

基于代谢组学和转录组学的不同生长年限下银杏萜类生物合成关键基因表达分析

刘志强², 高 崎^{1,2*}, 李 航², 江美芳², 陈 冉²

1. 上海中医药大学中药学院, 上海 201203

2. 上海上药杏灵科技药业股份有限公司, 上海 201703

摘要:目的 以不同发育阶段不同树龄的银杏叶(银杏幼树叶, 多年生银杏树叶)为对象, 分析次生代谢物变化规律, 测定黄酮及萜内酯含量, 分析萜类生物合成的通路和关键基因表达, 为提高银杏萜内酯产量提供分子生物学数据。方法 采用液相色谱-质谱联用(LC-MS)技术, 并进行高通量转录组学测序(RNA-seq), 从差异表达基因中寻找次生代谢物生物合成的关联酶基因。结果 代谢数据表明, 不同树龄, 不同发育阶段的银杏叶次生代谢物明显不同, 黄酮类成分, 老树叶片与幼树叶 4~7 月含量均呈下降趋势; 同发育阶段, 幼树叶含量高于老树叶片含量; 萜内酯类成分, 老树叶片 4~7 月含量略微呈上升趋势, 幼树叶略微呈缓降趋势; 同发育阶段, 幼树叶含量高于老树叶片含量。转录组分析出萌发期, 老树叶片与幼树叶片生长状态相似, 5~7 月老树叶与幼树叶基因表达发生较大变化。筛选出 49 条与萜类合成有关的候选基因, 分析后发现甲羟戊酸途径(mevalonate pathway, MVA)中老树叶表达高于幼树叶; 甲基赤藻糖醇磷酸途径(methylerythritol 4-phosphate pathway, MEP)中幼树叶表达高于老树叶。结论 幼树叶比老树叶药用活性成分含量更高, 适合作为药材。银杏叶内萜类合成途径可能主要依赖于 MEP 途径, 且相关基因表达量与代谢成分相互证明。

关键词: 银杏; 测序; 转录组; 代谢组; 萜类

中图分类号: R286.12 文献标志码: A 文章编号: 0253-2670(2022)04-1138-10

DOI: 10.7501/j.issn.0253-2670.2022.04.022

Expression analysis of key genes in terpenoid biosynthesis of *Ginkgo biloba* under different growth years based on metabolomics and transcriptome

LIU Zhi-qiang¹, GAO Qi^{1,2}, LI Hang², JIANG Mei-fang², CHEN Ran²

1. School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

2. Shanghai Xingling Science and Technology Pharmaceutical Co., Ltd., Shanghai 201703, China

Abstract: Objective *Ginkgo biloba* leaves (*G. nursery* leaves, perennial *G. biloba* leaves) at different development stages and ages were used as objects to determine the content of flavonoids and terpene lactones in their secondary metabolites, to analyze the pathways of terpene biosynthesis and the expression of key genes, and to provide molecular biopharmaceutical data for improving the yield of ginkgo terpenolactone. **Methods** Based on liquid chromatography-mass spectrometry (LC-MS) metabolomic technology, and high-throughput transcriptomics sequencing (RNA-seq) to find the related enzymes of secondary metabolite biosynthesis from differentially expressed genes. **Results** It was found that the secondary metabolites of *G. biloba* leaves at different ages and developmental stages were significantly different in metabolomics. Firstly, the content of flavonoids, both in the tree leaves and nursery leaves, was presented by a decreasing trend from April to July. While at the same developmental stage, the content in nursery leaves was higher than that in tree leaves. Secondly, the content of terpene lactones in tree leaves was presented by a slightly increasing trend from April to July, however, a slightly decreasing trend in nursery leaves. While at the same development stage, the content in nursery leaves was higher than that in tree leaves. Then analyzed in the transcriptome, the growth status of old leaves was similar with the nursery leaves in germination stage, and there was a relatively large change in the gene expressions of old leaves and nursery leaves in May, June and July. After screening out 49 candidate genes related to terpenoid synthesis and analyzing them, it was found that the expression of tree leaves in mevalonate (MVA) pathway pathway was higher than that in nursery leaves; The later was higher than the former in

收稿日期: 2021-08-06

基金项目: 上海市国资委企业技术创新和能级提升项目(2018010)

作者简介: 刘志强, 男, 硕士研究生, 研究方向为中药创新及其产业化发展。E-mail: 546895057@qq.com

*通信作者: 高 崎, 男, 硕士生导师, 教授, 研究方向为中药创新及其产业化发展。E-mail: gaoqi@xingling.com.cn

methylethylthritol 4-phosphate (MEP) pathway. **Conclusion** The nursery leaves have a higher content of medicinal active ingredients than tree leaves, which are suitable as medicinal materials. The terpenoid synthesis pathway in *G. biloba* leaves may mainly depend on the MEP pathway, and the expression levels of related genes and metabolic components are mutually proven.

Key words: *Ginkgo biloba* L.; transcription; sequencing; metabolic group; terpenoids

银杏 *Ginkgo biloba* L. 又名白果树, 为银杏科银杏属的单科单属落叶乔木。银杏作为传统中药, 具有抗氧化、抗凋亡、改善脑血流、抑制血小板活性等多种药理作用^[1], 可用于治疗脑卒中、脑梗死、冠心病稳定型心绞痛等疾病的治疗^[2-3]。《中国药典》2020 年版规定, 银杏内酯和银杏黄酮为银杏叶提取物的主要药用成分, 其中黄酮醇苷质量分数不少于 24%, 萜内酯不少于 6%^[4]。随着银杏叶提取物制剂的广泛使用, 如何提高银杏药用活性成分的含量成为一个新的问题。一般而言, 植物次生代谢物的积累与发育阶段有着密切联系, 研究银杏主要次生代谢物的变化, 为其生产实践提供理论依据。

目前, 有关银杏的基因研究也伴随着测序技术的发展也在被广泛的开展。2016 年完成了银杏的基因组测序, 为银杏转录组测序提供了可参考基因, 这使得银杏次生代谢物生物合成的研究变得更为便利, 特别是涉及到黄酮醇苷的类黄酮生物合成通路

以及相关的功能基因都被频繁报道^[5]。本实验旨在运用代谢组数据结合转录测序结果, 采用新型分析软件与更为科学的算法对不同树龄不同生长阶段的银杏叶进行比较分析, 筛选银杏叶萜类生物合成的关键表达基因, 为阐述其生物合成途径提供依据, 为银杏叶种植采摘提供理论数据。

1 材料

银杏叶样品采集自四川省都江堰市申都中药有限公司种植石羊村种植基地 (N30°5'40", E103°41'10"), 根据基地实验记录数据分别于 2019 年 4 月下旬、2019 年 5 月下旬、2019 年 6 月下旬、2019 年 7 月下旬选取 2 年生幼树 (Y4~7) 和 20 年生老树 (L4~7) 进行采摘, 以液氮速冻后放于 -80 °C 保存, 用于总 RNA 提取。为保证采取样品的一致性, 采集相同地点银杏叶, 干燥后用作化学成分分析。每份样品设置 3 个生物学重复。对照品信息见表 1。

表 1 对照品收集信息
Table 1 Reference collection information

化合物	CAS	批号	厂家
奎宁酸	77-95-2	K29D9B77039	上海源叶生物科技有限公司
原儿茶酸	99-50-3	H21J9Z64031	上海源叶生物科技有限公司
没食子儿茶素	970-74-1	1297/17878	上海诗丹德标准技术服务有限公司
白果内酯	33570-04-6	P08J9F52398	上海源叶生物科技有限公司
表儿茶素	490-46-0	C20N6Q5882	上海源叶生物科技有限公司
银杏内酯 J	107438-79-9	D29J9G53845	上海源叶生物科技有限公司
银杏内酯 C	15291-76-6	C28S6G3986	上海源叶生物科技有限公司
3-O-β-D-吡喃葡萄糖基槲皮素	482-35-9	P25J9F65872	上海源叶生物科技有限公司
3-O-[α-L-吡喃鼠李糖(1"→2")]-[α-L-吡喃鼠李糖(1"→6")]-β-D-吡喃葡萄糖基山柰酚	55804-74-5	8257	上海诗丹德标准技术服务有限公司
芦丁	153-18-4	Y16M9S61523	上海源叶生物科技有限公司
银杏内酯 A	15291-75-5	W25A7K13730	上海源叶生物科技有限公司
3-O-[α-L-吡喃鼠李糖(1"→2")]-[α-L-吡喃鼠李糖(1"→6")]-β-D-吡喃葡萄糖基异鼠李素	104472-68-6	Y14S8H43976	上海源叶生物科技有限公司
银杏内酯 B	15291-77-7	D15D8G50608	上海源叶生物科技有限公司
3-O-[β-D-吡喃葡萄糖基(1"→2")]-α-L-吡喃鼠李糖-槲皮素	143016-74-4	8301	上海诗丹德标准技术服务有限公司
3-O-β-D-吡喃葡萄糖基山柰酚	480-10-4	Y19M8H36474	上海源叶生物科技有限公司
3-O-[α-L-吡喃鼠李糖(1"→6")]-β-D-吡喃葡萄糖基山柰酚	17650-84-9	Y15N8H48277	上海源叶生物科技有限公司
3-O-[α-L-吡喃鼠李糖(1"→6")]-β-D-吡喃葡萄糖基异鼠李素	55033-90-4	5771	上海诗丹德标准技术服务有限公司
3-O-[β-D-吡喃葡萄糖基(1"→2")]-α-L-吡喃鼠李糖-异鼠李素	604-80-8	S08A8D33447	上海源叶生物科技有限公司
3-O-[α-L-吡喃鼠李糖(1"→6")]-β-D-吡喃葡萄糖基丁香亭	53430-50-5	8260	上海诗丹德标准技术服务有限公司
3-O-[β-D-吡喃葡萄糖基(1"→2")]-α-L-吡喃鼠李糖-山柰酚	142451-65-8	8308	上海诗丹德标准技术服务有限公司
3-O-[6-O-香豆酰基-β-D-吡喃葡萄糖基(1"→2")]-α-L-吡喃鼠李糖-槲皮素	143061-65-8	8244	上海诗丹德标准技术服务有限公司
3-O-[6-O-香豆酰基-β-D-吡喃葡萄糖基(1"→2")]-α-L-吡喃鼠李糖-山柰酚	111957-48-3	8258	上海诗丹德标准技术服务有限公司
木犀草素	491-70-3	2226/17876	上海诗丹德标准技术服务有限公司
芹菜素	520-36-5	404/17869	上海诗丹德标准技术服务有限公司
白果素	521-32-4	Y15J9L63710	上海源叶生物科技有限公司
异银杏素	548-19-6	Y10N8Y47776	上海源叶生物科技有限公司
金松双黄酮	521-34-6	P03M9F55041	上海源叶生物科技有限公司
银杏酸 (C13: 0)	20261-38-5	P11D8F50448	上海源叶生物科技有限公司
银杏酸 (C15: 1)	22910-60-7	Y24M9H57091	上海源叶生物科技有限公司
银杏酸 (C17: 1)	111047-30-4	Y23M9H62018	上海源叶生物科技有限公司

2 方法

2.1 LC-MS 测定

2.1.1 样品制备 将采摘的样品烘干磨粉后精确称量各个时期样品粉末 0.5 g, 置具塞离心管中, 加入 50% 甲醇 10 mL, 涡旋振荡 1 min, 超声提取 15 min, 再振荡 1 min, 离心 (转速 6000 r/min) 10 min, 分取上清液于 25 mL 量瓶中, 重复上述操作 1 次, 合并 2 次上清液, 用 50% 甲醇定容至 25 mL。

2.1.2 色谱条件^[6] 样品于 Waters Acquity UPLC/Xevo G2-XS Q-TOF 超高效液相色谱质谱联用仪进行测定, 每个样品平行 3 针, 保证平行实验。色谱柱为 Waters Acquity UPLC BEH C₁₈ 柱 (100 mm×2.1 mm, 1.7 μm), 柱温 45 °C, 体积流量 0.2 mL/min, 进样体积 1 μL。流动相为 0.2 mol/L 醋酸铵溶液 (含 0.1% 甲酸) (A) - 乙腈 (B), 梯度洗脱: 0~60 min, 98%~2% A; 60~70 min, 2% A; 70~71 min, 2%~98% A; 71~75 min, 98% A。

2.1.3 质谱条件 负离子模式, Sensitivity 模式 (分辨率为 30 000); 毛细管电压 -2.5 kV; 样品锥孔电压 40 V; 源偏移电压 80 V; 源温 120 °C; 脱溶剂温度 450 °C; 锥孔气 50 L/h; 脱溶剂气体积流量 800 L/h; 雾化气压力 600 kPa; 质量数校正范围 m/z 50~1000, 校正溶液为 0.5 mmol/L 甲酸钠溶液, 体积流量 20 μL/min; 实时校正 lock spray 为 1 ng/μL 的亮氨酸脑啡肽溶液, m/z 554.2615; 数据采集方法为 MSE, 数据类型为 continuum, 能量范围 10~40 V, 扫描时间为 0.2 s。采用 Waters UNIFI 软件进行相关分析。

2.1.4 方法学考察 按照参考文献方法^[6], 取银杏叶样品进行精密度、稳定性、重复性试验, 结果 RSD 均小于 3%, 表明该方法稳定可靠。

2.1.5 数据分析 通过在国内外数据库, 包括 CNKI 数据库、PubMed、ScienceDirect、Web of Science 等, 以“银杏酮酯”“银杏叶提取物”“银杏”“银杏叶”“银杏花”“银杏果”“Ginkgo biloba extract”“Ginkgo biloba extract”“Ginkgo”“Ginkgo biloba”“Ginkgo biloba flowers”“Ginkgo nut”为关键字检索相关文献, 汇总银杏叶中可能的化学成分, 将化学成分名称、分子式导入到 UNIFI 软件 (Waters 公司) 建立银杏酮酯化学成分数据库。将样本质谱分析所得的原始数据导入配套的 UNIFI 工作站中, 根据建立的化学成分数据库, 提取所有离子, 并对提取离子排除分析, 通过对照品对照, 进一步确证化合物。

2.2 RNA 提取与文库构建

使用 TRIzol® 试剂 (植物组织 RNA 纯化试剂) 按照制造商说明 (Invitrogen, Carlsbad, CA, 美国) 从银杏叶组织中提取总 RNA, 使用 DNase I (TaKara) 去除基因组 DNA。然后用 2100 生物分析仪 (安捷伦科技公司, 圣克拉拉, 美国) 测定总 RNA 质量的完整性和纯度, 并使用 rd-2000 (NanoDrop Thermo Scientific, Wilmington, DE, 美国) 进行定量。符合文库构建要求 ($A_{260}/A_{280} = 1.8 \sim 2.2$, $A_{260}/A_{230} \geq 2.0$, $RIN \geq 8.0$, $28S:18S \geq 1.0$, $> 2 \mu g$)。利用 Illumina TruSeq™ RNA 样品制备试剂盒 (San Diego, CA) 制备银杏双叶 RNA-seq 转录组文库后, 在 Illumina HiSeq xten 测序仪 (Illumina, San Diego, CA) 上进行 2×150 bp 的双端测序, 每个样本选择 3 份进行生物学重复。

2.3 转录组数据与参考基因比对及转录本组装

通过使用 SeqPrep (<https://github.com/jstjohn/SeqPrep>) 和 Sickle (<https://github.com/najoshi/sickle>) 对大量的 raw reads 进行了裁剪和质量控制得到 clean data, 用 HISAT2 (<http://ccb.jhu.edu/software/hisat2/index.shtml>) 与银杏基因组 (<http://gigadb.org/dataset/100209>) 进行序列比对^[7], 得到 mapped reads 并进行质量评估。

基于所选参考基因组序列, 使用 StringTie (<http://ccb.jhu.edu/software/stringtie/>) 软件对 Mapped Reads 进行拼接, 并与原有的基因组注释信息进行比较, 寻找原来未被注释的转录区, 发掘该物种的新转录本和新基因, 从而补充和完善原有的基因组注释信息^[8-9]。

2.4 基因功能分析及表达差异基因筛选

将获得的所有 Unigene 与 NCBI 蛋白数据库 nonredundant (NR) 进行比对, 以 E 值 $< 1 \times 10^{-5}$ 为阈值, 通过 BLAST 算法进行功能注释^[10]。同时, 将 Unigene 在 Swiss-Prot、Pfam、COG、GO 和 KEGG 各大数据库的功能注释进行汇总, 分析预测其功能分类和参与的生物学途径。

为了识别 2 个不同样本间的差异表达基因 (differential expression genes, DEGs), 使用 TPM 法衡量每个转录本表达水平。借用 RSEM (<http://deweylab.biostat.wisc.edu/rsem/>) 对基因进行了定量分析^[11]。利用 R 统计软件包 EdgeR (<http://www.bioconductor.org/packages/2.12/bioc/html/edgeR.html>) 进行差异表达分析。

此外,以 Bonferroni 为多重验证法,以 P 值 ≤ 0.01 为阈值,用 Goatools (<https://github.com/tanghaibao/Goatools>)和 KOBAS(<http://kobas.cbi.pku.edu.cn/home.do>)进行了 GO 功能富集和 KEGG 通路分析^[12-13]。

3 结果与分析

3.1 LC-MS 数据分析

3.1.1 银杏叶化学成分分析

为了对银杏叶的化学

成分进行研究,采用 50%甲醇溶液超声提取的方法,分别获得不同年限不同发育阶段银杏叶除银杏双黄酮及烷基酚类以外的其余成分。通过 UPLC Q-TOF/MS 分析得到银杏叶总离子流图(TIC),见图 1。根据总离子流图所得到的化合物精确相对分子质量、二级碎片信息,与对照品和实验室构建的数据库比对,共鉴定出 55 个化合物,其中 30 个已与对照品比对,结果见表 2。

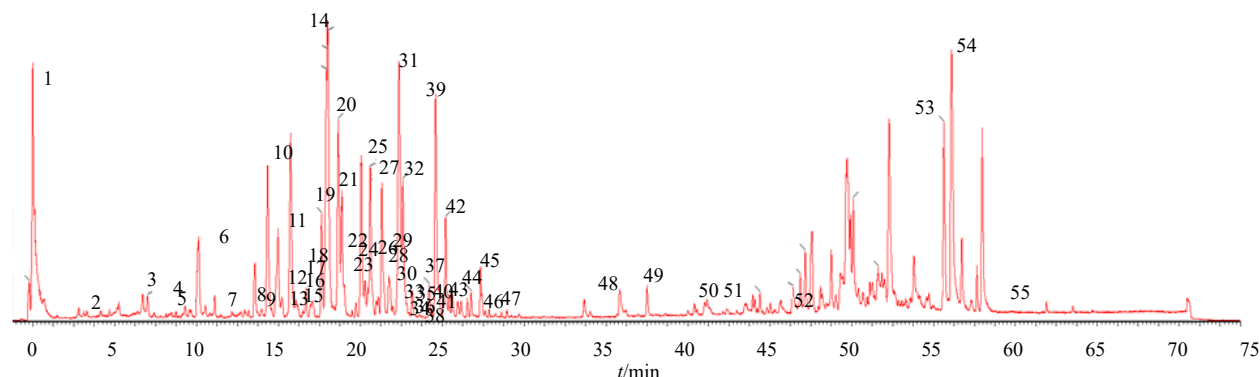


图 1 银杏叶总离子流图

Fig. 1 Total ion flow diagram of *G. biloba* leaves

表 2 银杏叶化合物质谱信息

Table 2 Mass spectrum information of *G. biloba* leaves compounds

峰号	化合物名称	分类	分子式	$m/z[M-H]^-$	误差($\times 10^{-6}$)	t_R/min	二级碎片
1*	奎宁酸	有机酸类	$C_7H_{12}O_6$	191.055 0	-2.8	1.25	173.05
2*	原儿茶酸	有机酸类	$C_7H_6O_4$	153.018 5	-2.1	4.97	135.03, 109.03
3*	没食子儿茶素	黄烷醇类	$C_{15}H_{14}O_7$	305.065 2	-2.0	7.96	243.03, 177.02, 137.02, 125.02, 109.03
4*	儿茶素	黄烷醇类	$C_{15}H_{14}O_6$	289.070 7	2.0	8.25	137.02, 125.02
5	3-O- β -D-吡喃葡萄糖基二氢山柰酚	黄酮类	$C_{21}H_{22}O_{11}$	449.107 8	0	8.73	287.06, 259.06
6*	白果内酯	萜内酯类	$C_{15}H_{18}O_8$	325.091 8	-0.6	11.36	237.11, 193.12, 163.11
7*	表儿茶素	黄烷醇类	$C_{15}H_{14}O_6$	289.070 6	-0.4	11.80	273.08, 137.02, 123.04
8	3-O- β -D-吡喃葡萄糖(1" $\rightarrow$ 2")-[α -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基山柰酚	黄酮类	$C_{33}H_{40}O_{21}$	771.196 5	-0.6	14.43	609.14, 462.08, 301.03
9*	银杏内酯 J	萜内酯类	$C_{20}H_{24}O_{10}$	423.128 1	-0.6	14.81	379.14, 367.14, 349.09
10*	银杏内酯 C	萜内酯类	$C_{20}H_{24}O_{11}$	439.123 4	-0.1	15.58	411.13, 383.13, 321.13
11	3-O- β -D-吡喃葡萄糖(1" $\rightarrow$ 2")-[α -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基山柰酚	黄酮类	$C_{33}H_{40}O_{20}$	755.202 3	-0.6	16.03	593.15, 463.09, 301.03
12	3-O- β -D-吡喃葡萄糖基杨梅黄酮	黄酮类	$C_{21}H_{20}O_{13}$	479.082 5	-1.1	16.09	316.02, 287.02, 271.02
13*	3-O- $[\alpha$ -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基杨梅黄酮	黄酮类	$C_{27}H_{30}O_{17}$	625.140 3	1.4	16.22	316.02
14	3-O- β -D-吡喃葡萄糖(1" $\rightarrow$ 2")-[α -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基山柰酚同分异构体	黄酮类	$C_{33}H_{40}O_{20}$	755.201 9	-0.3	16.99	593.15, 463.09, 301.03
15	3-O- $[\alpha$ -L-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-7-O- β -D-吡喃葡萄糖基槲皮素	黄酮类	$C_{42}H_{46}O_{23}$	917.233 3	-1.3	17.94	755.18, 609.14, 300.03
16*	3-O- β -D-吡喃葡萄糖基槲皮素	黄酮类	$C_{21}H_{20}O_{12}$	463.087 3	-0.4	18.86	300.03, 271.02, 255.03
17	3-O- β -D-吡喃鼠李糖基槲皮素	黄酮类	$C_{21}H_{20}O_{11}$	447.092 0	-0.4	18.86	285.04, 271.09, 151.00
18*	3-O- $[\alpha$ -L-吡喃鼠李糖(1" $\rightarrow$ 2")]- $[\alpha$ -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基山柰酚	黄酮类	$C_{33}H_{40}O_{19}$	739.207 2	0.3	18.88	447.09, 284.03
19*	芦丁	黄酮类	$C_{27}H_{30}O_{16}$	609.145 8	0	19.18	463.09, 300.03, 271.02
20*	银杏内酯 A	萜内酯类	$C_{20}H_{24}O_9$	407.133 3	-0.6	19.27	379.14, 351.15, 319.16
21*	3-O- $[\alpha$ -L-吡喃鼠李糖(1" $\rightarrow$ 2")]- $[\alpha$ -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基异鼠李素	黄酮类	$C_{34}H_{42}O_{20}$	769.218 1	-0.5	19.30	447.11, 284.05

续表 2

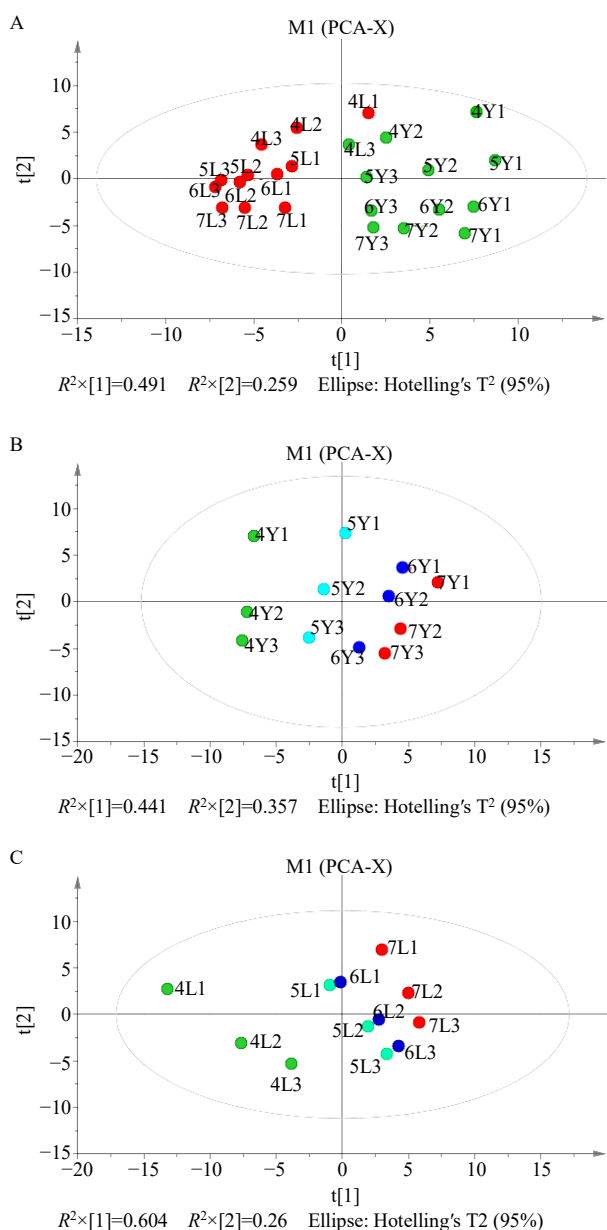
峰号	化合物名称	分类	分子式	$m/z[M-H]^-$	误差 ($\times 10^{-6}$)	t_R/min	二级碎片
22	3-O-[6-O- β -D-吡喃葡萄糖基(1" $\rightarrow$ 6")]-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-山奈酚	黄酮类	C ₄₂ H ₄₆ O ₂₂	901.238 8	-0.7	19.58	739.19, 593.15, 284.03
23*	银杏内酯 B	萜内酯类	C ₂₀ H ₂₄ O ₁₀	423.128 3	-0.2	19.89	395.13, 367.14, 349.13
24	7-O- β -D-吡喃葡萄糖基芹菜素	黄酮类	C ₂₁ H ₂₀ O ₁₀	431.096 9	-0.2	21.22	285.04, 255.03, 227.03
25*	3-O-[β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-槲皮素	黄酮类	C ₂₇ H ₃₀ O ₁₆	609.145 3	0.5	21.31	447.09, 300.03
26*	3-O- β -D-吡喃葡萄糖基山奈酚	黄酮类	C ₂₁ H ₂₀ O ₁₁	447.091 9	0.3	21.53	255.032, 27.03
27*	3-O-[α -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基山奈酚	黄酮类	C ₂₇ H ₃₀ O ₁₅	593.150 7	0.9	21.85	447.09, 285.04, 255.03
28*	3-O-[α -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基异鼠李素	黄酮类	C ₂₈ H ₃₂ O ₁₆	623.160 7	0.7	22.56	315.05, 271.02
29*	3-O-[α -L-吡喃鼠李糖(1" $\rightarrow$ 6")]- β -D-吡喃葡萄糖基丁香亭	黄酮类	C ₂₉ H ₃₄ O ₁₇	653.170 9	0.5	22.98	345.06
30	3'-O- β -D-葡萄糖基木犀草素	黄酮类	C ₂₁ H ₂₀ O ₁₁	447.091 9	-3.8	23.50	285.04, 256.03, 133.03
31*	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-槲皮素	黄酮类	C ₃₆ H ₃₆ O ₁₈	755.181 2	-3.6	23.60	609.15, 447.09, 300.03
32*	3-O-[β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-山奈酚	黄酮类	C ₂₇ H ₃₀ O ₁₅	593.150 8	-2.9	23.84	284.03
33	3-O-[β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-丁香亭	黄酮类	C ₂₉ H ₃₄ O ₁₇	653.171 1	-4.0	23.90	345.06
34	3-O- α -L-吡喃鼠李糖-山奈酚	黄酮类	C ₂₁ H ₂₀ O ₁₀	431.097 0	-2.1	24.39	268.04
35*	3-O-[β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-异鼠李素	黄酮类	C ₂₈ H ₃₂ O ₁₆	623.160 3	-2.6	24.41	314.04
36	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-槲皮素同分异构体 3	黄酮类	C ₃₆ H ₃₆ O ₁₈	755.181 1	-3.7	24.43	609.15, 447.09, 300.03
37*	槲皮素	黄酮类	C ₁₅ H ₁₀ O ₇	301.034 4	-2.7	24.8	271.02, 227.03, 151.00
38	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-槲皮素同分异构体 1	黄酮类	C ₃₆ H ₃₆ O ₁₈	755.183 4	0.1	25.45	609.15, 447.09, 300.03
39*	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-山奈酚	黄酮类	C ₃₆ H ₃₆ O ₁₇	739.185 3	-5.3	25.84	593.15, 284.03, 255.03
40*	木犀草素	黄酮类	C ₁₅ H ₁₀ O ₆	285.039 2	-0.1	26.05	133.03
41	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-异鼠李素	黄酮类	C ₃₇ H ₃₈ O ₁₈	769.196 2	-3.2	26.32	623.16, 314.04
42	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-槲皮素同分异构体 2	黄酮类	C ₃₆ H ₃₆ O ₁₈	755.182 9	-4.8	26.46	609.15, 447.09, 300.03
43	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-山奈酚同分异构体 1	黄酮类	C ₃₆ H ₃₆ O ₁₇	739.188 7	0.5	27.19	593.15, 284.03, 255.03
44	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-山奈酚同分异构体 2	黄酮类	C ₃₆ H ₃₆ O ₁₇	739.187 6	-7.6	28.02	593.15, 284.03, 255.03
45	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-山奈酚同分异构体 3	黄酮类	C ₃₆ H ₃₆ O ₁₇	739.188 2	1.0	28.59	593.15, 284.03, 255.03
46*	芹菜素	黄酮类	C ₁₅ H ₁₀ O ₅	269.044 5	-2.3	28.86	117.03
47	3-O-[6-O-香豆酰基- β -D-吡喃葡萄糖基(1" $\rightarrow$ 2")]- α -L-吡喃鼠李糖-山奈酚同分异构体 4	黄酮类	C ₃₆ H ₃₆ O ₁₇	739.188 1	-2.9	29.12	593.15, 284.03, 255.03
48	穗花杉双黄酮	双黄酮	C ₃₀ H ₁₈ O ₁₀	537.081 6	0.1	34.92	417.06, 375.05, 331.06
49*	白果素	双黄酮	C ₃₁ H ₂₀ O ₁₀	551.097 8	-3.9	37.10	519.10
50	红杉双黄酮	双黄酮	C ₃₁ H ₂₀ O ₁₀	551.097 2	-3.9	41.26	431.08, 389.07
51*	异银杏素	双黄酮	C ₃₂ H ₂₂ O ₁₀	565.113 7	-4.0	42.43	533.09, 389.07, 629.21
52*	金松双黄酮	双黄酮	C ₃₃ H ₂₄ O ₁₀	579.128 9	-1.7	46.90	547.10, 415.05
53*	银杏酸 (C13: 0)	烷基酚类	C ₂₀ H ₃₂ O ₃	319.226 3	-2.8	56.89	275.24
54*	银杏酸 (C15: 1)	烷基酚类	C ₂₂ H ₃₄ O ₃	345.241 9	-3.5	57.35	301.25
55*	银杏酸 (C17: 1)	烷基酚类	C ₂₄ H ₃₈ O ₃	373.272 9	-9.3	59.22	329.28

*已与对照品比对鉴定

*identified by comparison with the reference

3.1.2 多元统计分析 为了分析银杏叶不同发育阶段代谢产物组的动态变化, 对不同生长年限不同发育阶段的银杏叶进行非监督模式识别的主成分分析 (principal component analysis, PCA) 图 2 分别为多年生老树叶与幼苗叶片; 多年生老树叶不同生长阶段;

幼苗叶不同生长阶段的 PCA 分析散点图^[4]。图 2-A 显示出多年生老树叶紧密分布在 PC1 的负半轴, 而幼苗叶分布在正半轴; 图 2-B、C 表明多年生老树叶与幼苗叶 4~7 月逐渐从 PC1 负半轴过渡至 PC1 正半轴。结果表明, 无论是树龄还是生长阶段对于银杏叶的代



A-老树与幼树整体成分差异 B-幼树不同生长阶段成分差异 C-老树不同生长阶段成分差异, 字母前数字代表月份

A-global composition difference between old and young trees B-component differences at different growth stages of young trees C-composition differences of old trees at different growth stages, the number before the letter represents month

图 2 次生代谢物主成分分析散点图

Fig. 2 Scatter diagram of secondary metabolism principal component analysis

谢物质组成均会产生较大变化。为了进一步分析不同生长年限不同发育阶段银杏叶黄酮醇与萜内酯类成分的含量变化, 通过半定量方法对其进行相对含量测定。图 3 可明显看出各类成分的变化趋势, 黄酮类成分, 老树叶片与幼树叶 4~7 月含量均呈下降趋势; 同时期内, 幼树叶含量高于老树叶片量。萜内酯类成

分, 大树叶片 4~7 月含量略微呈上升势, 幼树叶略微呈缓降趋势; 同时期内, 幼树叶含量高于老树叶片量。

3.2 转录转录组数据分析注释分析

3.2.1 银杏转录测序质量分析及注释情况 通过 Illumina Hiseq xten 测序平台测序, 最后处理得到的 8 组样品数据统计如表 3 所示。各样品碱基质量超过 Q_{30} 的比例均在 90% 以上, 且 GC 含量均在 50% 左右, 测序组装效果好; 与银杏参考基因进行序列比对, 对比效率均在 90% 以上, 测序精度高。

组装拼接后共得 39 137 条 Unigene, 将转录组组装获得的所有基因和转录本与 6 大数据库 (NR、Swiss-Prot、Pfam、COG、GO 和 KEGG) 进行比对。从图 4 可以看出 NR 数据库可注释到 31 936 条; Swiss-Prot 数据库可注释到 25 235 条; Pfam 数据库可注释到 21 561 条; COG 数据库可注释到 28 922 条; GO 数据库可注释到 27 503 条; KEGG 数据库可注释到 13 504 条; 共注释到 32 358 条, 占 Unigene 总数的 82.68%。

3.2.2 样本间表达量分析 通过基因的表达矩阵, 可验证实验设计的合理性, 了解样本间特别是生物学重复间的相关性。图 5 可以表明 L4 和 Y4, Y5 和 Y6、Y7, L5 和 L6、L7 的基因表达相关性更高。说明在 4 月树叶处于萌发期, 老树叶片与幼树叶片生长状态相似, 5~7 月老树叶与幼树叶基因表达发生较大变化。

3.3 差异表达基因表达量比较及 KEGG 通路注释分析

选取不同树龄差异最大的 L7 与 Y7; 老树基因表达差异最大的 L4 与 L7; 幼树叶基因表达差异最大的 Y4 和 Y5 进行差异分析。发现 L7vsY7 存在 3811 条显著差异基因, 上升 2068 条, 下降 1743 条; 发现 L4vsL7 存在 4836 条显著差异基因, 上升 2527 条, 下降 2309 条; 发现 Y4vsY5 存在 3060 条显著差异基因, 上升 1510 条, 下降 1550 条, 结果如图 6 所示。将上述组别的差异表达基因分别进行 KEGG 通路注释, 由于银杏叶中的活性成分为萜内酯和黄酮醇类, 其中黄酮醇类成分为次级代谢产物, 因此主要分析“其他次级代谢产物生物合成 (biosynthesis of other secondary metabolites)”和“萜类和聚酮类化合物的代谢 (metabolism of terpenoids and polketides)”。为了更清楚的了解所关注的黄酮类成分及萜内酯类成分与差异表达基

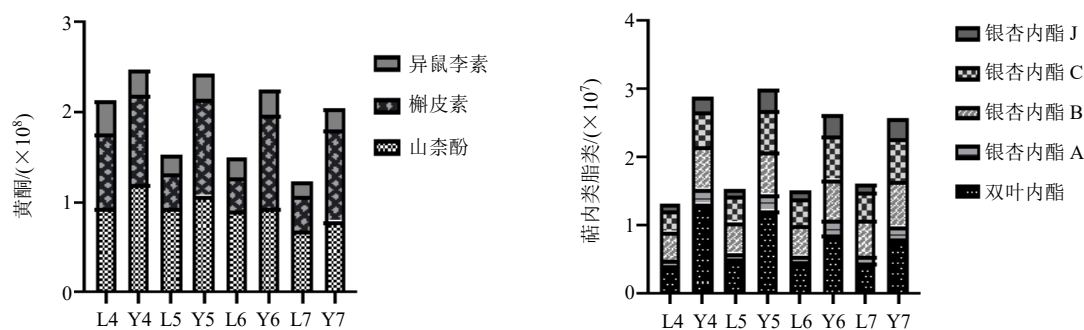


图 3 不同发育阶段不同树龄银杏黄酮类与银杏萜内酯类成分相对含量

Fig. 3 Contents of flavonoids and ginkgolide in different developmental stages of *G. biloba*

表 3 转录组测序数据质量分析

Table 3 Transcriptome sequencing data quality analysis

编号	clean reads	clean bases	误差率/%	Q20/%	Q30/%	GC/%	总 mapped 数 (占比/%)	multiple mapped 数 (占比/%)	Uniquely mapped 数 (占比/%)
L4	58897500	8829161903	0.0234	98.67	95.74	46.26	57 040 427 (96.85)	5 151 043 (8.75)	51 889 384 (88.10)
L5	52760088	7845051890	0.0234	98.65	95.76	46.84	50 411 574 (95.55)	8 014 104 (15.19)	42 397 470 (80.36)
L6	57113338	8544294953	0.0234	98.68	95.79	46.98	54 453 219 (95.34)	7 026 826 (12.30)	47 426 393 (83.04)
L7	42519962	6281708055	0.0245	98.23	94.67	45.10	40 438 620 (95.11)	3 486 004 (8.20)	36 952 616 (86.91)
Y4	50156736	7503478159	0.0233	98.74	95.90	46.00	48 811 805 (97.32)	3 122 339 (6.23)	45 689 466 (91.09)
Y5	46774336	6991040465	0.0231	98.79	96.05	45.90	45 382 803 (97.03)	3 295 340 (7.05)	42 087 463 (89.98)
Y6	58486076	8720165448	0.0233	98.71	95.84	46.69	56 448 409 (96.52)	4 790 414 (8.19)	51 657 995 (88.33)
Y7	47434370	6959127445	0.0242	98.34	94.98	46.24	45 389 674 (95.69)	3 925 925 (8.28)	41 463 749 (87.41)

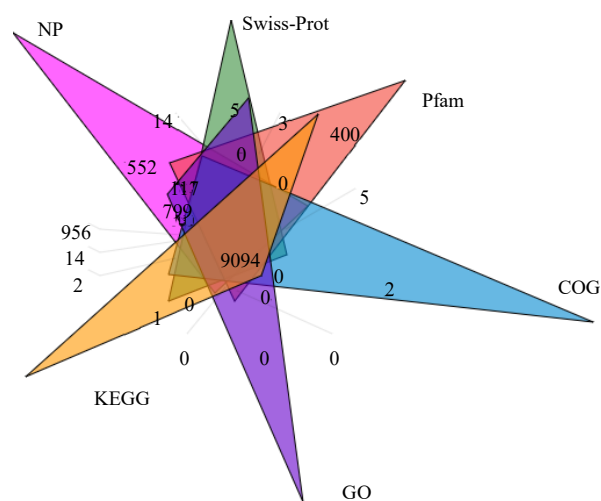


图 4 功能注释统计图

Fig. 4 Functional annotation statistical diagram

因是否在有关通路上存在相关性,对 3 组差异基因再进行 KEGG 富集分析,筛选出 flavonoid biosynthesis 和 terpenoid backbone biosynthesis 2 条代谢通路来分析相关性(表 4)。从结果可以看出幼树叶在不同发育阶段表达基因差异最大的 Y4 和 Y5 在黄酮和萜内酯的生物合成上均不具有显著性富集;

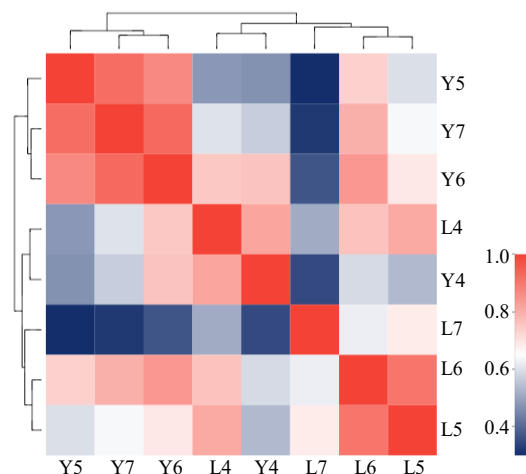


图 5 表达量分析

Fig. 5 Expression quantity analysis

老树叶在不同发育阶段表达基因差异最大的 L4 与 L7 在黄酮的生物合成上不具有显著性富集,在萜内酯的生物合成上存在显著性富集;不同树龄表达基因差异最大的 L7 与 Y7 在黄酮和萜内酯的生物合成上均具有显著性富集。根据 KEGG 通路注释结果图 7 可以看出, Y4vsY5、L4vsL7、L7vsY7 的差异表达基因都在这 2 类代谢途径中得到了注释。

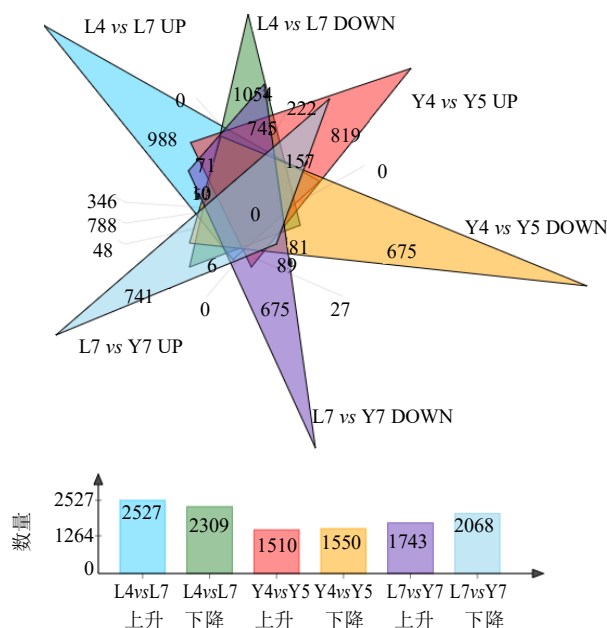


图 6 差异表达基因 Venn 图
Fig. 6 Venn diagram of differentially expressed gene

表 4 差异基因与通路相关性

Table 4 Correlation between different genes and pathways

组别	黄酮生物合成途径	萜类骨架生物合成途径
Y4vsY5	0.640	0.230
L4vsL7	0.330	0.036*
L7vsY7	0.025*	0.021*

表中数值表示差异基因和黄酮、萜类骨架生物合成途径的相关性，
* $P < 0.05$ 即差异显著，** $P < 0.01$ 即差异极显著

In the table represents the difference between genes and flavonoid and terpene biosynthesis of skeleton relationship between the way, * $P < 0.05$ says significant difference, ** $P < 0.01$ says that the difference is significant

3.4 银杏叶中萜类化合物生物合成相关基因表达分析

对 KEGG 数据库黄酮生物合成 (map00941) 和萜类生物合成 (map00900) 2 条途径注释分析, 发现相关蛋白均被多条 Unigen 注释到。注释情况见表 5。

有关银杏黄酮类生物合成相关基因表达分析已经被报道^[5], 这里着重对银杏萜类成分生物合成相

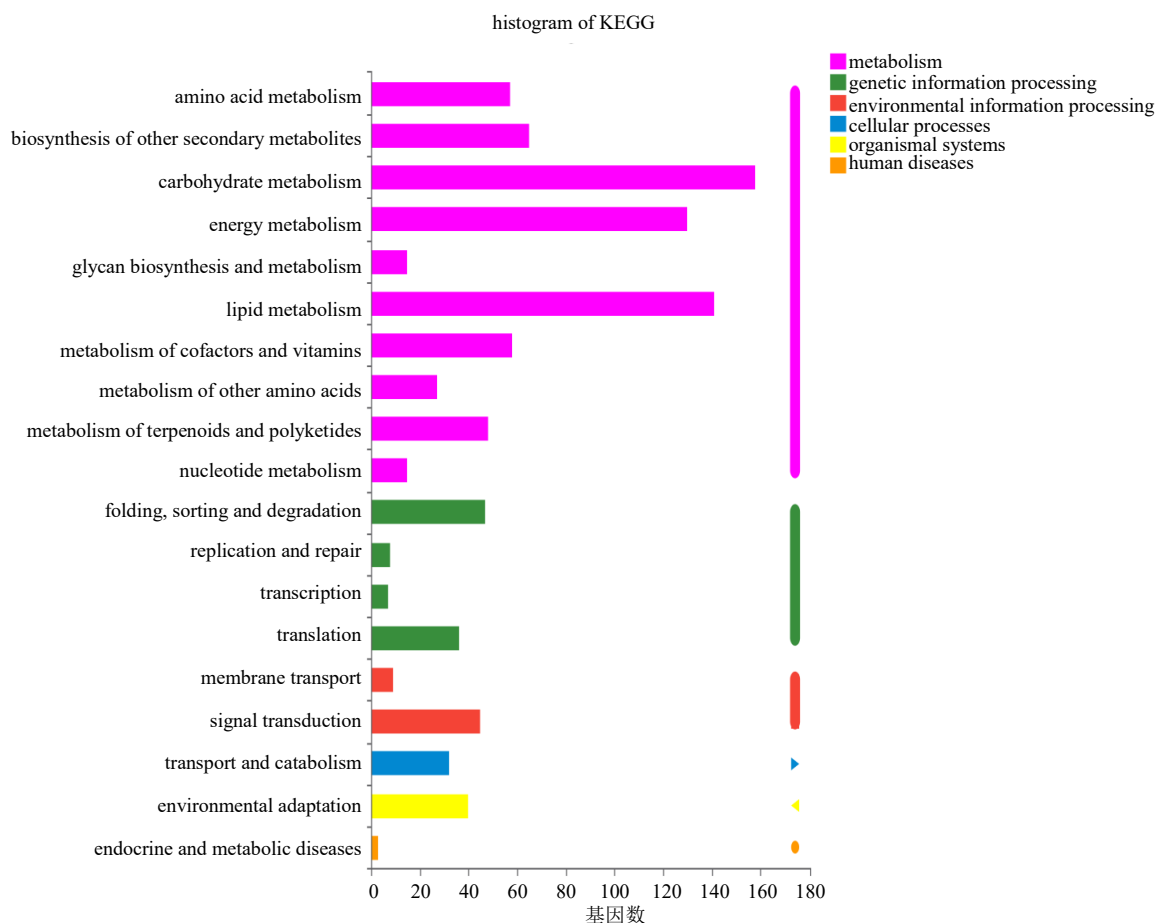


图 7 差异基因 KEGG 富集分析统计图
Fig. 7 Statistical map of enrichment analysis of different genes KEGG

表 5 萜类与黄酮类生物合成相关基因注释

Table 5 Comments on genes related to biosynthesis of terpenoids and flavonoids

通路	基因名称	中文名称	缩写	注释到的 Unigene 数
萜类 MVA 合成途径	acetyl-CoA C-acetyltransferase	乙酰 CoA 酰基转移酶	<i>AACT</i>	5
	hydroxymethylglutaryl-CoA reductase	羟基甲基戊二酰-CoA 还原酶	<i>HMGCR</i>	25
	mevalonate kinase	甲羟戊酸激酶	<i>MK</i>	1
	phosphomevalonate kinase	磷酸甲羟戊酸激酶	<i>PMVK</i>	4
	diphosphomevalonate decarboxylase	二磷酸甲羟戊酸脱羧酶	<i>MVD</i>	2
萜类 MEP 合成途径	1-deoxy-D-xylulose-5-phosphate synthase	1-脱氧-D-木酮糖-5-磷酸合酶	<i>DXS</i>	6
	1-deoxy-D-xylulose-5-phosphate reductoisomerase	1-脱氧-D-木酮糖-5-磷酸还原异构酶	<i>DXR</i>	1
	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	2-C-甲基-D-赤藓糖醇 4-磷酸胞苷酰转移酶	<i>CMS</i>	1
	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	4-二磷酸胞苷-2-C-甲基-D-赤藓糖醇激酶	<i>CMK</i>	2
	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	2-C-甲基-D-赤藓糖醇 2,4-环二磷酸合酶	<i>MCS</i>	2
黄酮类合成途径	chalcone synthase	查耳酮合酶	<i>CHS</i>	17
	flavonol synthase	黄酮醇合酶	<i>FLS</i>	19
	flavanone 3-dioxygenase	黄烷酮 3-羟化酶	<i>F3H</i>	8
	chalcone isomerase	查耳酮异构酶	<i>CHI</i>	3
	flavonoid 3',5"-hydroxylase	类黄酮 3',5"-羟化酶	<i>F3'5'H</i>	8
	flavonoid 3'-hydroxylase	类黄酮 3'-羟化酶	<i>F3'H</i>	5
	dihydroflavonol 4-reductase	二氢黄酮醇 4-还原酶	<i>DFR</i>	19
	anthocyanidin synthase	花色素合成酶	<i>ANS</i>	7
	anthocyanidin reductase	花色素还原酶	<i>ANR</i>	11

关表达基因进行分析。根据药典记载,银杏药用活性成分主要为总黄酮醇苷和萜内酯 2 类。其中,萜内酯为银杏特有成分,也是银杏叶提取物的质量控制检测标准。

萜类合成通路主要分为 MVA 及 MEP,通过对萜类合成通路基因进行筛选注释,发现在这 2 条通路上的关键基因均被注释,且存在显著性差异。本着转录研究为基因动态表达的原则,为了寻找银杏萜类合成的规律,本实验对不同树龄,不同生长阶段的银杏在这 2 条通路上的多个关键基因表达进行比较分析。发现这些基因虽然大多都被多个 Unigene 所注释,但 Unigene 其表达模

式趋于一致,故选取表达量最高的 Unigene 作为丰度衡量。

图 8 表现出, MVA 途径中 *AACT*、*HMGCR*、*MVD* 3 个基因表达量超过 200,且均为老树, *MK*、*PMVK* 表达较低。从同一发育阶段比较来看,老树 4~7 月均高于幼树叶;从不同发育阶段来看,老树呈上升趋势,在 6、7 月表达高于 4、5 月;幼树叶则出现缓降趋势,6、7 月表达低于 4、5 月。MEP 途径中 *DXR* 和 *MCS* 表达最高,从同一发育阶段来看,幼树叶 4~7 月表达均高于老树;从不同发育阶段看,幼树叶呈缓降趋势,老树呈上升趋势。

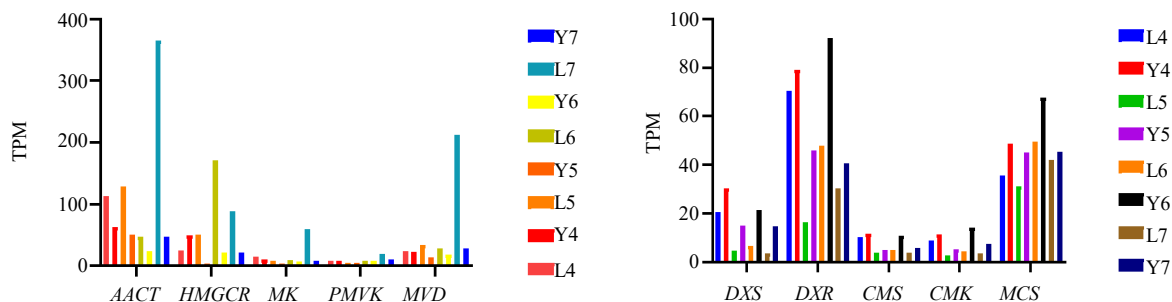


图 8 萜类生物合成途径功能基因表达量分析

Fig. 8 Functional gene expression analysis of terpenes biosynthesis pathway

4 讨论

4.1 不同树龄不同发育阶段银杏叶活性成分代谢差异

通过对不同树龄,不同生长阶段的银杏叶次级代谢物进行 PCA 分析,发现无论是树龄差异还是生长阶段的不同都会产生明显的差别。银杏叶提取物的药用活性成分主要是黄酮醇苷及银杏内酯 2 大类次级代谢物,通过对黄酮醇苷(槲皮素、异鼠李素、山柰素)和银杏内酯含量测定,发现幼树叶中 2 类成分均明显高于老树叶含量,这佐证了选取银杏幼树叶入药的合理性,由于采摘后重新生长的叶片代谢积累会发生较大改变,缺少后续积累数据,对于 7 月是否为最佳采摘时期有待进一步研究。

4.2 银杏萜类生物合成的关键基因表达

本次转录测序检测到表达基因共 39 137 个,其中已知基因 33 663 个,新基因 5474 个;表达转录本共 59 312 个,其中已知转录本 31 947 个,新转录本 27 365 个。根据基因表达分析发现幼树叶比老树叶代谢更为旺盛,幼树每个月整体基因表达都具有显著性差异,老树叶经历 4 个月的生长后才表现出整体基因表达差异。经过 KEGG 富集分析发现与生长阶段这一因素相比,树龄的差异导致银杏黄酮类和萜类的差异更为明显,这也与代谢成分测定的结果相吻合。通过 KEGG 通路注释,筛选出 49 条与银杏萜类生物合成的相关基因,并进一步分析了这些基因的表达状况。结果表明,银杏叶萜类合成可能更依赖于 MEP 通路,MEP 上的结构基因表达模式与代谢含量结果相对应,特别是 DXS 的表达模式几乎与银杏内酯变化一致。DXS 可能是银杏内酯生物合成的主要限速基因。利用基因工程技术,提高 DXS 的表达,从而提升银杏内酯的量是具有一定的可行性,这也是本研究的下一步工作。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 李思佳,耿剑亮,张悦,等.银杏药理作用研究进展[J].药物评价研究,2017,40(6):731-741.
- [2] 李春雷,张峰.维生素、叶酸联合银杏叶提取物对老年

脑卒中患者认知功能影响研究[J].中国生化药物杂志,2015,35(8):107-109.

- [3] 张鹏飞,廖丽君,谭玉萍.银杏叶提取物治疗慢性阻塞性肺疾病作用机制的研究进展[J].中国全科医学,2017,20(15):1906-1910.
- [4] 中国药典[S].四部.2015:1491-1493.
- [5] 刘伟,王俊焱,李萌,等.基于转录组测序的银杏类黄酮生物合成关键基因表达分析[J].中草药,2018,49(23):5633-5639.
- [6] 何厚洪,杜昕,金辉辉,等.银杏叶提取物 HPLC 指纹图谱研究及共有峰的 LC-MS 鉴定[J].药物分析杂志,2017,37(6):1107-1112.
- [7] Grabherr M G, Haas B J, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome[J].*Nat Biotechnol*, 2011, 29(7): 644-652.
- [8] 鲁光辉,李新峰,冯亮,等.银杏内酯 B 对癫痫大鼠血脑屏障保护作用的研究[J].药物评价研究,2020,43(4):765-772.
- [9] Conesa A, Götz S, García-Gómez J M, et al. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research[J].*Bioinformatics*, 2005, 21(18): 3674-3676.
- [10] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes[J].*Nucleic Acids Res*, 2000, 28(1): 27-30.
- [11] Li B, Dewey C N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome[J].*BMC Bioinformatics*, 2011, 12: 323.
- [12] Robinson M D, McCarthy D J, Smyth G K. edgeR: A bioconductor package for differential expression analysis of digital gene expression data[J].*Bioinformatics*, 2010, 26(1): 139-140.
- [13] Xie C, Mao X Z, Huang J J, et al. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases[J].*Nucleic Acids Res*, 2011, 39: W316-W322.
- [14] 秦伟瀚,花雷,郭延垒,等. UPLC-Q-TOF-MS 结合代谢组学分析冬虫夏草不同部位的差异性[J].中国实验方剂学杂志,2018,24(21):69-76.

[责任编辑 时圣明]