原有假设检验的显著性P值(P-value)进行校正得到q

值,以FDR为差异表达基因(differentially expressed genes, DEGs)筛选的关键指标,对转录本进行KEGG代谢通路分析及差异基因表达分析。

3 结果与分析

3.1 不同生长年限蒙古黄芪三萜皂苷含量差异

对不同生长年限蒙古黄芪,按组分别制样,于“2.1”项下色谱条件进行含量测定。对不同生长年限蒙古黄芪中三萜皂苷含量进行对比分析,结果表明三年生蒙古黄芪各种皂苷成分含量均有不同程度提高,其中黄芪甲苷、黄芪皂苷I、黄芪皂苷II、黄芪皂苷III、异黄芪皂苷I、异黄芪皂苷II的平均质量分数较两年生分别上升42.50%、86.05%、95.82%、100.00%、27.11%、37.81%。两年生蒙古黄芪总皂苷质量分数(6种黄芪三萜皂苷含量之和)为16.8~27.1 mg/g,平均质量分数为22.57 mg/g;三年生蒙古黄芪总皂苷质量分数为28.7~39.0 mg/g,平均质量分数为35.57 mg/g;蒙古黄芪总皂苷平均质量分数三年生较两年生高出13.0 mg/g,为两年生的1.58倍。进一步对不同生长年限蒙古黄芪样品中三萜皂苷含量进行统计学分析验证,结果表明不同生长年限蒙古黄芪中三萜皂苷含量存在显著差异($P < 0.05$),三年生蒙古黄芪三萜皂苷含量显著升高。

3.2 转录组测序与序列拼接

对不同生长年限蒙古黄芪样本进行转录组测序,每个样本重复3次,共生成282 321 486个clean reads,包含有42.22 Gb的clean base。Q20的碱基百分比分布在98.57%~97.71%,Q30的碱基百分比分布在94.88%~95.29%,GC碱基量为46.10%~49.86%(表1)。结果表明蒙古黄芪转录组测序数据质量较高,用Trinity软件将高质量数据进行De novo组装,共得到369 790条transcripts和336 068条Unigene, N50长度分别为1293、904 bp, N90长度分别为275、261 bp,组装完整性较好。长度分布在200~500 bp的Unigene最多,有227 595条,占Unigene总数的67.72%(表2)。将Trinity拼接得到的转录组作为参考序列,将各样品的clean reads与参考序列进行比对,得到各个样品的mapped reads,其样品匹配度均在76.0%以上,表明测序数据组装拼接效果较好。

3.3 蒙古黄芪 Unigenes 的功能注释

将得到的336 068条Unigene与各大核苷酸和蛋白质数据库进行比对,发现Unigene在NR数据库和PFAM数据库中基因匹配率最高分别为97.18%、

表 1 各样品转录组测序基础数据

Table 1 RNA-seq data of samples

样本	总数量/条	高质数量/条	GC/%	Q20/%	Q30/%	匹配量/条	匹配度/%
两年生	45 314 992	45 314 978	46.10	98.67	95.13	41 325 741	79.15
两年生	51 107 200	51 107 182	48.43	98.66	95.13	48 933 509	79.14
两年生	44 917 600	44 917 590	49.41	98.71	95.29	43 240 988	78.00
三年生	41 558 952	41 558 946	49.86	98.57	94.88	35 704 002	76.92
三年生	49 987 650	49 987 630	49.84	98.66	95.16	42 819 783	76.88
三年生	49 435 172	49 435 160	49.86	98.64	95.11	42 470 996	76.91

表 2 转录本拼接情况统计

Table 2 Statistics of transcripts and Unigenes after sequence assembly

指标	200~500 bp/条	500~1000 bp/条	1~2000 bp/条	>2000 bp/条	总数/条	平均长度/条	N50/bp	N90/bp
转录本	227 606	62 273	50 872	29 039	369 790	738.8	1293	275
Unigene	227 595	57 424	34 283	16 766	336 068	617.2	904	261
编码基因数	3216	17 399	29 038	16 009	65 662	/	/	/

85.94%。其次是 Swiss-Prot 数据库和 GO 数据库，在 Swiss-Prot 数据库、GO 数据库、PFAM 数据和 NR 数据库被同时注释的 Unigene 为 16 916 条。在 5 个数据库被共同注释到的 Unigene 有 21 194 条，占总 Unigene 的 6.31% (表 3 和图 1)。

统计在 NR 数据库中比对到的注释信息，共得到 61 285 条 Unigene，其中蒙古黄芪与双子叶植物纲 (Dicotyledoneae) 相似序列匹配度最高为 60 059

表 3 序列注释统计

Table 3 Statistics of sequence annotation

数据库	数量/条	比例/%
NR	61 285	97.180
Swiss-Prot	39 827	63.153
GO	39 166	62.104
KEGG	23 630	37.469
PFAM	54 194	85.936
eggNOG/COG	36 599	58.034

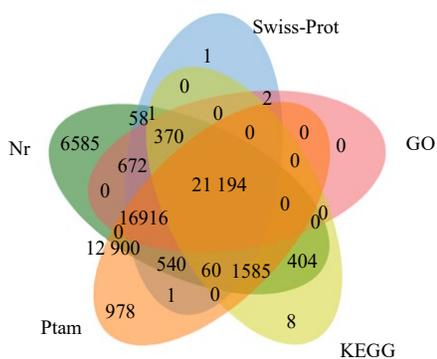


图 1 序列注释韦恩图

Fig. 1 Venn diagram of transcript annotation in different databases

条 (98%); 其次为豆目 (Fabales Bromhead) 39 222 条 (64%), 豆科 (Fabaceae) 41 061 条 (67%), 鹰嘴豆属 *Cicer* Linn. 18 385 条 (30%)、苜蓿属 *Medicago* L. 12 870 条 (21%)。进行物种信息对比发现，蒙古黄芪与植物鹰嘴豆 *Cicer arietinum* Linn. 相似序列匹配度最高为 21 450 条 (35%)，其次是蒺藜状苜蓿 *Medicago truncatula* Gaertn. 15 321 条 (25%)、欧洲栓皮栎 *Quercus suber* L. 4290 条 (7%)、百脉根 *Lotus japonicus* Linn. 4290 条 (7%)、木豆 *Cajanus cajan* (Linn.) Millsp. 3677 条 (6%)、稻 *Oryza sativa* L. 3064 条 (5%)。这些注释信息可为蒙古黄芪基因相关研究提供参考 (图 2)。

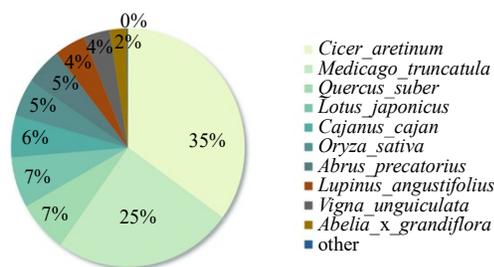


图 2 同源序列物种比对

Fig. 2 Species distribution annotated in NR database gene_ontology_blast

3.4 GO、eggNOG 与 KEGG 注释分析

通过 GO 基因本体论数据库注释，进一步将蒙古黄芪基因产物功能和概念阐释^[27]。将蒙古黄芪的 Unigenes 比对到 GO 数据库中，发现共有 39 166 条基因序列得到注释。总体分为 3 大类：分子功能 (molecular function, 81 027 条, 30.20%)、细胞组

分 (cellular component, 88 136 条, 32.85%) 和生物过程 (biological process, 99 156 条, 36.95%)。注释到具有分子功能的 Unigenes 共分为 15 种, 其中结合蛋白 (binding, 17 842 条) 和催化活性 (catalytic activity, 16 395 条) 比对到的基因数目最多。注释到细胞组分的 20 个组别中, 细胞 (cell,

24 644 条)、细胞器 (organelle, 19 296 条) 和细胞部分 (cell part, 24 578 条) 所占的比例最高。在比对到生物过程的 26 个分组中新陈代谢过程 (metabolic process, 16 150 条)、细胞过程 (cellular process, 20 126 条) 和生物调控 (biological regulation, 7978 条) 基因数目最多 (图 3)。

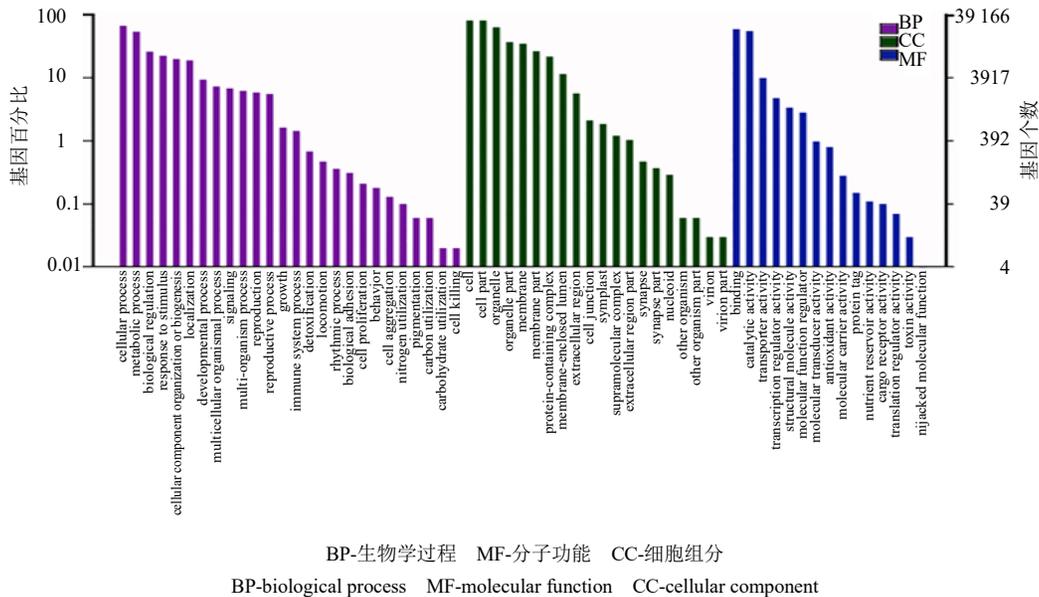


图 3 转录组的 GO 注释统计
Fig. 3 GO annotation of the transcriptome

将蒙古黄芪的 Unigenes 在 eggNOG/GO 数据库中进行信息比对, 共注释到了 36 559 条基因, 分为 24 类 (图 4), 其中已知功能且条目数最多的有翻译后修饰/蛋白质周转 (posttranslational modification, protein turnover) 2982 条、碳水化合物的运输和新陈代谢 (carbohydrate transport and metabolism) 2553 条、信号转导机制 (signal transduction mechanisms) 2442 条。

KEGG 注释分析发现蒙古黄芪共有 23 630 个 Unigene 成功比对到 84 条通路中, 主要在代谢 (metabolism, 6594 条 67.42%)、遗传信息处理 (genetic information processing, 2 195 条, 22.44%)、细胞过程 (cellular processes, 690 条, 7.06%)、有机系统 (organismal systems, 179 条, 1.83%)、环境信息处理 (environmental information processing, 122 条, 1.25%) 5 个大的分类中 (图 5)。富集到代谢过程的 Unigene 最多, 其中有 20 条代谢途径共 2911 条 Unigene 与植物的次生代谢产物合成相关; 聚集在氨基酰基-tRNA 生物合成 (aminoacyl-tRNA biosynthesis, 228 条) 的 DEGs 数目最多, 后富集

较多 Unigene 的 5 条代谢途径依次是苯丙素生物合成 (phenylpropanoid biosynthesis, 226 条)、苯丙氨酸、酪氨酸和色氨酸生物合成 (phenylalanine, tyrosine and tryptophan biosynthesis, 152 条)、萜类骨架生物合成 (terpenoid backbone biosynthesis, 122 条)、不饱和脂肪酸的生物合成 (biosynthesis of unsaturated fatty acids, 114 条)、泛酸盐和辅酶 A 生物合成 (pantothenate and CoA biosynthesis, 107 条) (表 4)。

3.5 不同生长年限蒙古黄芪差异表达基因分析

差异丰度基因筛选, 以差异倍数 |FoldChange| ≥ 2 且 FDR (false discovery rate) < 0.01 作为筛选标准, 判断转录本上调和下调表达情况, 筛选到蒙古黄芪转录本差异表达基因共 48 261 条, 其中三年生蒙古黄芪上调表达 30 138 条, 下调表达 18 123 条。进一步对差异表达基因进行 KEGG 代谢通路富集分析 (图 6), 结果共有 10 061 条 DEGs 被富集到 84 条 pathway 上, 在富集较显著的 15 条通路中, 差异代谢基因主要富集在萜类骨架生物合成、二萜类生物合成 (diterpenoid biosynthesis)、氨基酰基-tRNA 生物

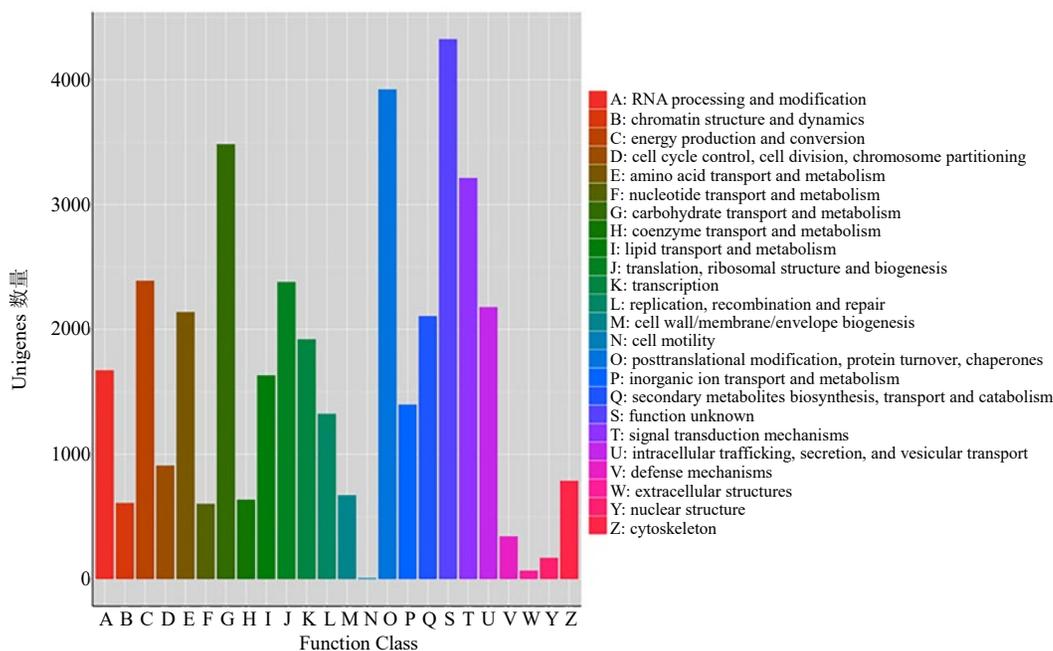


图 4 基因 eggNOG/COG 注释分类统计图

Fig. 4 EggNOG/COG annotation classification statistical map of genes

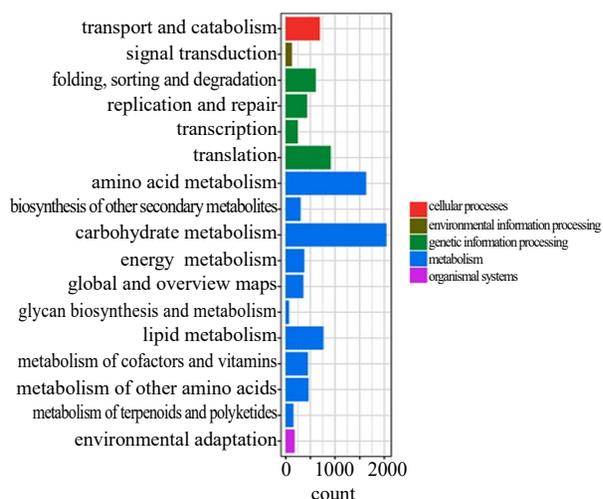


图 5 基因 KEGG 二级分类统计

图 5 KEGG secondary classification statistics of genes

合成中。三年生蒙古黄芪在氨基酰基-tRNA 生物合成途径富集到 228 条 Unigenes 的背景下，富集到差异代谢基因有 183 条，99 条显著上调；在萜类骨架生物合成途径富集到 122 条 Unigenes 的背景下，富集到 100 条差异代谢基因，其中有 53 条 DEGs 显著上调；在二萜类生物合成途径富集到的 14 条 Unigenes 的背景下，在富集到 13 条差异代谢基因，全部表达下调。

3.6 三萜皂苷骨架生物合成通路差异表达基因分析

不同生长年限蒙古黄芪的萜类骨架生物合成和二萜类生物合成通路中共筛选到 15 种差异表达关键

表 4 差异次生代谢物合成途径

Table 4 Differential secondary metabolite synthesis pathways

差异代谢途径名称	代谢途径号	Unigenes/条
氨基酰基-tRNA 生物合成	ko00970	228
苯丙素生物合成	ko00940	226
苯丙氨酸、酪氨酸和色氨酸生物合成	ko00400	152
萜类骨架生物合成	ko00900	122
不饱和脂肪酸的生物合成	ko01040	114
泛酸盐和辅酶 A 生物合成	ko00770	107
缬氨酸、亮氨酸和异亮氨酸生物合成	ko00290	106
脂肪酸生物合成	ko00061	100
异喹啉生物碱的生物合成	ko00950	98
赖氨酸生物合成	ko00300	84
泛醌和其他萜类醌生物合成	ko00130	76
叶酸生物合成	ko00790	75
类黄酮生物合成	ko00941	44
糖鞘脂生物合成 globo 系列	ko00603	30
内酰胺生物合成	ko00261	26
其他类 O-聚糖生物合成	ko00514	25
二苯乙烯、二芳基庚烷类和姜辣素生物合成	ko00945	20
角质、木质和蜡质的生物合成	ko00073	17
二萜类生物合成	ko00904	14
倍半萜和三萜类生物合成	ko00909	15

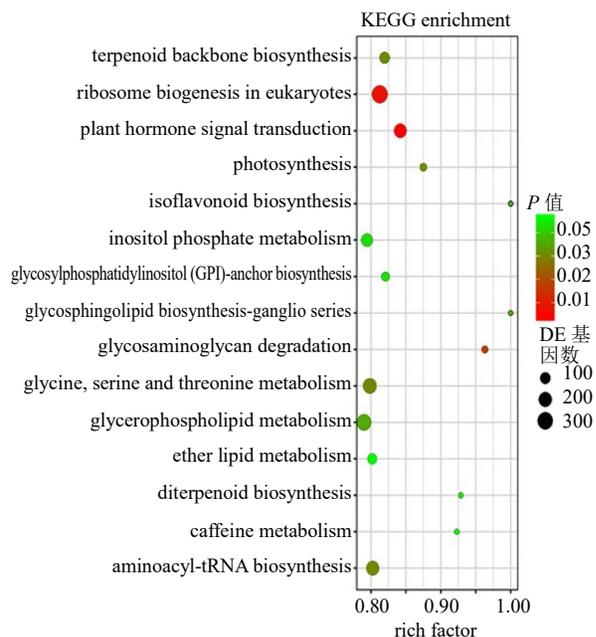


图6 差异基因 KEGG 途径富集

Fig. 6 KEGG pathways enriched for differentially expressed genes

酶基因，分别为六异戊二烯二磷酸合酶（hexaprenyl-diphosphate synthase, *hexPS*, 2条 Unigene）、酰基辅酶 A-胆固醇酰基转移酶（acetyl-CoA C-acetyltransferase, *ACAT*, 10条 Unigene）、羟甲基戊二酰辅酶 A 合酶（hydroxymethylglutaryl-CoA synthase, *HMGCS*, 5条 Unigene）、羟甲基戊二酰辅酶 A 还原酶（hydroxymethylglutaryl-CoA reductase, *HMGCR*, 13条 Unigene）、甲羟戊酸激酶（mevalonate kinase, *MVK*, 4条 Unigene）、磷酸甲羟戊酸激酶（phosphomevalonate kinase, *MVAK2*, 4条 Unigene）、二磷酸甲羟戊酸脱羧酶（diphosphomevalonate decarboxylase, *MVD*, 4条 Unigene）、异戊烯二磷酸 Delta-异构酶（isopentenyl-diphosphate delta-isomerase, *IDI*, 3条 Unigene）、过二磷酸合酶（farnesyl diphosphate synthase, *FDPS*, 11条 Unigene）、牻牛儿基牻牛儿基焦磷酸合酶（geranylgeranyl diphosphate synthase, *GGPS1*, 6条 Unigene）、蛋白香叶烯基转移酶（protein farnesyltransferase subunit beta, *FNTB*, 7条）、内肽酶（endopeptidase, *STE24*, 4条 Unigene）、蛋白质-S-异戊二烯半胱氨酸 O-甲基转移酶（protein-S-isoprenylcysteine O-methyltransferase, *STE14*, 2条 Unigene）、异戊二烯半胱氨

酸氧化酶（prenylcysteine oxidase/ farnesylcysteine lyase, *FCLY*, 6条）、二转聚戊二烯基二磷酸合酶（ditranspolycis-polyprenyl diphosphate synthase, *DHDDS*, 9条）。其中 *hexPS* 均上调表达, *ACAT*、*HMGCS*、*HMGCR*、*MVK*、*MVAK2*、*MVD*、*IDI*、*FDPS*、*GGPS1*、*FNTB*、*STE24*、*STE14*、*FCLY*、*DHDDS* 既有上调又有下调表达的 Unigene, 其差异表达基因聚类热图（图7）。

3.7 三萜皂苷生物合成后修饰基因 *CYP*、*UGT* 挖掘及表达分析

不同生长年限蒙古黄芪皂苷合成通路中共筛选到8种 *CYP* 和2种 *UGT* 差异表达关键酶基因。分别为参与丙酮酸代谢和氨基苯甲酸酯降解的酰基磷酸酶基因（acylphosphatase, *acyP*）；参与苯丙烷生物合成的阿魏酸-5-羟化酶基因（ferulate-5-hydroxylase, *CYP84A/F5H*）；参与类固醇生物合成的甾醇 14 α -去甲基化酶基因（sterol 14 α -demethylase, *CYP51*）和甾醇 22-去饱和酶基因（*ERG5/CYP61A*, sterol 22-desaturase）；参与色氨酸代谢、氨基苯甲酸酯降解的细胞色素 P450/NADPH-细胞色素 P450 还原酶基因（cytochrome P450/NADPH-cytochrome P450 reductase, *cypD_E/CYP102A/CYP505*）；参与坏死性凋亡和细胞衰老通路的肽基-脯氨酰异构酶基因（peptidyl-prolyl isomerase D, *PPID/CYPD*）；剪接体酶肽基脯氨酰异构酶 H [peptidyl-prolyl isomerase H, *PPIH/CYPH* (cyclophilin H)]；参与视黄醇代谢的细胞色素 P450 家族 26 亚家族 A 酶基因（cytochrome P450 family 26 subfamily A, *CYP26A*）。2种 *UGT* 关键酶基因分别是：糖蛋白葡萄糖基转移酶基因（UDP-glucose:glycoprotein glucosyltransferase, *HUGT*）；*SKP1* 的 G2 等位基因抑制因子（suppressor of G2 allele of *SKP1*, *SUGT1/SGT1*）。筛选的10种三萜皂苷生物合成后修饰基因, 在三年生蒙古黄芪中的表达均上升（表5）。

4 讨论

本研究用高通量测序技术, 对不同生长年限蒙古黄芪新鲜植物根进行转录组测序分析, 质控和拼接后得到了 336 068 条 Unigenes, 在 5 个数据库（NR、Swiss-Prot、GO、KEGG、PFAM、）被共同注释到 21 194 条 Unigene, 占总 Unigene 的 6.31%, 极大地丰富了蒙古黄芪的转录组信息和生物学信息, 扩充了黄芪的数据库信息。进一步对其基因进行分析, 发现在 KEGG 数据库中有 23 630 个 Unigenes

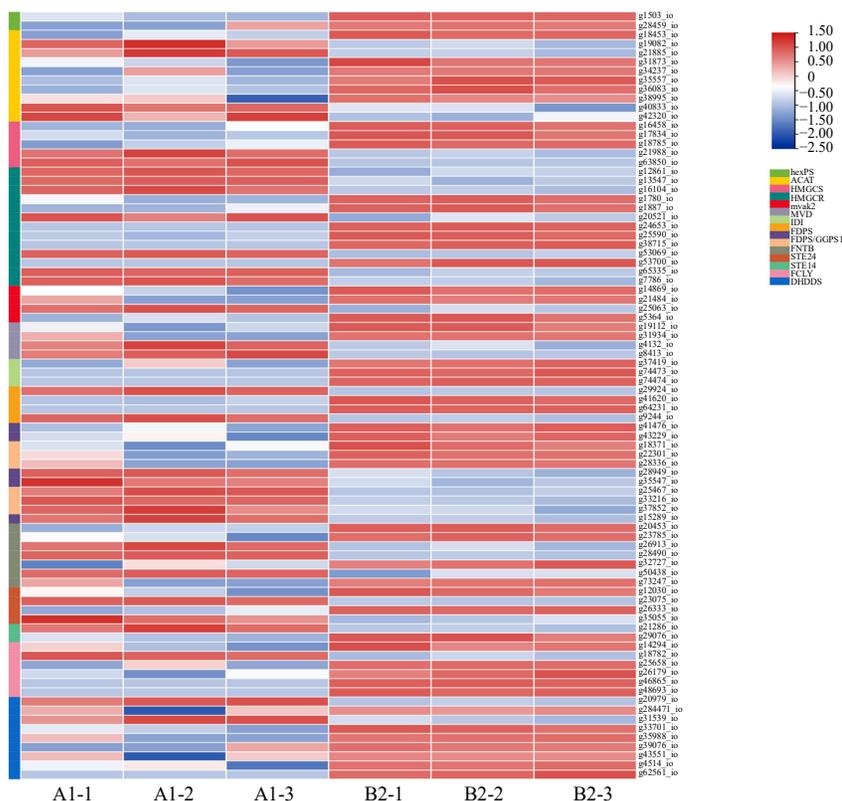


图 7 差异代谢基因聚类热图

Fig. 7 Clustering heat map of differential metabolic genes

表 5 黄芪皂苷生物合成的 CYP、UGT 基因

Table 5 Astragalus saponin biosynthesis of CYP, UGT genes

基因缩写	EC 编号	Unigene 编号	B vs A 上/下调
acyP	3.6.1.7	g105592_i0	上调
PPID, CYPD	5.2.1.8	g36068_i0	上调
PPIH, CYPH	5.2.1.8	g80318_i0	上调
CYP26A	1.14.14.-	g90817_i0	上调
CYP51	1.14.14.154/ 1.14.15.36	g20181_i0	上调
CYP84A	1.14.-	g14282_i0; g18644_i0; g33209_i0; g46321_i0	上调
CYP61A	1.14.19.41	g15568_i0; g20329_i0; g39339_i0; g16273_i0; g23980_i0;	上调
cypD_E/ CYP102A/ CYP505	1.14.14.1/ 1.6.2.4	g34946_i0; g35649_i0; g4314_i0; g44357_i0; g50033_i0; g9403_i0;	上调
HUGT	2.4.1.-	g24183_i0; g24970_i0; g44101_i0;	上调
SUGT1	/	g32060_i0; g35103_i0;	上调

成功比对到 84 条通路中，两、三年生蒙古黄芪至少有 10 061 条差异表达基因，主要表现在氨基酰-tRNA 生物合成、萜类骨架生物合成、二萜类生物合成途径中。

蒙古黄芪的 Unigene 在 Nr 数据库获得注释比例最高 (98.17%)，注释到的同源信息主要为双子叶植物纲(98%)，其次为豆目(64%)、蝶形花科(67%)、鹰嘴豆属 (30%)、苜蓿属 (21%)，这可为蒙古黄

芪的物种起源、植物进化、遗传变异、物种分类等提供重要参考。进一步进行物种序列相似性比对发现，蒙古黄芪与植物鹰嘴豆 (35%)、蒺藜状苜蓿 (25%)、欧洲栓皮栎 (7%)、百脉根 (7%)、木豆 (6%) 的相似序列匹配度较高，这些物种的基因信息可为蒙古黄芪的基因功能研究提供参考。对蒙古黄芪已知序列但未知结构功能的蛋白质，可以从这些植物中寻找它的相似序列，从而推测出相似列表达产

生的蛋白质结构和功能^[28-29]。

进一步对两、三年生蒙古黄芪差异表达基因进行 KEGG 代谢通路分析,发现差异代谢基因主要富集在萜类骨架生物合成和二萜类生物合成中,其中三年生蒙古黄芪 *hexPS* 酶基因均上调表达, *ACAT*、*HMGCS*、*HMGCR*、*MVK*、*MVAK2*、*MVD*、*IDI*、*FDPS*、*GGPS1*、*FNTB*、*STE24*、*STE14*、*FCLY*、*DHDDS* 既有上调表达又有下调表达的 *Unigene*。其中 *hexPS* 作为萜类合成的关键基因,主要参与萜类基础骨架合成,以满足黄芪三萜皂苷前体物质的供应,对黄芪三萜皂苷的合成起着支撑性作用^[30]。三年生蒙古黄芪的 *hexPS* 基因表达明显上调,可能与其三萜皂苷合成量大,对萜类基础骨架需求量大有关,这与本研究对两年生、三年生蒙古三萜皂苷含量验证结果一致^[31]。*ACAT*、*FNTB*、*STE24*、*STE14*、*FCLY*、*DHDDS* 主要参与萜骨架的合成与修饰,其表达既有上调又有下调,但上调表达数明显大于下调表达数,这可能与萜骨架合成与修饰过程中一些负反馈调节基因的参与有关,当植物合成的某些有机物积累到一定量时,负反馈调节机制会降低其正向合成基因的表达^[32]。*HMGCR* 是 *HMGCS* 经甲羟戊酸途径合成 C5 类异戊二烯中间体的负反馈调节基因,二者双向调节异戊二烯中间体的合成速率,故当 C5 类异戊二烯中间体合成达到一定量时,*HMGCS* 基因会受到 *HMGCR* 的反馈调节而出现短暂的表达减弱。*MVK* 将高能供体分子磷酸基团转移到特定靶分子“激活”或“能化”甲羟戊酸底物分子,然后 *MVAK2* 将甲羟戊酸磷酸化,*MVD* 对二磷酸甲羟戊酸进行脱羧反应,生成的异戊烯二磷酸再由 *IDI*、*FDPS*、*GGPS1*、*FNTB* 等酶基因经一系列反应生成三萜皂苷前体物质异戊烯焦磷酸(isopentenyl pyrophosphate, IPP)^[33]。黄芪三萜皂苷的生物合成首先经过萜类骨架生物合成途径生成萜类前体物质,然后经过甲羟戊酸(mevalonate)途径或 2-C-甲基-D-赤藓醇-4-磷酸(2-C-methyl-D-erythritol-4-phosphate)途径磷酸化,生成异戊烯焦磷酸^[34],IPP 再由法尼基焦磷酸合酶、鲨烯合酶、鲨烯环氧酶、环阿屯醇合酶经一系列反应生成三萜皂苷的前体环阿屯醇(cycloartenol)^[35-36],环阿屯醇再经一系列的氧化、还原、糖基化等后修饰,最终生成黄芪皂苷类化合物。

关于黄芪三萜皂苷生物合成的后修基因目前少有报道,因此,本研究进一步从蒙古黄芪转录组测

序结果中挖掘参与黄芪三萜皂苷生物合成后修饰的 CYP、UGT 酶基因。从不同生长年限蒙古黄芪皂苷合成通路中共筛选到 8 种 CYP 差异表达关键酶基因,分别为 *acyP*、*PPID/CYPD*、*PPIH/CYPH*、*CYP26A*、*CYP51*、*CYP84A*、*CYP61A*、*cypD_E/CYP102A/CYP505*; 2 种 UGT 差异表达关键酶基因,分别是 HUGT、SUGT1。共筛选的 10 种 CYP、UGT 酶基因在三年生蒙古黄芪的表达均明显增强,本研究对蒙古黄芪三萜皂苷含量进行测定,三年生蒙古黄芪三萜皂苷含量明显上升,这可能与黄芪三萜皂苷生物合成的后修饰酶基因 CYP、UGT 的表达调控有关^[37]。其中 *CYP84A* 可通过苯丙烷生物合成途径合成木质素单体^[38];*CYP61A* 主要通过类固醇生物合成角鲨烯 2,3-环氧化物^[39];*CYP51* 主要通过胆固醇生物合成、麦角钙化醇生物合成、植物甾醇生物合成途径合成角鲨烯 2,3-环氧化物^[40];*HUGT* 主要参与内质网中的蛋白质加工修饰^[41];这些 CYP、UGT 酶基因均可作为蒙古黄芪三萜皂苷合成后加工与修饰的候选基因。

本研究通过对蒙古黄芪转录组测序和生物信息学分析,极大地丰富了蒙古黄芪的生物学信息,挖掘了参与黄芪三萜皂苷的合成通路及关键基因,探讨了关于黄芪三萜皂苷生物合成的后修饰酶基因 CYP、UGT 及其表达规律,为后续深入研究蒙古黄芪的植物育种、遗传变异、生长发育、有效成分合成提供理论基础。

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. 2020: 315.
- [2] 顾志荣,葛斌,许爱霞,等. 基于本草考证的黄芪功效主治及用药禁忌挖掘 [J]. 中成药, 2018, 40(11): 2524-2530.
- [3] 马艳春,胡建辉,吴文轩,等. 黄芪化学成分及药理作用研究进展 [J]. 中医药学报, 2022, 50(4): 92-95.
- [4] 徐世一,刘秀波,陆佳欣,等. 黄芪活性成分抗肿瘤作用机制的研究进展 [J]. 中草药, 2022, 53(23): 7613-7623.
- [5] 孙淑英,陈贵林. 内蒙古黄芪产业化现状、问题及对策建议 [J]. 分子植物育种, 2018, 16(15): 5126-5133.
- [6] 李洪,王彧超,王瑞军,等. 黄芪种质资源研究与利用 [J]. 安徽农学通报, 2020, 26(20): 34.
- [7] 向璐,张巧艳,赵琦明,等. 黄芪-当归化学成分、药理作用及临床应用的研究进展 [J]. 中草药, 2022, 53(7): 2196-2213.

- [8] 马园园, 王静, 罗琼, 等. 黄芪总皂苷药理作用研究进展 [J]. 辽宁中医药大学学报, 2020, 22(7): 153-157.
- [9] Xiao L M, Cao P H, Luo Z H, *et al.* Cycloartane-type triterpenoids from the root of *Astragalus membranaceus* var. *mongholicus* [J]. *J Asian Nat Prod Res*, 2020, 22(10): 905-913.
- [10] Wang Y M, Liu L, Ma Y K, *et al.* Chemical Discrimination of *Astragalus mongholicus* and *Astragalus membranaceus* based on metabolomics using UHPLC-ESI-Q-TOF-MS/MS approach [J]. *Molecules*. 2019, 24(22): 4064.
- [11] 王玲玲, 杨路存, 熊丰, 等. 不同栽培密度和采收期对蒙古黄芪生长发育和产量的影响 [J]. 分子植物育种, 2019, 17(23): 7962-7968.
- [12] 辛博, 谢景, 王文全. 不同生长年限黄芪药材中总多糖和总黄酮含量的测定 [J]. 中医药信息, 2015, 32(5): 31-34.
- [13] 孙海, 金桥, 吴虎平, 等. 环境因素对黄芪产量和品质的影响 [J]. 特产研究, 2019, 41(3): 118-122.
- [14] 武悦, 陈阳, 王星哲, 等. 扁茎黄芪转录组测序及生物信息学分析 [J]. 华北农学报, 2022, 37(1): 42-49.
- [15] 王强, 胡金鹏, 潘雅清, 等. 蒙古岩黄芪全长转录组测序及微卫星位点特征分析 [J]. 分子植物育种, 2021, 19(24): 8109-8120.
- [16] Park Y J, Thwe A A, Li X H, *et al.* Triterpene and flavonoid biosynthesis and metabolic profiling of hairy roots, adventitious roots, and seedling roots of *Astragalus membranaceus* [J]. *J Agric Food Chem*, 2015, 63(40): 8862-8869.
- [17] 薛守宇, 朱涛, 李冰冰, 等. 转录组和代谢组联合分析在植物中的应用研究 [J]. 山西农业大学学报: 自然科学版, 2022, 42(3): 1-13.
- [18] Gaydosik A M, Tabib T, Domsic R, *et al.* Single-cell transcriptome analysis identifies skin-specific T-cell responses in systemic sclerosis [J]. *Ann Rheum Dis*, 2021, 80(11): 1453-1460.
- [19] Möller-Levet C S, Laing E E, Archer S N, *et al.* Author correction: Diurnal and circadian rhythmicity of the human blood transcriptome overlaps with organ- and tissue-specific expression of a non-human primate [J]. *BMC Biol*, 2022, 20(1): 98.
- [20] 孙娟, 蒋晓娟, 王亚东, 等. 基于高通量转录组测序探讨参苓白术散缓解小鼠溃疡性结肠炎的作用机制 [J]. 中国中药杂志, 2022, 47(22): 6155-6163.
- [21] 寇佩雯, 刘长乐, 许珂珂, 等. 不同年份防风转录组学分析及有效成分生物合成关键基因挖掘 [J]. 中国中药杂志, 2022, 47(17): 4609-4617.
- [22] 王瑞娴, 李川. 全长转录组测序技术在非模式植物转录组学研究中的应用 [J]. 分子植物育种, 2019, 17(2): 502-508.
- [23] 于倩, 孔德民, 王宇恒, 等. 基于转录组的黄芪种子萌发过程的生化分析 [A] // 2016年全国植物生物学大会摘要集 [C]. 武汉: 全国植物生物学协会, 2016: 160.
- [24] 贺润丽, 杨晓琳, 段迎, 等. 转录组测序分析膜荚黄芪紫茎花青素生物合成的关键基因 [J]. 中药材, 2021, 44(1): 47-53.
- [25] Wu X T, Li X T, Wang W, *et al.* Integrated metabolomics and transcriptomics study of traditional herb *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao reveals global metabolic profile and novel phytochemical ingredients [J]. *BMC Genomics*, 2020, 21(Suppl 10): 697.
- [26] 王亚丽, 田曼, 李江, 等. HPLC-DAD-ELSD 法同时测定黄芪中 10 个成分的含量 [J]. 中南药学, 2018, 16(9): 1268-1271.
- [27] Chen L, Zhang Y H, Wang S P, *et al.* Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways [J]. *PLoS One*, 2017, 12(9): e0184129.
- [28] Lan Y M, Rosen G, Hershberg R. Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains [J]. *Microbiome*, 2016, 4(1): 18.
- [29] Ietswaart R, Gyori B M, Bachman J A, *et al.* GeneWalk identifies relevant gene functions for a biological context using network representation learning [J]. *Genome Biol*, 2021, 22(1): 55.
- [30] Kuwahara Y, Nakajima D, Shinpo S, *et al.* Identification of potential genes involved in triterpenoid saponins biosynthesis in *Gleditsia sinensis* by transcriptome and metabolome analyses [J]. *J Nat Med*, 2019, 73(2): 369-380.
- [31] Li Y T, Li J X, Diao M X, *et al.* Characterization of a group of UDP-glycosyltransferases involved in the biosynthesis of triterpenoid saponins of *Panax notoginseng* [J]. *ACS Synth Biol*, 2022, 11(2): 770-779.
- [32] 张甜甜, 郑炳松, 袁虎威, 等. 植物挥发性有机物合成与代谢途径及其释放与感知调控机制的研究进展 [J/OL]. 天然产物研究与开发: 1-20[2022-10-29]. <http://www.trcw.ac.cn/fileup/1001-6880/PDF/2022101102.pdf>
- [33] 钱程程, 赵历强, 杨亚焱, 等. 三叶木通转录组分析及三萜皂苷生物合成途径关键酶基因的发掘 [J]. 植物科学学报, 2022, 40(3): 378-389.
- [34] Yao L, Lu J, Wang J, *et al.* Advances in biosynthesis of triterpenoid saponins in medicinal plants [J]. *Chin J Nat Med*, 2020, 18(6): 417-424.

- [35] Xia P G, Zheng Y J, Liang Z S. *Panax notoginseng* structure and location studies on Key enzymes in saponins biosynthesis [J]. *Int J Mol Sci*, 2019, 20: 23.
- [36] Mo G Y, Huang F, Fang Y, *et al.* Transcriptomic analysis in *Anemone flaccida* rhizomes reveals ancillary pathway for triterpene saponins biosynthesis and differential responsiveness to phytohormones [J]. *Chin J Nat Med*, 2019, 17(2): 131-144.
- [37] Zhang Y M, Chen Q C, Huang Y H, *et al.* Gene excavation and expression analysis of CYP and UGT related to the post modifying stage of gypenoside biosynthesis in *Gynostemma pentaphyllum* (Thunb.) Makino by comprehensive analysis of RNA and proteome sequencing [J]. *PLoS One*, 2021, 16(12): e0260027.
- [38] Ali M A, Alhemaïd F, Farah M A, *et al.* Transcriptome characterization of *Larrea tridentata* and identification of genes associated with phenylpropanoid metabolic pathways [J]. *PLoS One*, 2022, 17(3): e0265231.
- [39] Long N B, Zhong G W. The C-22 sterol desaturase Erg5 is responsible for ergosterol biosynthesis and conidiation in *Aspergillus fumigatus* [J]. *J Microbiol*, 2022, 60(6): 620-626.
- [40] Zhao M Z, Lin Y P, Wang Y F, *et al.* Transcriptome analysis identifies strong candidate genes for ginsenoside biosynthesis and reveals its underlying molecular mechanism in *Panax ginseng* C. A. Meyer [J]. *Sci Rep*, 2019, 9(1): 615.
- [41] Hou M Q, Wang R F, Zhao S J, *et al.* *Panax* ginsenosides in genus and their biosynthesis [J]. *Acta Pharm Sin B*, 2021, 11: 1813-1834.

[责任编辑 时圣明]