

• 药材与资源 •

基于全长转录组挖掘八角果实中木脂素生物合成关键基因

李金梅^{1,2}, 覃艳红^{1,2}, 黄 鼎^{1,2}, 陈建桦^{1,2}, 明如宏^{1,2}, 姚绍嫦^{1,2}, 李良波^{1,2}, 谭 勇^{1,2*}, 黄荣韶^{1,2*}

1. 广西中医药大学药学院, 广西 南宁 530200

2. 广西中医药大学 广西壮瑶药重点实验室, 广西 南宁 530200

摘要: **目的** 通过分析八角 *Illicium verum* 不同组织部位的基因转录水平差异, 挖掘八角果实中木脂素生物合成途径相关酶基因, 为八角木脂素生物合成途径关键酶基因研究提供参考。 **方法** 利用 PacBio 和 Illumina 测序平台对八角的茎、嫩叶、老叶、花及果实进行全长转录组测序与比较转录组学分析, 通过 WGCNA 分析挖掘具有组织特异表达的木脂素生物合成途径关键酶基因, 利用 qRT-PCR 对候选基因的表达情况进行验证。 **结果** 经过纠错去除冗余后共获得 32 996 条高质量转录本, 平均长度 2 033 bp, N₅₀ 长度 2 343 bp, 其中 31 248 条在公共数据库获得注释。共鉴定到 39 条参与木脂素生物合成的转录本, 涉及 10 个关键酶基因, 其中在八角果实中高表达的有 14 条转录本, 涉及 6 个酶基因。八角果实中与其他组织相比, 差异表达的基因富集在苯丙烷代谢途径中。通过 WGCNA 分析鉴定到 15 个共表达模块, 在果实特异表达模块 (MEred) 中挖掘到木脂素合成关键酶基因 4CL、CCR 共 3 条转录本。候选基因 PAL (transcript11647), C4H (transcript26532), 4CL (transcript12186), CCR (transcript27851) 和 CAD (transcript27743) 在不同组织中的表达情况与转录组数据一致。 **结论** 为八角分子生药学研究提供了可靠的转录组数据库, 也为八角果实中木脂素生物合成途径关键酶基因研究提供了研究基础。

关键词: 八角; 全长转录组; 木脂素; 生物合成; 加权基因共表达网络分析**中图分类号:** R286.12 **文献标志码:** A **文章编号:** 0253-2670(2024)23-8153-09**DOI:** 10.7501/j.issn.0253-2670.2024.23.021Full-length transcriptome analysis reveals key genes involved in lignan biosynthesis in fruits of *Illicium verum*LI Jinmei^{1,2}, QIN Yanhong^{1,2}, HUANG Ding^{1,2}, CHEN Jianhua^{1,2}, MING Ruhong^{1,2}, YAO Shaochang^{1,2}, LI Liangbo^{1,2}, TAN Yong^{1,2}, HUANG Rongshao^{1,2}

1. Institute of Pharmacy, Guangxi University of Chinese Medicine, Nanning 530200, China

2. Guangxi Key Laboratory of Zhuang and Yao Ethnic Medicine, Guangxi University of Chinese Medicine, Nanning 530200, China

Abstract: Objective The differences in the gene expression of *Illicium verum* in different tissues and enzyme genes of lignan biosynthesis pathway in fruits of *I. verum* were analyzed to provide a basis for the study of key enzyme genes in the biosynthetic pathway of lignin in *I. verum*. **Methods** PacBio and Illumina sequencing platforms were used to perform full-length transcriptome sequencing of stems, tender leaves, mature leaves, flowers and fruits of *I. verum*, and comparative transcriptome analysis of different tissues were carried out. Key enzyme genes of lignan biosynthesis pathway with tissue-specific expression were excavated by weighted gene co-expression network analysis (WGCNA), and the expression of candidate genes was verified by qRT-PCR. **Results** A total of 32 996 high-quality transcripts were obtained after error correction and redundancy removal, with an average length of 2 033 bp and an N₅₀ length of 2 343 bp, of which 31 248 were annotated in the public database. A total of 39 transcripts involved in lignan biosynthesis were identified, involving 10 key enzyme genes. Among them, 14 transcripts were highly expressed in the fruits of *I. verum*, involving six enzyme genes. Compared with other tissues, the differentially expressed genes in fruits of *I. verum* were enriched in the phenylpropanoid metabolic pathway. A total of 15 co-expression modules were identified by WGCNA, among which three transcripts

收稿日期: 2024-08-23**基金项目:** 广西科技重大专项项目 (桂科 AA22096029); 广西创新驱动发展专项资金项目 (桂科 AA23023035); 广西中医药大学-柳药集团青年科技创新能力提升计划专项项目 (GL-2023-01)**作者简介:** 李金梅 (1998—), 女, 在读硕士, 从事中药资源学研究。E-mail: 2281953548@qq.com***通信作者:** 谭 勇, 博士, 教授, 硕士生导师, 主要从事中药资源保护和中药种植研究。E-mail: xjty321@163.com

黄荣韶, 博士, 教授, 博士生导师, 主要从事中药资源研究。E-mail: hrshao802@163.com

of the key enzyme gene of lignan biosynthesis pathway (4CL and CCR) were excavated in the fruit-specific expression module (MEred). The expression of candidate genes *PAL* (transcript11647), *C4H* (transcript26532), *4CL* (transcript12186), *CCR* (transcript27851) and *CAD* (transcript27743) in different tissues was consistent with the transcriptome data. **Conclusion** Our study provides a reliable transcriptome database for the study of molecular pharmacognosy of *I. verum*, and also provides a research foundation for the investigation of key enzyme genes of lignan biosynthesis in *I. verum* fruits.

Key words: *Illicium verum* Hook. f.; full-length transcriptome; lignan; biosynthesis; weighted gene co-expression network analysis

八角 *Illicium verum* Hook. f. 为木兰科多年生木本植物，以干燥成熟果实入药，具有温阳散寒，理气止痛的功效^[1]。作为“桂十味”地道药材，八角的种植和加工是我国广西许多地区富民强县的特色支柱产业之一。目前广西八角种植面积 36.7 万公顷，年产量高达 18 万 t，占全国八角总产量 85% 和世界八角总产量 70% 以上^[2-3]。现代研究表明八角果实中含有多种活性成分，包括挥发油、萜类、黄酮类、木脂素、有机酸等化合物^[4-5]。木脂素作为八角果实的主要活性成分之一，是一类重要的可供药用的植物次生代谢产物，具有抗病毒、抗肿瘤、抗菌和抗氧化等活性^[6]。曾武^[7]发现八角乙醇提取物中三苯基-倍半木脂素具有良好的抗菌活性。叶凤梅^[8]从八角筛选出的 3 个木脂素化合物 GS-4、GS-20、GS-45 均对细菌和肿瘤细胞的生长表现出不同程度的抑制活性。目前，关于八角的研究主要集中在化学成分和药理活性方面，而对其活性成分生物合成及其调控方面的研究则相对较少。

植物中木脂素的生物合成途径一般分为以下 3 个阶段^[9]。第 1 阶段是对香豆酰辅酶 A 的合成：在苯丙氨酸转氨酶（phenylalanine ammonia-lyase, PAL）的催化作用下将苯丙氨酸转化为肉桂酸，接着在肉桂酸 4-羟化酶（cinnamic acid 4-hydroxylase, C4H）的催化作用下，将反式肉桂酸转化成对香豆酸；最后在 4-香豆酰辅酶 A 连接酶（4-coumarate CoA ligase, 4CL）的催化作用下生成对香豆酰辅酶 A；在第 2 个阶段是木脂素类化合物的前体松柏醇的合成：莽草酸/奎宁酸羟基肉桂酰转移酶（shikimate/quinate hydroxycinnamoyl transferase, HCT）及对香豆酸 3-羟化酶（p-coumarate 3-hydroxylase, C3H）作用于香豆酸生成咖啡酰辅酶 A；接着在咖啡酰辅酶 A 氧甲基转移酶（caffeoyl-CoA O-methyltransferase, CCoAOMT）的作用下产生阿魏酰辅酶 A；肉桂酰辅酶 A 还原酶（cinnamoyl-CoA reductase, CCR）作用于阿魏酰辅酶 A 产生松柏醛；肉桂醇脱氢酶（cinnamyl alcohol

dehydrogenase, CAD）催化松柏醛脱氢转变为松柏醇；第 3 个阶段由松柏醇合成结构多样的木脂素单体：松柏醇通过过氧化物酶（peroxidases, POD）氧化聚合形成木脂素多聚物，而松柏醇经聚合蛋白酶（dirigent, DIR）作用生成松脂醇，松脂醇再经过松脂醇还原酶（pinoresinol-lariciresinol reductase, PLR）催化生成落叶松脂素和开环异落叶松脂素，最后再经过羟基化、甲基化、糖基化等一系列催化反应形成一系列的木脂素单体物质。

转录组测序技术是一种快速发现和鉴定表达基因的方法。二代测序技术（next-generation sequencing, NGS）已经广泛应用于没有参考基因组的非模式物种的次生代谢生物合成途径解析以及相关基因表达分析的研究中。通过利用 NGS 技术，学者们分析了一系列药用植物有效成分的生物合成途径，并成功鉴定了相关次生代谢途径中的关键酶基因，如槐花中黄酮类物质^[10]、朱砂根中三萜皂苷^[11]、钩藤中钩藤碱^[12]及苍术根中倍半萜^[13]的生物合成途径分析及关键酶的挖掘与鉴定等。鉴于 NGS 技术获得的序列读取长度短，导致拼接出来的序列组装完整度较低以及一些重要信息丢失，基因预测的完整性和准确性也受到了限制，进而影响了非模式物种相关基因的深入挖掘以及功能鉴定工作。随着测序技术的进步，基于三代测序平台可以直接获取转录本的高质量全长序列，刚好可以弥补 NGS 技术的缺点，为无参考基因组的非模型植物提供高质量的参考序列。

目前八角基因组信息仍未破解，全长转录组数据库也未见报道，制约了其分子药学相关研究的发展，且关于八角木脂素生物合成途径仍知之甚少。本研究基于 PacBio 三代测序平台构建了八角高质量全长转录本数据库，通过对八角茎、嫩叶、老叶、花、果实 5 个不同组织的比较转录组学分析，结合 WGCNA 挖掘八角果实中木脂素生物合成途径关键酶基因，并通过 qRT-PCR 对其表达水平进行验证，为揭示八角果实中木脂素类成分生物合成的

分子机制奠定基础。

1 材料

本研究试验材料八角种植于桂平市香山八角林场(22.93°N, 110.25°E)。经广西中医药大学黄荣韶教授鉴定为八角 *I. verum* Hook. f.。随机选取 3 株八角, 采集的茎、嫩叶、老叶、花、果实, 材料采摘后立即洗净、除杂, 并立即放于液氮中速冻, 置-80℃超低温冰箱中保存, 用于 RNA 的提取。

2 方法

2.1 RNA 提取和测序

使用艾德莱 EASYspin Plus 多糖多酚复杂植物 RNA 快速提取试剂盒(RN51)分别提取八角茎、嫩叶、老叶、花、果实的总 RNA, 按说明书进行操作, 再用 1%的电泳凝胶检验提取 RNA 的完整性。基于 PacBio 和 Illumina 测序平台, 全长转录组测序部分由武汉贝纳科技有限公司完成。

2.2 数据分析

为确保基因序列重新组装和后续分析的准确性, 将原始读段、低质量读段、模糊读段和重复序列在组装前删除。先统计原始数据, 再进行数据过滤和质量评估, 然后对 clean reads 进行转录本拼接, 结合软件 BLAST ($E\text{-value} \leq 1 \times 10^{-5}$) 将 unigene 与 Nr、KOG、Pfam、Swiss-Prot、KEGG、GO 等数据库进行基因功能注释、KEGG 代谢通路分析等。依据转录组注释以及基因表达分析结果, 挖掘其中木脂素类生物合成途径相关基因进行进一步分析。运用 TBtools 绘制基因不同组织相对表达热图, 分析候选基因在八角不同组织部位的表达模式。

2.3 加权基因共表达网络分析 (weighted gene co-expression network analysis, WGCNA)

使用 R 软件 WGCNA 包进行加权基因共表达网络分析, 将所有转录本的 RPKM 值输入后进行分析。具体参数如下: 软阈值设置为 12, 最小模块大小设置为 30, 合并阈值为类似的模块设置为 0.35。根据动态聚类方法将具有相同表达模式的基因分类为一个模块。

2.4 qRT-PCR 验证差异表达基因 (DEGs)

利用 Primer Premier5.0 软件设计的基因特异性引物进行 qRT-PCR 验证, 由南宁捷尼斯生物科技有限公司合成(表 1)。利用 TaKaRa PrimeScript™ RT reagent Kit (AM81775A) 试剂盒将从八角茎、嫩叶、老叶、花、果实等不同组织样品中提取的 RNA 反转录成 cDNA。采用 ChamQ Universal SYBR qPCR Master Mix 试剂盒, 在荧光定量 PCR 仪 (Bio-Rad CFX96™ Real-time System) 上进行 qRT-PCR 检测基因表达; 采用 *Actin* 作为内参基因, 每个样品 3 个生物学重复, 采用 $2^{-\Delta\Delta C_t}$ 法计算 DEGs 的相对表达量。

3 结果与分析

3.1 八角全长转录组组装与质量分析

通过 PacBio Sequel II 测序平台对八角的茎、嫩叶、老叶、花、果 5 个组织的混合样品进行测序, 共获得 619 750 条环形一致性序列 (Circular consensus sequencing, CCS), 平均长度为 2 149 bp。将全长非嵌合序列 (full-length non-concatemer, FLNC) 转录组数据经过去冗余纠错后共获得 32 996 条高质量全长转录本序列, 平均长度 2 033 bp, N50 长度 2 343 bp。采用 BUSCO 软件对全长转录本序

表 1 qRT-PCR 引物序列
Table 1 Primers for qRT-PCR

基因	引物名称	引物序列 (5'-3')
PAL	transcript11647-F	TGAACAAGGGCACGGACAG
	transcript11647-R	CTTGCTTGGTTCTCCTATGCG
C4H	transcript26532-F	AGCCAAAGTAGAGGCCAACG
	transcript26532-R	TCCAAGTTCTGCACCAACCTG
4CL	transcript12186-F	GTGAAATCGTGAATGGGAAGC
	transcript12186-R	CGGCTCAACACCACAACCTCT
CCR	transcript27851-F	AAGAACTGGTACTGTTACGGGAAAG
	transcript27851-R	CTAGCATTGATCGTTGGCTGTAG
CAD	transcript27743-F	CAGCACAAGACTCAACAGGGAT
	transcript27743-R	CAACTCCACAACCTCCCCTACAA
Actin	actin-F	TGAGACCTTCAATGTCCCTGC
	actin-R	CACCATCACCAGAATCCAGGA

列进行质量完整性评估, 其完整性达到了 75.1%。为了进一步评估八角转录本的质量, 将得到的全长转录本序列作为参考序列, 把八角各组织样品的 clean reads 在参考序列上做 mapping 分析, 各组织样品的比对率都在 78%~82%。综上表明本研究获得的八角全长转录本序列质量较高、数据可靠, 可以满足后续候选基因筛选分析要求。

3.2 八角基因功能注释和功能分类

为获得八角全长转录本序列全面的功能注释信息, 将八角所有的转录本与非冗余蛋白质库(non-redundant protein sequence database, NR)、蛋白综合数据库(universal protein resource, Uniprot)、基因本体论(gene ontology, GO)、真核同源群数据库(eukaryotic orthologous groups of protein, KOG)、Pathway、蛋白质家族数据库(The protein families database, Pfam)、京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG) 7 大数据库比对, 获得注释信息。结果表明, 共有 31 248 条全长转录本得到注释, 占全部转录本的 94.70% (表 2), 其中在 Uniprot 和 NR 数据库中注释得到的转录本数量最多, 分别为 31 063 和 31 021 条转录本, 占比为 94.14%和 94.01%, 而在 KOG 数据

库中注释到的转录本数量最少, 共注释到 6 600 条转录本序列, 占比为 20.00%。根据 Nr 数据库的比对结果, 统计比对率最高的前 10 个物种, 其中与八角同源性序列比较高的物种有莲 *Nelumbo nucifera* L. (5 672, 26.51%)、沉水樟 *Cinnamomum micranthum* f. *kanehirae* L. (5 284, 24.7%)、博落回 *Macleaya cordata* (2 781, 13%)、无油樟 *Amborella trichopoda* L. (2 666, 12.46%), (图 1)。

表 2 八角转录本的功能注释结果

Table 2 Functional annotation results of transcripts in *Illicium verum*

注释数据库	被注释的转录本数	占比/%
Swiss-Prot	31 063	94.14
Nr	31 021	94.01
Pfam	28 804	87.30
GO	25 508	77.31
KEGG	17 808	53.97
Pathway	10 554	31.99
KOG	6 600	20.00
总的被注释转录本数	31 248	94.70
总的转录本数	32 996	100.00

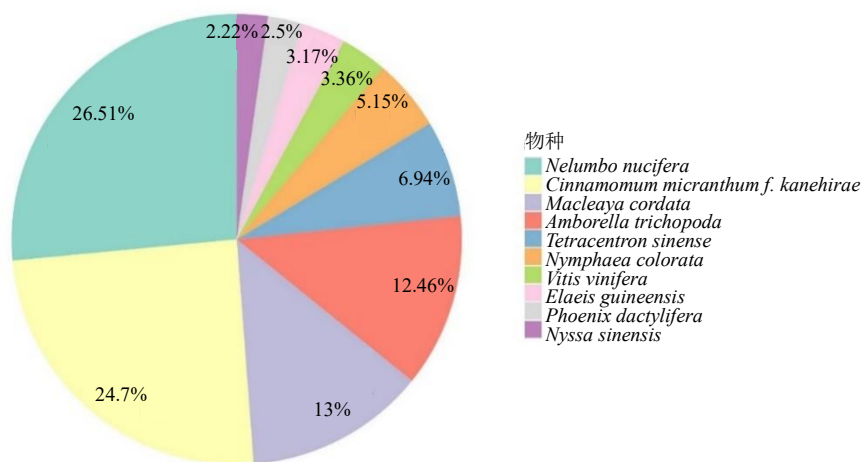


图 1 NR 库中的主要物种分布

Fig. 1 Distribution of main species in NR database

3.3 八角不同组织差异表达基因分析

鉴于八角的主要药用部位为果实, 因此本研究将八角果实与其他 4 个组织分别进行比较转录组学分析。利用表达定量得到的转录本在各个样品中的表达量 reads count 数据, 进行八角不同组织间基因差异表达分析, 筛选阈值为 $\text{padj} < 0.05$ 且

$\log_2[\text{Fold Change}] \geq 1$ 。比较转录组学分析结果表明, 八角的茎与果实相比, 差异表达基因总数为 7 128 个, 其中 4 013 个基因表达上调, 3 115 个基因表达下调; 花与果实相比, 差异表达基因总数为 10 352 个, 其中 5 225 个基因表达上调, 5 127 个基因表达下调; 嫩叶与果实相比, 差异表达基因总数为 12 036

个, 其中 5 659 个基因表达上调, 6 377 个基因表达下调; 老叶与果实相比, 差异表达基因总数为 12 590 个, 其中 5 645 个基因表达上调, 6 945 个基因表达下调 (图 2-A)。其中, 在这 4 个比较组中, 共同的差异基因共有 1 795 个, KEGG 富集分析结果显示, 这些差异基因主要富集在次生代谢途径 (biosynthesis of other secondary metabolites), 包括

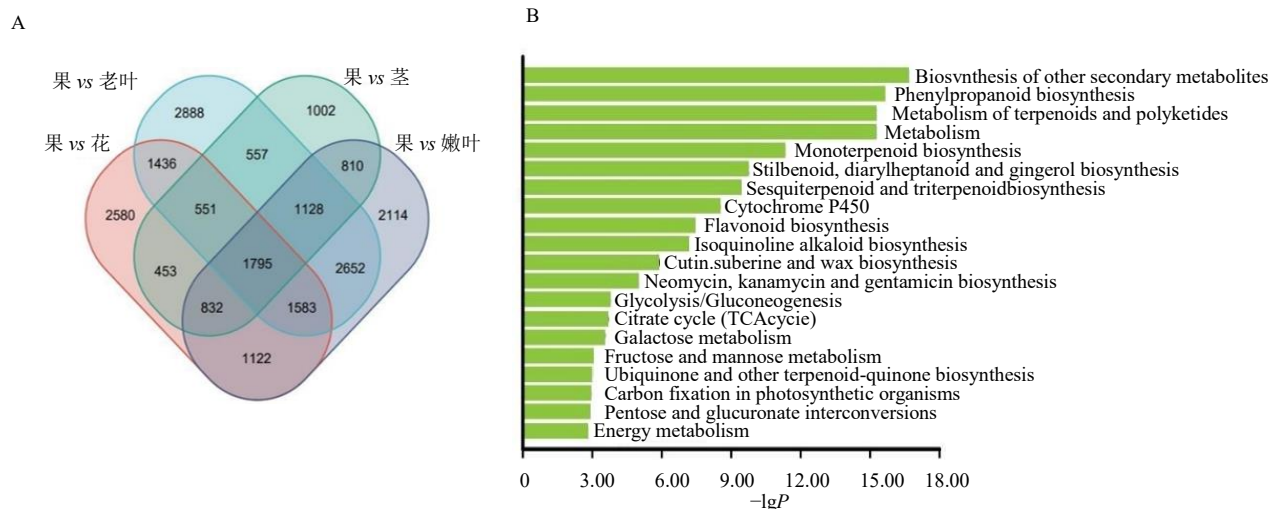


图 2 八角 5 个组织中的差异基因韦恩图 (A) 及 KEGG 富集分析 (B)

Fig. 2 Venn diagram (A) and KEGG enrichment analysis (B) of differential genes in five organs of *I. verum*

3.4 木脂素类物质生物合成相关基因挖掘与分析

鉴于植物中木脂素的生物合成途径目前已经研究得比较清楚, 本研究根据八角全长转录本数据库的基因注释结果, 总共从数据库中鉴定到 39 条与木脂素生物合成相关的转录本, 涉及 10 个关键酶基因, 其中包括 *PAL* (2 条转录本), *C4H* (1 条转录本), *4CL* (5 条转录本), *CHT* (2 条转录本), *C3H* (1 条转录本), *CCoAMT* (2 条转录本), *CCR* (13 条转录本), *CAD* (4 条转录本), *DIR* (3 条转录本) 以及 *PLR* (6 条转录本), 见表 3。

为了进一步分析上述关键酶基因在八角不同组织的表达水平, 将上述 39 条转录本的 FPKM 值代表基因的相对表达水平, 绘制不同组织部位的基因表达量热图 (图 3)。如图 3 所示, 木脂素生物合成相关的关键酶基因有多种表达模式: 在某个组织中高表达, 例如 *PAL* (Transcript10678)、*4CL* (Transcript17628、Transcript19404)、*CCR* (Transcript27658、Transcript31478、Transcript29852、Transcript35302、Transcript34431)、*DIR* (Transcript39041)、*PLR* (Transcript30736、Transcript17904) 在茎中高表达, 共有 11 条转录本,

苯丙烷代谢途径 (phenylpropanoid biosynthesis)、萜类和聚酮类化合物代谢途径 (metabolism of terpenoids and polyketides)、单萜生物合成途径 (monoterpenoid biosynthesis)、倍半萜和三萜生物合成途径 (sesquiterpenoid and triterpenoid biosynthesis)、异喹啉生物碱生物合成途径 (isoquinoline alkaloid biosynthesis) 等 (图 2-B)。

涉及 5 个酶基因; *HCT* (Transcript25508、Transcript13501), *CCoAMT* (Transcript35752)、*CCR* (Transcript23829)、*DIR* (Transcript35387)、*PLR* (Transcript29348、Transcript32836) 在嫩叶中高表达, 共有 7 条转录本, 涉及 5 个酶基因。值得关注的是, 在八角果实中高表达的基因中, 木脂素生物合成途径相关酶基因的转录本高表达数目最多, 共有 14 条, 涉及 6 个酶基因, 分别为 *PAL* (Transcript11647)、*C4H* (Transcript26532)、*4CL* (Transcript19915、Transcript12186)、*C3H* (Transcript19822)、*CCR* (Transcript33158、Transcript20745、Transcript34592、Transcript2785、Transcript32070、Transcript28835、Transcript32797)、*CAD* (Transcript27743、Transcript29397)。

部分转录本在多个组织部位高表达, 如 *CCR* 中 Transcript28835、Transcript32797、Transcript35302 和 Transcript34431 转录本在茎、嫩叶及果实中表达量相对较高; *CAD* 中 Transcript30196 和 Transcript27743 转录本在嫩叶、老叶及果实中表达量相对较高; *PLR* 中 Transcript32836 转录本在茎、叶及老叶中表达量相对较高。

表 3 参与八角木脂素生物合成的关键酶转录本数
Table 3 Key enzyme transcripts in lignan synthesis pathway of *I. verum*

编号	关键酶名称	缩写	转录本数目
1	苯丙氨酸氨裂合酶	PAL	2
2	肉桂酸 4-羟化酶	C4H	1
3	4-香豆酯辅酶 A 连接酶	4CL	5
4	对羟基肉桂酰辅酶 A	HCT	2
5	对香豆酸 3-羟化酶	C3H	1
6	咖啡酰辅酶 A 氧甲基转移酶	CCoAMT	2
7	肉桂酰辅酶 A 还原酶	CCR	13
8	肉桂醇脱氢酶	CAD	4
9	过氧化物酶	DIR	3
10	松脂醇还原酶	PLR	6

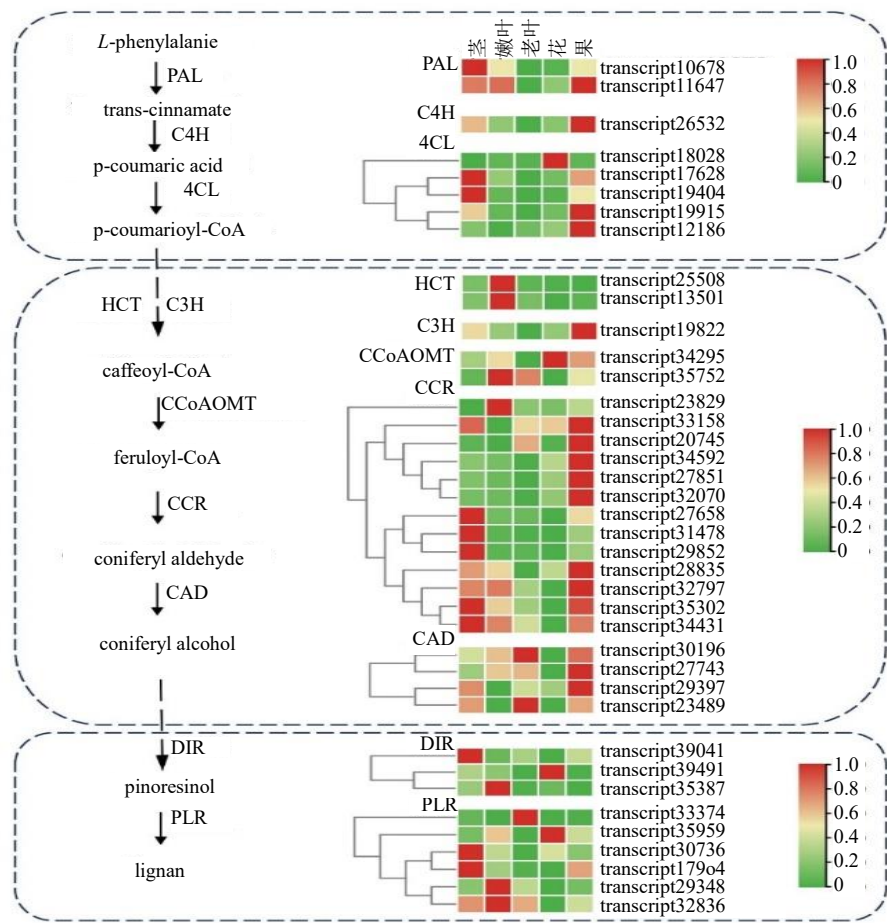


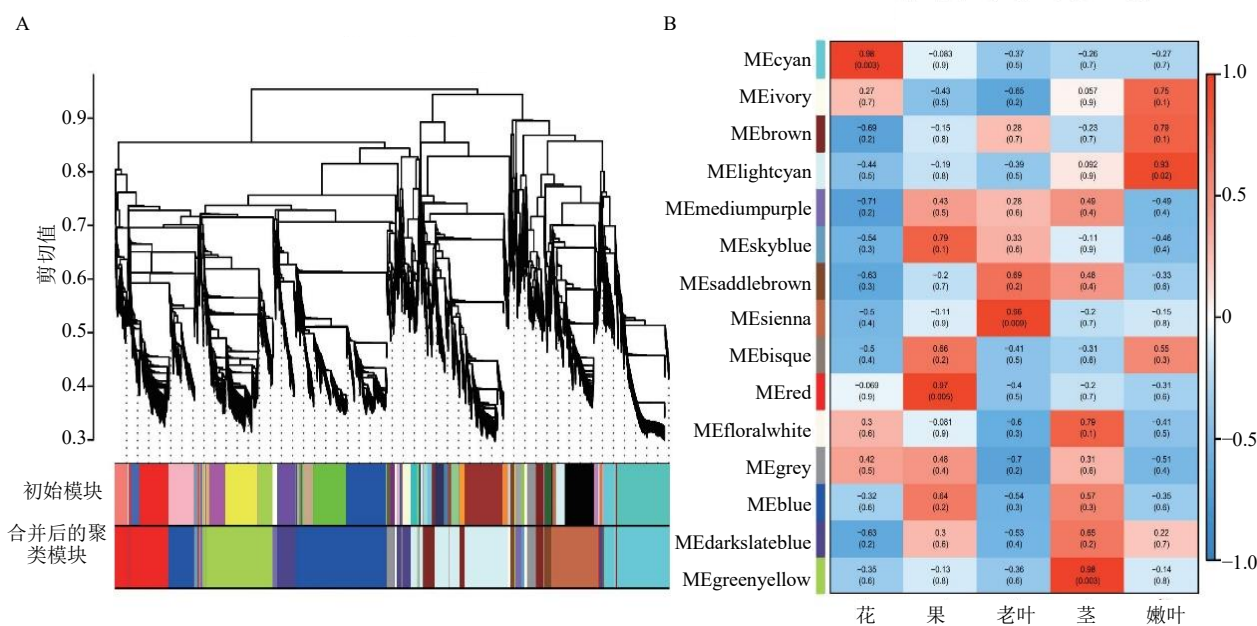
图 3 八角木脂素生物合成途径及其关键酶表达量热图

Fig. 3 Lignan biosynthesis pathway in *I. verum* and heat map of key enzyme expression

3.5 WGCNA

为了更深入地挖掘八角果实中木脂素生物合成途径的核心基因，通过 WGCNA 方法构建八角茎、嫩叶、老叶、花、果实 5 个不同组织部位的加权基因共表达网络，按照动态剪切树法的标准划分模块，共划分为 15 个基因共表达网络模块，每个模块用不同的聚类树以及颜色代表（图 4-A）。

根据不同共表达模块与 5 个组织部位进行相关性分析的结果，发现 9 个组织特异性表达模块（图 4-B）。其中，MEfloralwhite 模块和 MEgreenyellow 模块中的基因与八角茎高度相关，相关性分别为 0.79 和 0.98；MEivory 模块、MEbrown 模块和 MELightcyan 模块中的基因与八角嫩叶高度相关，相关性分别为 0.75、0.79 和 0.93；MESienna 模块中的



A-基因层次聚类树；B-基因模块与八角不同组织部位关联热图。

-A -gene hierarchical clustering tree; B-association heat map of gene module and different tissue parts of *I. verum*.

图 4 八角中不同组织部位 DEGs 的 WGCNA

Fig. 4 WGCNA of DEGs in different tissues of *I. verum*

基因与八角老叶高度相关，相关性为 0.98；MEcyan 模块中基因与八角花高度相关，相关性为 0.98；MESkyblue 模块和 MERed 模块中的基因与八角果实高度相关，相关性分别为 0.79 和 0.97。为了验证基因共表达网络和组织特异性模块两者结果的可靠性，对与果实相关性最高的 MERed 模块中与木脂素生物合成相关基因进行挖掘和分析。结果显示，MERed 模块中有 1 672 个基因，其中 4CL (Transcript12186)、CCR (Transcript27851, Transcript33158) 在该模块中被鉴定到，这与通过比较转录组学分析挖掘到的果实中高表达的木脂素生物合成相关酶基因是基本一致的，表明通过 WGCNA 分析对果实中木脂素生物合成关键酶基因的挖掘是非常可靠的。

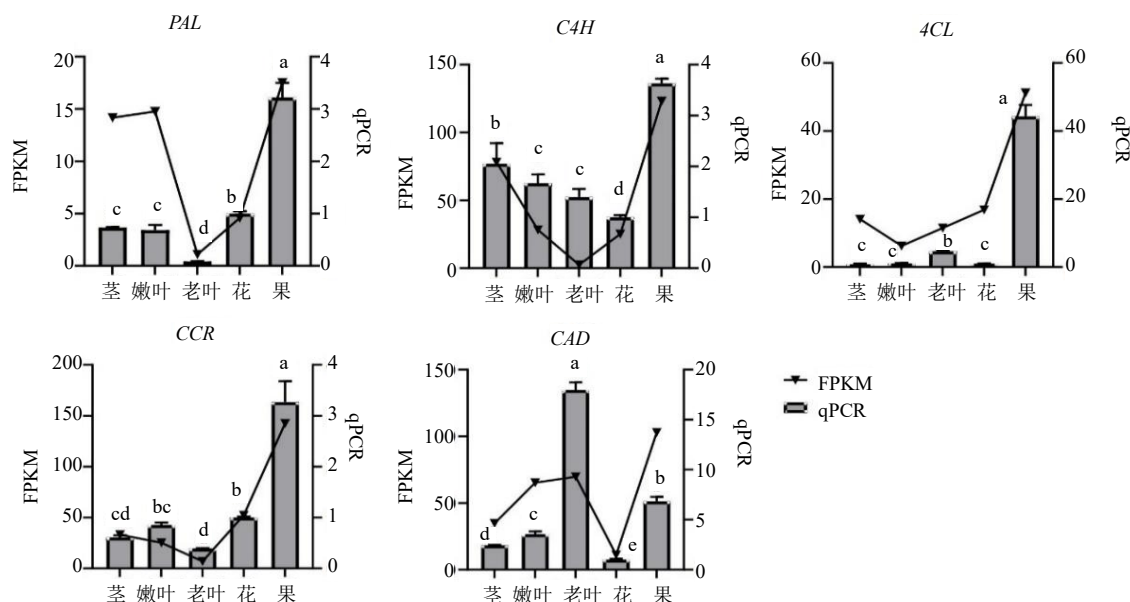
3.6 qRT-PCR 验证基因的表达

为验证转录组数据的准确性，根据 WGCNA 分析结果，采用 qRT-PCR 对八角果实中特异性表达的木脂素生物合成关键酶基因 4CL (Transcript12186)、CCR (Transcript27851) 的表达水平进行验证。另外 PAL、C4H、CAD 作为植物木脂素生物合成途径中的关键酶基因，因此，有必要选取部分在八角果实中表达量高的转录本进行 qRT-PCR 验证。结果显

示，木脂素生物合成关键酶基因 4CL (Transcript12186)、CCR (Transcript27851) 均特异地在果实中高表达，PAL (Transcript11647)、C4H (Transcript26532) 和 CAD (Transcript27743) 在果实的表达量最高，且在其他组织如茎、嫩叶、老叶中也检测到较高的表达水平。通过 Duncan 法进行多重比较结果发现，除 CAD (Transcript27743) 外，其余 4 个基因在果实中的表达量显著高于其它被测组织，且各组织间的相对表达水平也存在差异。关键酶基因 CAD (Transcript27743) 则在老叶中高表达，与其它组织相比具有显著性差异。此外，这 5 个基因在不同组织部位的表达水平与转录组数据中 FPKM 值的表达模式趋势保持一致，进一步验证了转录组分析结果的可靠性 (图 5)。

4 讨论

八角是我国重要药食两用植物资源，广西素有“八角之乡”的美称，是我国乃至世界八角的主要产区。从中医理论来看，八角味辛性温，归肝、肾、脾、胃经，具有温阳散寒、理气止痛之效。现代药理学研究也表明八角中的木脂素具有良好的药理活性，如保肝降酶、抗炎、抗肿瘤、抗菌、抗氧化等活性^[14]。目前对八角的研究主要集中在其化学成



不同小写字母表示组间差异显著 ($P < 0.05$)。

Marked with different lowercase letters indicates significant difference between groups ($P < 0.05$).

图 5 八角木脂素生物合成途径关键酶基因 qRT-PCR 验证

Fig. 5 qRT-PCR verification of key enzyme genes in lignan synthesis pathway of *I. verum*

分的提取、鉴定及药理活性研究等方面,对其果实主要活性物质木脂素生物合成的分子机制方面研究较少^[15]。不仅如此,木脂素在植物中的合成过程也较为复杂,尽管上游途径的相关酶基因已经研究得较为清楚,但是木脂素聚合以及各种单体形成过程中的羟基化、甲基化、糖基化催化过程仍有待深入解析。本研究构建了八角全长转录组数据库,通过比较转录组学以及 WGCNA 分析挖掘八角果实中木脂素生物合成途径关键酶基因,为进一步研究八角果实中木脂素的合成与调控机制提供了理论支撑。

近年来,基因组学、转录组学、代谢组学等组学数据及分析工具的开发,为研究植物中复杂的代谢网络提供了基础和平台^[16]。八角基因组十分庞大(10 Gb 左右,课题组开展了基因组 survey 工作),为开展八角基因组的测序工作带来了相当大的挑战以及相当高的成本,目前转录组测序已经是非常成熟的技术手段,为研究非模式植物基因功能提供相对经济且可靠的基因遗传信息。在最近前人的研究中,八角转录组测序均是基于二代转录组技术,获得的转录组数据库中转录本的平均长度为 1 212 bp, N₅₀ 长度 1 987 bp; 以及平均长度为 681 bp, N₅₀ 长度为 1 054 bp^[17-18]。本研究对八角的不同组织部

位进行三代全长转录组测序,对测序数据进行筛选、拼接后得到 32 996 条全长转录本,转录本的平均长度为 2 033 bp, N₅₀ 长度 2 343 bp,通过功能基因注释,其中有 31 248 条在 Nr、KOG、Pfam、Uniprot、KEGG、GO 等数据库中得到注释,相比于之前八角的转录组数据库质量有了极大的提升,所获得的全长转录本数据库则进一步丰富了八角组学数据,为八角分子生物学研究提供了更为可靠的基因信息。

药用植物次生代谢合成途径关键催化酶基因鉴定和功能研究是解析药效成分生物合成机制的关键。本研究基于八角高质量的全长转录组数据库,共筛选出可能参与木脂素生物合成 10 个关键酶基因的 39 条全长转录本,分别为 *PAL*、*C4H*、*4CL*、*HCT*、*C3H*、*CCoAMT*、*CCR*、*CAD*、*DIR* 和 *PLR*。通过比较转录组学分析,发现八角果实中与其他组织相比,差异表达的基因富集在苯丙烷代谢途径中。值得关注的是,木脂素生物合成途径的上游基因如 *PAL*、*C4H*、*4CL*、*C3H*、*CCR* 均有转录本在八角果实中高表达,而下游基因如 *DIR* 和 *PLR* 基因更多的是在八角的茎、花、叶中表达,而在果实中的表达量较低(图 3),这可能是由于这 2 个基因在八角果实不同发育时期的表达差异导致的,类似

的情况在五味子和佛手瓜中也存在^[19-20]。本研究筛选出的转录本在八角木脂素生物合成途径中的各个关键催化过程均有涉及,这些候选基因为后续八角中木脂素生物合成途径解析、基因功能验证及木脂素生物合成的中间体和衍生物的挖掘提供了基础。

综上,本研究首次利用 PacBio RS II 测序平台构建了八角的三代全长转录组数据库,推导了木脂素成分的生物合成途径并对关键催化酶基因进行了鉴定,通过比较转录组和 WGCNA 分析了木脂素生物合成途径关键酶基因在不同组织中的表达情况并挖掘到了果实中特异表达的关键酶基因,为进一步研究八角果实中木脂素成分生物合成的分子机理奠定了基础。

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: 5.
- [2] Liao W J, Luo J, Zou D X, *et al.* First report of anthracnose caused by *Colletotrichum siamense* on *Illicium verum* in China [J]. *Plant Dis*, 2023, 25: 36.
- [3] Zou Q Y, Huang Y Y, Zhang W Y, *et al.* A comprehensive review of the pharmacology, chemistry, traditional uses and quality control of star anise (*Illicium verum* hook. F.): An aromatic medicinal plant [J]. *Molecules*, 2023, 28(21): 7378.
- [4] Li W, Wu Z N, Xia Y P, *et al.* Antiviral and antioxidant components from the fruits of *Illicium verum* hook.f. (chinese star anise) [J]. *J Agric Food Chem*, 2022, 70(12): 3697-3707.
- [5] 邓富声, 贾学恭, 黄园园, 等. 八角属植物萜类成分及其药理活性研究进展 [J]. *中国药业*, 2022, 31(17): 128.
- [6] Jang W Y, Kim M Y, Cho J Y. Antioxidant, anti-inflammatory, anti-menopausal, and anti-cancer effects of lignans and their metabolites [J]. *Int J Mol Sci*, 2022, 23(24): 15482.
- [7] 曾武. 新型自由基诱导剂的合成及其抗肿瘤抗菌疗效研究 [D]. 郑州: 郑州大学, 2021.
- [8] 叶凤梅. 贡山八角枝叶中化学成分及抗菌抗肿瘤活性研究 [D]. 上海: 上海交通大学, 2015.
- [9] Ražná K, Harenčár L, Kučka M. The involvement of microRNAs in plant lignan biosynthesis-current view [J]. *Cells*, 2022, 11(14): 2151.
- [10] 潘媛, 陈大霞, 宋旭红, 等. 基于转录组测序挖掘国槐花蕾黄酮类物质生物合成相关基因 [J]. *中国中药杂志*, 2018, 43(13): 2682-2689.
- [11] 路艳, 马明东, 曹慧, 等. 水杨酸诱导下朱砂根转录组分析及三萜皂苷生物合成途径关键酶基因挖掘 [J]. *华北农学报*, 2023, 38(2): 106-119.
- [12] Yang M Q, Yao B W, Lin R M. Profiles of metabolic genes in *Uncaria rhynchophylla* and characterization of the critical enzyme involved in the biosynthesis of bioactive compounds-(iso)rhynchophylline [J]. *Biomolecules*, 2022, 12(12): 1790.
- [13] 王艺萌, 王知斌, 孙延平, 等. 苍术属植物中倍半萜类化合物化学结构和生物活性研究进展 [J]. *中草药*, 2021, 52(1): 299-309.
- [14] Xue L J, Liu K, Yan C X, *et al.* *Schisandra* lignans ameliorate nonalcoholic steatohepatitis by regulating aberrant metabolism of phosphatidylethanolamines [J]. *Acta Pharm Sin B*, 2023, 13(8): 3545-3560.
- [15] 侯振丽, 胡爱林, 石旭柳, 等. 八角茴香的化学成分及生物活性研究进展 [J]. *中药材*, 2021, 44(8): 2008-2017.
- [16] 陈泓宇, 董树廷, 郭妙弦, 等. 茯苓基因组中 MYB 转录因子基因家族鉴定及表达分析 [J]. *中草药*, 2023, 54(1): 245-253.
- [17] 陈博雯, 李开祥, 曾祥艳, 等. 基于转录组的八角挥发油合成相关基因挖掘与分析 [J]. *广西植物*, 2023, 43(2): 303-314.
- [18] 曾祥艳, 李开祥, 黄开顺, 等. 两种种源八角花色差异性的转录组分析 [J]. *分子植物育种*, 2021, 15: 1-16.
- [19] 李海燕, 刘久石, 魏如滨, 等. 五味子 PLR 基因及其启动子的克隆与表达分析 [J]. *中草药*, 2020, 51(18): 4747-4754.
- [20] 赵倩, 谭国飞, 钟秀来, 等. 佛手瓜 SePLR 基因克隆和生物学分析 [J]. *贵州农业科学*, 2022, 50(8): 78-86.

[责任编辑 时圣明]