

文章编号: 1000-5862(2015)06-0570-05

基于转录组高通量测序分析白光对杜仲愈伤组织中绿原酸含量的影响

张俊娥¹, 邓华锋², 郑文强¹, 江京¹, 金聪¹, 赵夫凯¹

(1. 江西师范大学生命科学学院, 江西省亚热带植物资源保护与利用重点实验室, 江西 南昌 330022)

2. 江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 利用 Illumina HiSeq™ 2500 平台对光暗培养的杜仲愈伤组织进行了转录组高通量测序, 利用 BLAST 软件进行差异表达基因的功能注释和富集分析, 就白光(光强为 12 000 lux, 16 h 光照, 8 h 黑暗)对杜仲愈伤组织中绿原酸含量的影响进行了研究. 结果表明: 通过 Trinity 软件合并组装后共获得 62 030 个 Unigenes, 通过 BLASTX 比对, 共获得 25 417 (40.98%) 个有注释信息的 Unigenes. 通过对 KEGG 通路进行深入分析表明, 白光(光强为 12 000 lux, 16 h 光照, 8 h 黑暗)培养杜仲愈伤组织 18 d 与绿原酸合成相关的 3 种酶基因上调表达, 它们分别为苯丙氨酸解氨酶(EC 4.3.1.24, phenylalanine ammonia-lyase, PAL)、肉桂酸-4-羟基化酶(EC 1.14.13.11, trans-cinnamate 4-monooxygenase, C4H)、奎宁酸羟基肉桂酰转移酶(EC 2.3.1.133, Shikimate *o*-hydroxyl-cinnamoyltransferase, HCT). 由此推断, 白光能促进杜仲愈伤组织中绿原酸的积累. 研究结果为获取高产绿原酸资源和杜仲遗传分析提供有价值的资料.

关键词: 杜仲; 愈伤组织; 白光; 绿原酸; 转录组高通量测序

中图分类号: S 567 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.06.05

0 引言

杜仲(*Eucommia ulmoides* Oliv.) 是我国特有的经济树种, 属于杜仲科杜仲属植物, 是一种以药用为主并具有多种用途的树种, 具有很高的药用价值和营养保健功效^[1]. 杜仲的重要次生代谢物苯丙酸类的绿原酸具有明显抗菌作用, 对消化系统、血液系统和生殖系统均有药理作用. 目前, 由于新药开发以及功效的确认, 市场对杜仲的有效活性成分的需求量极大, 但生态环境却因过度开采遭到严重的破坏, 野生杜仲资源数量急剧减少^[2]. 为了更好地解决这一矛盾, 筛选杜仲愈伤组织的有效活性成分高产克隆系并开展其活性成分积累调控机理研究势在必行. 本文利用转录组高通量测序技术就白光对杜仲愈伤组织中绿原酸产量的影响进行分析, 为获取高产绿原酸资源提供理论依据.

1 材料与方法

1.1 实验材料

取杜仲幼苗嫩叶表面消毒后, 接种在培养基 MS + 1.0 mg · mL⁻¹ 2,4-D + 1.0 mg · mL⁻¹ NAA + 1.0 mg · mL⁻¹ 6-BA 诱导愈伤组织. 20 d 后长出愈伤组织. 然后把杜仲愈伤组织接种在培养基 B₅ + 0.5 mg · mL⁻¹ NAA + 1.0 mg · mL⁻¹ 6-BA, 设计 2 种培养条件: 1) 光照培养, 光强 12 000 lux, 16 h 光照, 8 h 黑暗; 2) 24 h 黑暗, 作为对照. 2 种培养条件的温度均为 (26 ± 1) °C. 愈伤组织在这 2 种条件下培养 18 d 后, 用液氮迅速冷冻, 然后放到 -80 °C 保存备用.

1.2 实验方法

把冻存的杜仲愈伤组织送至北京百迈客生物科技有限公司进行测序和分析.

收稿日期: 2015-06-24

基金项目: 国家自然科学基金(31100396)和江西省亚热带植物资源保护和利用重点实验室开放基金(YRD201404)资助项目.

作者简介: 张俊娥(1972-), 女, 山西榆社人, 副教授, 博士, 主要从事植物生物技术与遗传育种研究.

1.2.1 RNA 样品提取和检测 采用艾德莱试剂盒即“RN40-EASYspin 植物 microRNA 快速提取试剂盒”提取杜仲愈伤组织的 RNA. RNA 样品的浓度、纯度和总量利用 nanodrop 2000(Thermo 基因有限公司) 进行检测. 样品的完整性利用 Agilent 2100 进行检测.

1.2.2 cDNA 文库构建和质控 文库制备采用 NEB 试剂盒. 用带有 Oligo(dT) 的磁珠富集真核生物 mRNA; 加入 Fragmentation Buffer 将 mRNA 进行随机打断; 以 mRNA 为模板 ,用 6 碱基随机引物合成第 1 条 cDNA 链 ,然后加入缓冲液、dNTPs、RNase H 和 DNA Polymerase I 合成第 2 条 cDNA 链 ,利用 AMPure XP beads 纯化 cDNA; 纯化的双链 cDNA 再进行末端修复、加 A 尾并连接测序接头 ,然后用 AMPure XP beads 进行片段大小选择; 最后通过 PCR 富集得到 cDNA 文库. 分别使用 Qubit 2. 0 和 Agilent 2100 对文库的浓度和插入片段大小进行检测 ,使用 Q-PCR 方法对文库的有效浓度进行准确定量 ,以保证文库质量.

1.2.3 上机测序 利用 Illumina HiSeq™ 2500 平台进行高通量测序 ,测序读长为 PE100.

1.2.4 数据过滤 截除 Reads 中的测序接头以及引物序列; 过滤低质量值数据 ,确保数据质量.

1.2.5 数据组装 采用 Trinity^[3] 软件组装.

1.2.6 差异表达分析 使用 EBSeg^[4] 进行差异表达分析 ,获得 2 个样品之间的差异表达基因集. 在差异表达分析过程中采用了公认有效的 Benjamini-Hochberg 方法对原有假设检验得到统计学意义上的显著性 p 值(p -value) 进行校正 ,并最终采用校正后的 p 值 ,即 FDR (False Discovery Rate) 作为差异表达基因筛选的关键指标 ,以降低对大量基因的表达值进行独立的统计假设检验带来的假阳性. 在筛选过程中 ,将 $FDR < 0. 01$ 且差异倍数 FC (Fold Change) ≥ 2 作为筛选标准. 其中 FC 表示 2 样品(组) 间表达量的比值.

2 结果与分析

2.1 组装结果统计

经过测序质量控制 ,共得到 13.00 Gb Clean Data ,各样品 Q_{30} 碱基百分比均不小于 88. 11% . 白光条件(编号: T01) 下 ,CG 含量为 47. 85% ; 黑暗条件(编号: T02) 下 ,CG 含量为 48. 11% . 利用 Trinity 软件进行合并组装共获得到 139 014 条 Transcripts 和 62 030 条 Unigenes ,Transcript 和 Unigene 的 N_{50} 分别为 1 920 和 1 372 ,组装完整性较高. 具体组装信息如表 1.

表 1 组装结果统计表

长度范围/bp	Contig	Transcript	Unigene
200 ~ 300	6 957 978(99. 33%)	31 307(22. 52%)	23 896(38. 52%)
300 ~ 500	19 541(0. 28%)	24 708(17. 77%)	14 818(23. 89%)
500 ~ 1000	12 933(0. 18%)	26 987(19. 41%)	9 983(16. 09%)
1 000 ~ 2 000	9 237(0. 13%)	31 629(22. 75%)	8 186(13. 20%)
> 2 000	5 085(0. 07%)	24 383(17. 54%)	5 147(8. 30%)
总数	7 004 774	139 014	62 030
总长度	311 669 071	157 794 421	45 678 674
N_{50} 长度	44	1 920	1 372
平均长度	44. 49	1 135. 10	736. 40

2.2 Unigene 功能注释

使用 BLAST^[5] 软件将组装获得的 62 030 条 Unigene 序列与 Nr (Non-redundant protein data-base) ^[6]、Swiss-Prot(Annotated protein sequence data-base) ^[7]、GO(Gene ontology) ^[8]、COG(Clusters of orthologous groups of protein) ^[9]、KOG(euKaryotic orthologous groups of protein) ^[10]、KEGG(Kyoto Encyclopedia of genes and genomes) ^[11] 数据库比对 ,预测完 Unigene 的氨基酸序列之后使用 HMMER^[12] 软件

与 Pfam(Protein family) ^[13] 数据库比对 ,获得 Unigene 的注释信息. 通过 BLASTX 比对 ,共获得 25 417 (40. 98%) 个有注释信息的 Unigenes. 其中 ,7 620 条 Unigene 注释到 COG ,占 12. 28% ; 14 857 条 Unigene 注释到 GO ,占 23. 95% ; 4 794 条 Unigene 注释到 KEGG ,占 7. 73% ; 14 140 条 Unigene 注释到 KOG ,占 22. 80% ; 16 174 条 Unigene 注释到 Pfam ,占 26. 07% ; 16 297 条 Unigene 注释到 Swiss-prot ,占 26. 24% ; 25 167 条 Unigene 注释到 Nr ,占 40. 57% . 具体注释信息见表 2.

表2 Unigene 注释信息

注释数据库	被注释的 基因个数	长度 ≥ 300 bp 的基因个数	长度 $\geq 1\,000$ bp 的基因个数
COG 注释	7 620	7 127	5 339
GO 注释	14 857	12 586	7 489
KEGG 注释	4 794	4 361	2 979
KOG 注释	14 140	12 069	7 715
Pfam 注释	16 174	14 988	10 557
Swiss-Prot 注释	16 297	14 351	9 081
nr 注释	25 167	21 180	12 349
All 注释	25 417	21 320	12 368

2.3 差异表达分析

采用 EBSeg^[14] 软件进行差异表达分析, 获得 2 个样品之间的差异表达基因集. 结果表明, 共获得差异表达基因 1 986 个, 其中在白光条件下上调表达的基因有 1 139 个, 下调表达的基因 847 个. 图 1 直观地呈现了杜仲愈伤组织 2 个样品的基因表达量总体分布.

2.4 与绿原酸合成相关的代谢途径

从图 2 可以看出, 与绿原酸合成相关的酶有 3 种, 分别为苯丙氨酸解氨酶 (EC 4.3.1.24, phenylalanine ammonia-lyase, PAL)、肉桂酸 4-羟基化酶 (EC 1.14.13.11, trans-cinnamate 4-monooxygenase, C4H)、奎宁酸羟基肉桂酰转移酶 (EC 2.3.1.133, Shikimate o-hydroxyl-cinnamoyltransferase, HCT). 这 3

种酶基因均上调表达. PAL 是杜仲绿原酸合成途径中上游的关键酶, PAL 在白光条件下比黑暗条件下

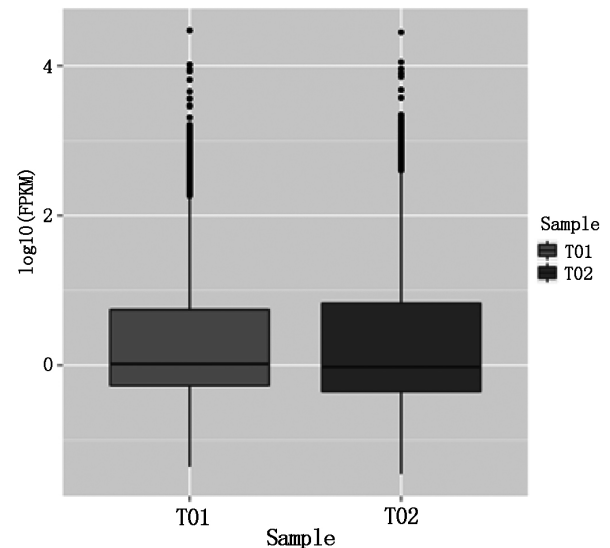


图1 杜仲愈伤组织 2 个样品基因表达量的 FPKM 箱线图

上调表达 3.94 倍, C4H 在白光条件下比黑暗条件下上调表达 2.10 倍, HCT 在白光条件下比黑暗条件下上调表达 2.20 倍 (见图 3). 结果表明, 白光培养杜仲愈伤组织 18 d, 其绿原酸积累能力增强. 由此可见, 白光 (光强为 12 000 lux, 16 h 光照 8 h 黑暗) 可以明显增强杜仲愈伤组织中绿原酸的含量.

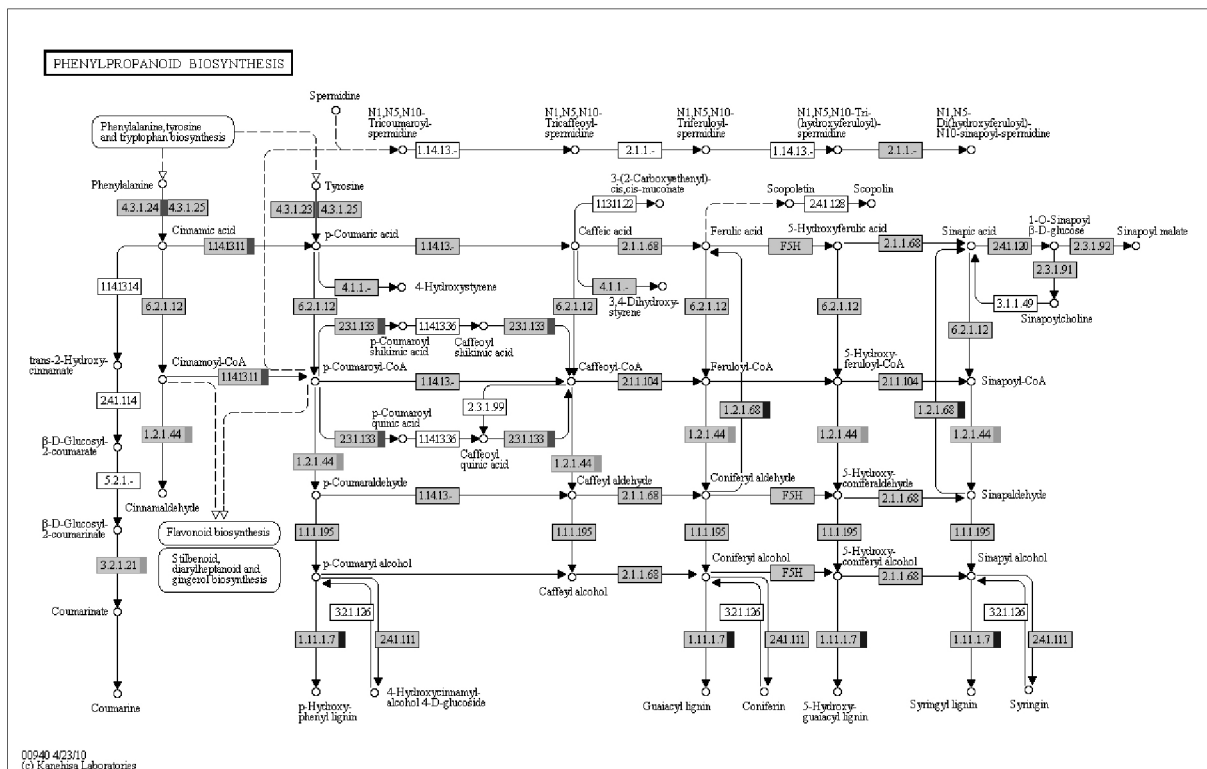


图2 绿原酸合成相关的代谢途径

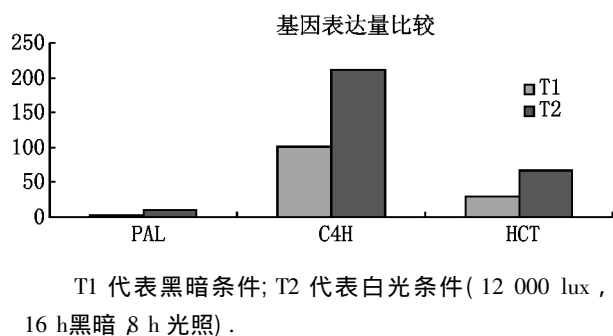


图3 光暗培养的杜仲愈伤组织中与绿原酸合成相关的3种酶基因表达量比较

3 讨论

高通量测序技术具有分析结果快、准确、信息全、灵敏度高和自动化的特点。自2009年以来,我国研究人员陆续对不同的中药材进行了高通量转录组测序和分析。先后对青蒿^[14]、西洋参^[15]、甘草^[16]、丹参^[17]、人参^[18]、虎杖^[19]、红花^[20]、金银花^[21]、杜仲的果实和叶片^[22]等中药材进行了转录组分析,这些研究可以从基因层面上阐述中药有效成分的代谢通路,发掘得到新的功能基因;通过转录水平的调节,提高药用成分的产量与活性,为高通量发掘新型活性药物成分奠定了坚实的基础。

李铁柱等^[23-24]通过对杜仲果实和叶片转录组分析表明,其绿原酸的合成途径共涉及4种酶基因,分别为苯丙氨酸脱氢酶(PAL)、肉桂酸4-羟化酶(C4H)、香豆酰-CoA连接酶(4CL)和莽草酸羟基肉桂酰转移酶(HCT),它们在幼果和叶片中表达量各有差异。本文利用Illumina HiSeq™ 2500平台对光暗培养的杜仲愈伤组织进行了转录组高通量测序,通过对KEGG通路进行深入分析,结果表明,白光(光强为12 000 lux,16 h光照,8 h黑暗)培养杜仲愈伤组织18 d,与绿原酸合成相关的3种酶基因上调表达,它们分别为苯丙氨酸解氨酶(EC 4.3.1.24, PAL)、肉桂酸4-羟化酶(EC 1.14.13.11, C4H)、奎宁酸羟基肉桂酰转移酶(EC 2.3.1.133, HCT)。该结果表明:白光(光强为12 000 lux,16 h光照,8 h黑暗)培养杜仲愈伤组织18 d,其绿原酸生产能力提高。本研究为获取高产绿原酸资源和杜仲遗传分析提供有价值的资料。

4 参考文献

[1] 张康健,马希汉. 杜仲次生代谢物与人类健康[M]. 咸阳:西北农林科技大学出版社,2009.

- [2] Yao Xiaohong, Deng Jianyun, Huang Hongwen. Genetic diversity in *Eucommia ulmoides* (Eucommiaceae), an endangered traditional Chinese medicinal plant [J]. Conservation Genetics 2012, 13(6): 1499-1507.
- [3] 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.
- [4] Leng Ning, Dawson J A, Thomson J A, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments [J]. Bioinformatics 2013, 29(8): 1035-1043.
- [5] Altschul S F, Madden T L, Schäffer A A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs [J]. Nucleic Acids Res, 1997, 25(17): 3389-3402.
- [6] Deng Yangyang, Li Jianqi, Wu Songfeng, et al. Integrated nr database in protein annotation system and its localization [J]. Computer Engineering 2006, 32(5): 71-74.
- [7] Apweiler R, Amos Bairoch A, Wu Cathy H, et al. UniProt: the universal protein knowledgebase [J]. Nucleic Acids Res 2004, 32 (Database issue): D115-D119.
- [8] Ashburner M, Ball C A, Blake J A, et al. Gene ontology: tool for the unification of biology [J]. Nat Genet 2000, 25(1): 25-29.
- [9] Tatusov R L, Galperin M Y, Natale D A, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution [J]. Nucleic Acids Res 2000, 28(1): 33-36.
- [10] Koonin E V, Fedorova N D, Jackson J D, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes [J]. Genome Biol, 2004, 5(2): R7.
- [11] Kanehisa M, Goto S, Kawashima S, et al. The KEGG resource for deciphering the genome [J]. Nucleic Acids Res 2004, 32 (Database issue): D277-D280.
- [12] Finn R D, Bateman A, Clements J, et al. Pfam: the protein families database [J]. Nucleic Acids Res 2014, 42 (Database issue): D222-D230.
- [13] Eddy S R. Profile hidden Markov models [J]. Bioinformatics, 1998, 14(9): 755-763.
- [14] Wang Wei, Wang Yejun, Zhang Qing, et al. Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing [J]. BMC Genomics, 2009, 10(1): 465.
- [15] Sun Chao, Li Ying, Wu Qing, 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 Ge-

- nomics 2010 ,11(1) : 262.
- [16] Li Ying ,Luo Hongmei ,Sun Chao ,et al. EST analysis reveals putative genes involved in glycyrrhizin biosynthesis [J]. BMC Genomics 2010 ,11(1) : 268.
- [17] 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.
- [18] Chen Shilin ,Luo Hongmei ,Li Ying ,et al. 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng* [J]. Plant Cell Reports 2011 ,30(9) : 1593-1601.
- [19] Hao Dacheng ,Ma Pei ,Mu Jun ,et al. De novo characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum* [J]. Science China Life Sciences 2012 ,55(5) : 452-466.
- [20] Huang Lulin ,Yang Xiao ,Sun Pei ,et al. The first Illumina-based de novo transcriptome sequencing and analysis of safflower flowers [J]. Plos One 2012 ,7(6) : e38653.
- [21] Yuan Yuan ,Song Lipu ,Li Minhui ,et al. Genetic variation and metabolic pathway intricacy govern the active compound content and quality of the Chinese medicinal plant *Lonicera japonica* thumb [J]. BMC Genomics 2012 ,13(1) : 195.
- [22] 李铁柱 杜红岩 刘慧敏 等. 杜仲果实和叶片转录组数据组装及基因功能注释 [J]. 中南林业科技大学学报 , 2012 ,32(11) : 122-130.
- [23] 李铁柱 杜红岩 刘慧敏 等. 杜仲幼果和成熟果实转录组数据组装及基因功能注释 [J]. 中南林业科技大学学报 2012 ,32(10) : 9-17.
- [24] 李铁柱 杜红岩 朱高浦. 杜仲绿原酸生物合成途径相关基因的差异表达 [J]. 经济林研究 2013 ,31(4) : 32-38.
- [25] 万倩 钟政荣 刘志刚. 粉尘螨 Der f14 基因克隆与分子特征的分析 [J]. 江西师范大学学报: 自然科学版 , 2015 ,39(2) : 166-170.

The Effects of White Light on Chlorogenic Acid Contents in *Eucommia Ulmoides Calli* Based on Transcriptome High-Throughput Sequencing

ZHANG June¹ ,DENG Huafeng² ,ZHENG Wenqiang¹ ,JIANG Jing¹ ,JIN Cong¹ ,ZHAO Fukai¹

(1. College of Life Science ,Key Laboratory of Protection and Utilization of Subtropical Plant Resources of Jiangxi Province ,
Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China;

2. School of Computer and Information Engineering ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China)

Abstract: In this study ,*Eucommia ulmoides calli* for 18d under white light (light intensity was 12 000 lux ,16 h light and 8 h darkness) were cultured and 24 h darkness culture was as control. Then transcriptome high-throughput sequencing for the *Eucommia ulmoides calli* were performed using the Illumina HiSeqTM 2500 sequencing platform. Functional annotation and enrichment analysis of differentially expressed genes were also carried out. Effect of white light on the content of chlorogenic acid of *Eucommia ulmoides calli* was studied. The results showed de novo assembly generated 62 030 unigenes. 25 417 (40. 98%) annotated unigenes were obtained by selecting BLAST parameter (E -value $\leq 10^{-5}$) and HMMER parameter (E -value $\leq 10^{-10}$). Metabolic pathway analysis revealed that 3 unigenes were predicted to be responsible for the chlorogenic acid ,the 3 unigenes were up-regulated and in the calli under light culture ,which was meaning that the production capacity of chlorogenic acid increased. The three enzymes dominated chlorogenic acid production in *Eucommia ulmoides calli*. The three enzymes were phenylalanine ammonia-lyase (EC 4. 3. 1. 24 ,PAL) ,trans-cinnamate 4-monooxygenase (EC 1. 14. 13. 11 ,C4H) and Shikimate *o*-hydroxyl-cinnamoyltransferase (EC 2. 3. 1. 133 ,HCT) . The study showed that the white light could improve the production capacity of chlorogenic acid components of *Eucommia ulmoides calli* and represents a valuable resource for future genomic studies on *Eucommia ulmoides*.

Key words: *Eucommia ulmoides* oliv. ; callus; white light; chlorogenic acid; transcriptome high-throughput sequencing

(责任编辑: 刘显亮)