

苗药八爪金龙转录组测序与次生代谢产物合成相关基因的挖掘

刘 畅¹, 俸婷婷^{1,2}, 刘雄伟¹, 丁晶鑫¹, 石 慧¹, 潘 婕¹, 周 英^{1,2*}

1. 贵州中医药大学药学院, 药食两用资源应用与开发研究中心, 中药材开发技术研究中心, 贵州 贵阳 550025

2. 贵州省药食同源植物资源开发工程技术研究中心, 贵州 贵阳 550025

摘 要: **目的** 对八爪金龙根进行转录组测序, 挖掘次生代谢合成相关基因, 探索八爪金龙次生代谢产物生物合成的分子基础。**方法** 采用 Illumina HiSeq 4000 高通量测序技术, 对八爪金龙根进行转录组测序。使用 Trinity 软件对获得的 Unigenes 数据进行过滤组装, 运用 KEGG 数据库对 Unigenes 进行注释。**结果** 共获得 52 249 条 Unigenes, 其中 31 391 条被公共数据库成功注释, 1507 条 Unigenes 被注释到次生代谢产物合成。通过对转录组数据深入挖掘发现, 参与八爪金龙苯丙素生物合成的 Unigenes 共有 126 条, 参与萜类化合物骨架生物合成的 Unigenes 数量为 73 条, 参与黄酮类化合物生物合成 Unigenes 有 58 条, 参与次生代谢后修饰的 Unigenes 共有 253 条。筛选出 9 条 Unigenes 编码 4 个与八爪金龙香豆素生物合成的关键酶, 23 条 Unigenes 编码 8 个与黄酮生物合成的关键酶, 39 条 Unigenes 编码 19 个与萜类生物合成的关键酶, 140 条 CYP450 基因和 113 条 UGT 基因可能参与次生代谢物的修饰。微卫星识别软件 (microsatellite identification tool, MISA) 分析发现八爪金龙转录组包含 17 400 个简单重复序列 (simple sequence repeats, SSR)。**结论** 通过高通量转录组测序, 初步揭示了参与八爪金龙次生代谢产物合成相关的基因, 为进一步研究八爪金龙次生代谢产物合成途径关键酶的功能及其调控机制奠定了基础。

关键词: 八爪金龙; 转录组; 次生代谢; 生物合成; 基因

中图分类号: R282.12

文献标志码: A

文章编号: 0253-2670(2021)05-1434-14

DOI: 10.7501/j.issn.0253-2670.2021.05.025

Transcriptome analysis and identification of related genes involved in secondary metabolism biosynthesis in *Ardisia crispa*

LIU Chang¹, FENG Ting-ting^{1,2}, LIU Xiong-wei¹, DING Jing-xin¹, SHI Hui¹, PAN Jie¹, ZHOU Ying^{1,2}

1. School of Pharmacy/Research Center for Application and Development of Medicine and Food Dual-use Resources/Research Center for Development of Traditional Chinese Medicine, Guizhou University of Traditional Chinese Medicine, Guiyang 550025, China

2. Guizhou Engineering Center for Innovative Traditional Chinese Medicine and Ethnic Medicine, Guizhou University, Guiyang 550025, China.

Abstract: Objective To explore related genes of the secondary metabolic biosynthesis and their molecular basis of *Ardisia crispa* root by using high-throughput sequencing. **Methods** The high-throughput RNA-seq technology (Illumina HiSeq 4000) was used to conduct a transcriptomic analysis of *A. crispa* root. Unigenes splicing was performed using Trinity software and annotated with KEGG database to predict the specific pathways. **Results** A total of 52 249 Unigenes were obtained, of which 31 391 were annotated in the public database, 1507 Unigenes were annotated for pathways of secondary metabolism biosynthesis. Based on the bioinformatic analysis, 126 Unigenes were involved in the phenylpropanoid biosynthesis, 73 Unigenes might participate in terpenoid backbone biosynthesis, 58 Unigenes were involved in the flavonoid biosynthesis, and 253 Unigenes might participate in secondary metabolism post-modification. A total of 9, 23, 39 Unigenes were screened out from the transcriptome database, which encoded 4, 8, 19 key metabolic enzymes associated with coumarin, flavone and terpenoid biosynthesis, respectively. 140 CYP450 genes and 113

收稿日期: 2020-08-09

基金项目: 国家重点研发计划项目 (2018YFC1708100); 贵州省高层次创新型人才培养项目 (黔科合人才[2015]4032 号); 贵州省科技厅学术新苗项目 (黔科合平台人才[2018]5766 号-9); 贵州中医药大学博士启动基金 (贵中医博士启动 [2019]04 号)

作者简介: 刘 畅 (1988—), 女, 讲师, 研究方向为分子生药学。E-mail: 19liuchang@163.com

***通信作者:** 周 英 (1971—), 女, 教授, 研究方向为中药药效物质基础与质量控制研究和中药新药的研究开发。E-mail: yingzhou71@126.com

UGT genes may be involved in the modification of secondary metabolites. MISA prediction found 17 400 simple sequence repeats (SSRs). **Conclusion** The high-throughput transcriptome sequencing revealed the genes involved in the synthesis of flavonoids and triterpenes, which lays a foundation for further studies on function and regulatory mechanisms of key enzymes involved in the synthesis of secondary metabolites in *A. crispa*.

Key words: *Ardisia crispa* (Thunb.) A. DC.; transcriptome; secondary metabolism; biosynthesis; gene

八爪金龙 *Ardisia crispa* (Thunb.) A. DC. 是紫金牛科紫金牛属植物, 药用部位为根, 别名朱砂根、百两金、八爪龙、八爪根、铁雨伞、高八爪、开喉箭等, 该药材在贵州分布广、蕴藏量大, 被苗族奉为喉科良药; 具有清热解毒、散瘀止痛、祛风除湿之功效, 用于治疗咽喉肿痛、扁桃体炎、心胃气痛、劳伤吐血、跌扑损伤、风湿骨痛^[1]。现代药理表明八爪金龙具有镇痛、抗炎、抗癌、抗关节炎、抗白血病^[2-5]等作用, 是苗药验方开喉剑喷雾剂的主要成分^[6]。

八爪金龙含有三萜苷类、黄酮类、异香豆素、叶绿素、挥发油、三萜类、酚类、醌类、强心苷、有机酸、鞣质、氨基酸、糖类等多种化学成分^[7-10]。其中香豆素类的岩白菜素为其主要有效成分, 黄酮类物质, 如汉黄芩素、千层纸素、汉黄芩素、黄芩苷^[11]与岩白菜素协同起到止咳化痰的作用。随着研究的深入, 发现三萜苷或苷元类化合物是其化学成分和活性成分, 目前已经从八爪金龙中分离得到 28 个三萜皂苷或皂苷元类成分^[12-14]。但目前对八爪金龙的研究主要集中在化学成分^[15]、药理活性^[16-17]、分子鉴别^[18]等方面, 而对八爪金龙次生代谢成分分子生物合成相关基因发掘和利用却鲜有报道。

转录组高通量测序技术逐渐成为研究药用植物天然活性成分生物合成相关功能基因挖掘及其表达规律的重要手段, 如青蒿 *Artemisia carvifolia* Buch. - Ham. ex Roxb. Hort. Beng. ^[19]、西洋参 *Panax quiquefolium* L. ^[20]、人参 *Panax ginseng* C. A. Meyer ^[21]、金银花 *Lonicera japonica* Thunb. ^[22]等都已经完成了转录组测序分析, 积累了一批与药用次生代谢产物合成调控相关的基因。八爪金龙作为一种民族药, 具有重要的应用价值, 但其功能基因研究基础十分薄弱。因此, 本研究利用 Illumina Hiseq 测序平台对八爪金龙根进行转录组测序, 以期获得与其有效成分合成相关的基因信息, 为进一步挖掘与克隆八爪金龙新的功能基因奠定基础。

1 材料与方法

1.1 材料

苗药八爪金龙根部样品采于贵阳花溪区,

经贵州中医药大学魏升华教授鉴定为苗药八爪金龙 *A. crispa* (Thunb.) A. DC.。取八爪金龙根部用锡箔纸包裹并标记好后放置于液氮中冷冻, 冻存于 -80 °C 冰箱备用, 用于后续 RNA 的提取。

1.2 RNA 的提取及测序

将样品在冰上融化后粉碎, 充分混匀并离心, 取适量上清, 用 Trizol 试剂分离提取总 RNA, 用 Oligotex mRNA 试剂盒对 RNA 进行纯化。利用 Agilent 2100 检测 RNA 样品的浓度和完整性, 12 个苗药八爪金龙 RNA 样品均满足转录组测序的建库要求。满足建库要求的 RNA 样品送至北京百迈克生物科技有限公司, 使用 Illumina HiSeq 4000 平台进行转录组测序。

1.3 数据过滤及组装

对转录组测序得到原始测序数据 (raw reads) 后进行数据过滤, 去除掉低质量、包含接头和未知碱基 N 含量过高的 reads, 得到高质量测序数据 (clean reads)。利用 Trinity 软件^[23]对 clean reads 进行重头组装, 用 cd-hit 软件去除完全一样的序列, 然后使用 tgi1 进行聚类, 合并相似度大于 90%, overlap 长度大于 35 的序列, 最后得到 Unigenes。

1.4 Unigenes 的注释和分类

使用 BLAST^[24]软件将 Unigenes 序列与非冗余蛋白序列 (NCBI non-redundant protein sequences, NR)、注释和修订蛋白序列数据库 (a manually annotated and reviewed protein sequence database, Swiss-Prot)、基因本体论 (gene ontology, GO)、直系同源蛋白簇 (clusters of orthologous groups, COG)、真核同源群簇 (clusters of euKaryotic orthologous groups, KOG)、京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG)^[25-31]数据库比对; 使用 KOBAS3.0^[32]得到 Unigenes 在 KEGG 中的 KEGG Orthology 结果; 预测完 Unigenes 的氨基酸序列之后使用 HMMER 软件^[33]与 Pfam (protein family)^[34]数据库比对, 获得 Unigenes 的注释信息。

1.5 基因表达量分析

采用 Li 等^[35]将测序得到的 reads 与 Unigenes

库进行比对,根据比对结果,结合 RSEM (RNA-Seq by Expectation Maximization) [36]进行表达量水平估计。利用 FPKM (fragments per kilobase of transcript per million mapped reads) [37]值表示对应 Unigenes 的表达丰度。FPKM 是每百万 reads 中来自比对到某一基因每千碱基长度的 reads 数目,是转录组测序数据分析中常用的基因表达水平估算方法。FPKM 能消除基因长度和测序量差异对计算基因表达的影响。

1.6 SSR 位点分析

利用鉴定单重复序列的软件 MISA (microsatellite identification tool) 对转录组数据中的 Unigenes 进行 SSR 位点的检测。按照单碱基、双碱基、三碱基重复分别至少 12、6、5 次,四碱基、五碱基、六碱基都不少于 4 次的标准

进行检索分析。最后对获得的 SSR 数据进行分类统计分析。

2 结果与分析

2.1 转录组测序与序列组装

利用 Illumina HiSeq 4000 测序平台对 12 个苗药八爪金龙根部样品进行转录组测序,共获得 89.73 Gb 高质量的 clean reads,其 Q₃₀ 平均值达到 95.05%,平均每个样品的 GC 含量占总碱基数的 45.78% (表 1)。利用 Trinity 对 clean reads 进行序列组装,共获得 52 249 条 Unigenes,总长度为 75 282 082 nt,平均长度为 1440 nt, N₅₀ 为 2336 nt,组装完整性较高。长度分布在 300~500 nt 的 Unigenes 最多,有 15 065 条,占 Unigenes 总数的 28.83%;长度大于 1000 nt 的 Unigenes 有 25 190 条;长度大于 2000 nt 的 Unigenes 有 13 283 条,占 Unigenes 总数的 25.42%。

表 1 样品测序数据统计

Table 1 Transcriptome data output quality

样品	clean reads/条	base-number/bp	GC/%	Q ₃₀
<i>Ardisia crispa</i> _01	27 425 579	8 195 040 960	45.67	94.85
<i>Ardisia crispa</i> _02	26 562 212	7 936 440 054	45.63	94.95
<i>Ardisia crispa</i> _03	22 037 780	6 582 090 830	45.65	95.15
<i>Ardisia crispa</i> _04	22 346 460	6 674 112 542	45.90	94.66
<i>Ardisia crispa</i> _05	19 247 338	5 742 837 384	45.23	95.27
<i>Ardisia crispa</i> _06	22 099 726	6 599 176 204	45.93	95.02
<i>Ardisia crispa</i> _07	26 117 088	7 794 477 474	45.80	94.92
<i>Ardisia crispa</i> _08	22 726 418	6 781 723 034	45.77	94.77
<i>Ardisia crispa</i> _09	30 932 371	9 243 781 820	46.24	95.30
<i>Ardisia crispa</i> _10	26 259 169	7 832 051 982	45.93	95.10
<i>Ardisia crispa</i> _11	26 490 511	7 916 611 478	45.79	95.33
<i>Ardisia crispa</i> _12	28 251 394	8 430 053 900	45.87	95.29

2.2 序列比对及基因注释

对 52 249 条 Unigenes 进行 COG、GO、KEGG、KOG、Pfam、SwissProt、eggNOG、NR 8 大数据库比对。结果表明共有 31 391 条 Unigenes 被注释到数据库中,占全部 Unigenes 的 60.07%;其中 NR 数据库中注释到的 Unigenes 最多,有 31 179 条,占总 Unigenes 的 59.67%;其次是 eggNOG 数据库,注释到 29 953 条 Unigenes (57.32%);COG、GO、KEGG、KOG、Pfam 和 SwissProt 数据库,分别注释到 10 988 (21.03%)、19 195 (36.73%)、12 775 (24.45%)、18 741 (35.86%)、22 793 (43.62%) 和 22 173 (42.43%) 条 Unigenes。Unigenes 注释同源基因的物种分布如图 1 所示,在相似序列匹配度较高的物种中,葡萄 *Vitis vinifera* L. 所占比例最高,

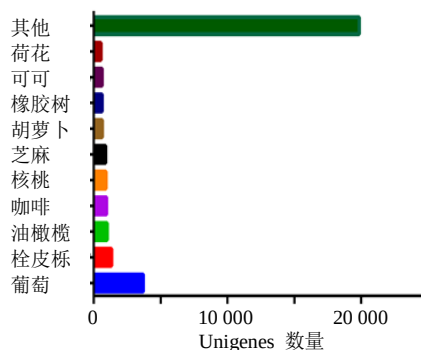


图 1 八爪金龙转录组 Unigenes 与 NR 数据库匹配物种分布
Fig. 1 Species distribution of *A. crispa* transcriptome Unigenes against NR database

3717 条;其次为栓皮栎 *Quercus suber* Bl. (1334 条)、油橄榄 *Olea europaea* L. (1002 条)、咖啡 *Coffea*

canephora L. (960 条)、核桃 *Juglans regia* L. (928 条)、芝麻 *Sesamum indicum* L. (902 条)、胡萝卜 *Daucus carota* Linn. var. *sativa* Hoffm. (640 条)、橡胶树 *Hevea brasiliensis* (Willd. ex A. Juss.) Muell. Arg. (634 条)、可可 *Theobroma cacao* L. (631 条)、荷花 *Nelumbo nucifera* Gaertn. (546 条)，其他匹配物种的 Unigenes 为 19 848 条。

2.3 Unigenes 的 KOG 分类

对苗药八爪金龙转录组数据进行 KOG 数据库功能注释和分类，共有 18 741 条 Unigenes 被注释到 25 种 KOG 分类中，有 4339 条 Unigenes 被注释到“一般功能预测 (general function prediction only)”，是注释最多 Unigenes 的类群，占总数的 23.15%；其次是“翻译后修饰，蛋白质周转，伴侣 (posttranslational modification, protein turnover, chaperones)” 2197 条，“信号转导机制 (signal transduction mechanisms)” 1632 条，“翻译，核糖体结构和生物发生 (translation, ribosomal structure and biogenesis)” 1252 条。注释到“碳水化合物运输和代谢 (carbohydrate transport and metabolism)” 类群的 Unigenes 有 1075 条，803 条 Unigenes 注释到“脂质转运与代谢 (lipid transport and metabolism)”，729 条 Unigenes 注释到“氨基酸转运与代谢 (amino acid transport and metabolism)”，546 条 Unigenes 注释到

“无机离子转运与代谢 (inorganic ion transport and metabolism)”，681 条 Unigenes 注释到“次生代谢物的生物合成、运输和分解代谢 (secondary metabolites biosynthesis, transport and catabolism)” (图 2)。

2.4 Unigenes 的 GO 和 KEGG 分析

利用 GO 数据库对 Unigenes 进行功能注释，19 195 条 Unigenes 被注释到细胞组成 (cellular component)，分子功能 (molecular function) 和生物过程 (biological processes) 3 大类。在细胞组成中 Unigenes 主要聚集在细胞 (cell) 和细胞部分 (cell part) 2 个类群，分别注释了 8902 和 8875 条；分子功能中聚集最多的 2 个类群是催化活性 (catalytic activity) 和结合蛋白 (binding)，分别注释到 9627 和 8961 条 Unigenes；在生物过程中聚集最多的 2 个类群是代谢过程 (metabolic process) 和细胞过程 (cellular process)，分别注释了 10 006 和 9388 条 Unigenes (图 3)。

通过 KEGG 数据库对八爪金龙 Unigenes 进行注释，共注释到 128 个 KEGG 标准代谢通路。按照基因注释量大小依次排序，选取前 20 个代谢通路 (表 2)，主要注释到代谢通路 (metabolic pathways)、次生代谢生物合成 (biosynthesis of secondary metabolites)、核糖体 (ribosome)、碳代谢 (carbon metabolism)、氨基酸生物合成 (biosynthesis

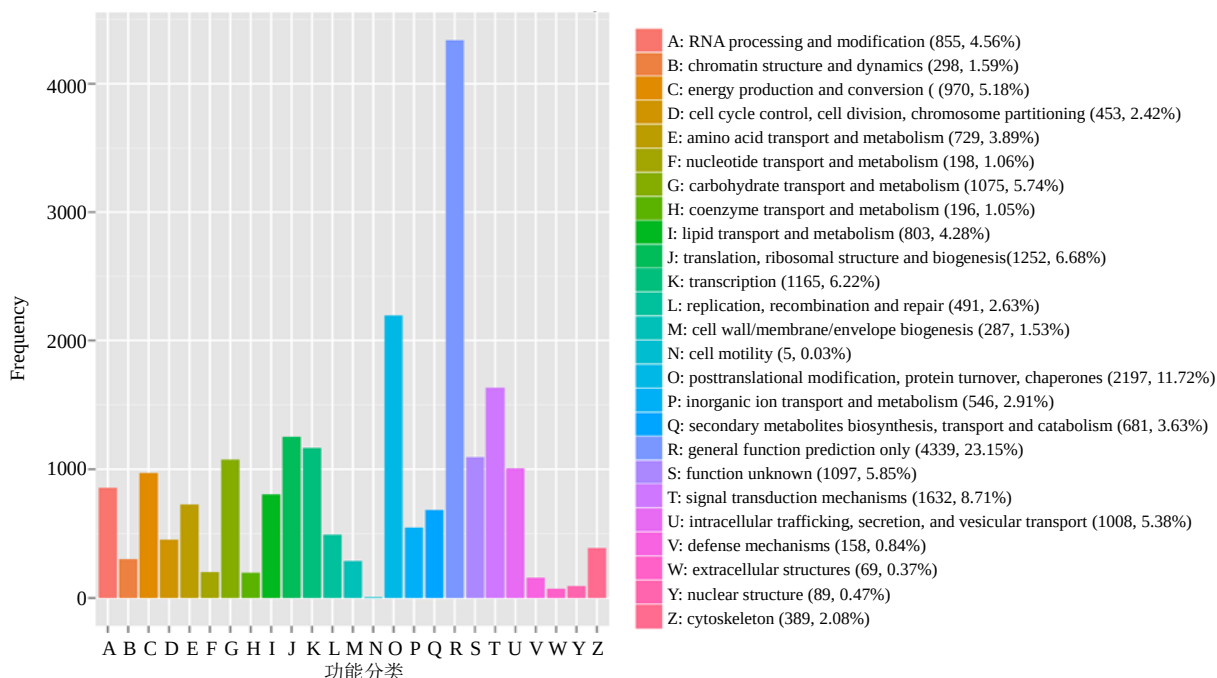


图 2 八爪金龙转录组的 KOG 功能分类

Fig. 2 KOG functional classifications of *A. crispa* transcriptome

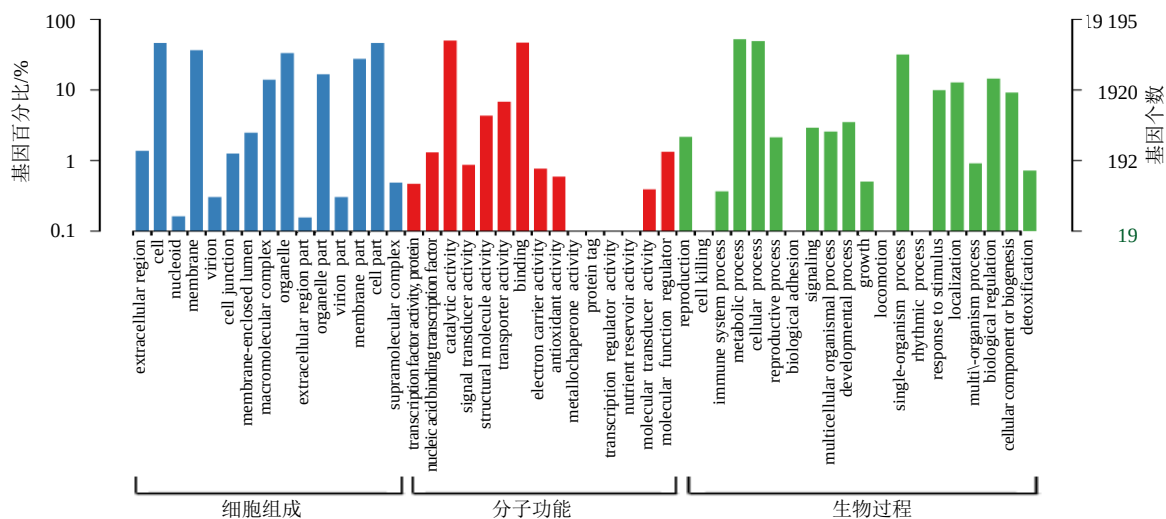


图 3 八爪金龙转录组的 GO 功能分类

Fig. 3 GO functional classifications of *A. crispa* transcriptome

表 2 八爪金龙转录组 Unigenes KEGG 通路分析统计

Table 2 KEGG functional classifications of *A. crispa* transcriptome

编号	通路	通路_id	Unigenes 数量
1	代谢通路 (metabolic pathways)	ko01100	2812
2	次生代谢物生物合成 (biosynthesis of secondary metabolites)	ko01110	1507
3	核糖体 (ribosome)	ko03010	652
4	碳代谢 (carbon metabolism)	ko01200	500
5	氨基酸生物合成 (biosynthesis of amino acids)	ko01230	416
6	内质网蛋白加工 (protein processing in endoplasmic reticulum)	ko04141	354
7	剪接体 (spliceosome)	ko03040	326
8	RNA 转运 (RNA transport)	ko03013	313
9	植物激素信号转导 (plant hormone signal transduction)	ko04075	289
10	内吞作用 (endocytosis)	ko04144	285
11	嘌呤代谢 (purine metabolism)	ko00230	264
12	淀粉和蔗糖代谢 (starch and sucrose metabolism)	ko00500	264
13	氧化磷酸化 (oxidative phosphorylation)	ko00190	256
14	糖酵解/糖异生 (glycolysis/gluconeogenesis)	ko00010	245
15	氨基酸和核苷酸代谢 (amino sugar and nucleotide sugar metabolism)	ko00520	199
16	泛素介导的蛋白水解作用 (ubiquitin mediated proteolysis)	ko04120	198
17	植物病原体相互作用 (plant-pathogen interaction)	ko04626	197
18	信使 RNA 监测通路 (mRNA surveillance pathway)	ko03015	192
19	嘧啶代谢 (pyrimidine metabolism)	ko00240	187
20	RNA 损害 (RNA degradation)	ko03018	182

of amino acids)、内质网蛋白加工 (protein processing in endoplasmic reticulum)、剪接体 (spliceosome)、RNA 转运 (RNA transport)、植物激素信号转导 (plant hormone signal transduction) 等通路。

2.5 八爪金龙次生代谢生物合成途径相关酶的鉴定

KEGG 通路分析发现 Unigenes 参与苯丙素、萜类、黄酮类、类胡萝卜素、玉米素、生物碱等生物

合成相关的 18 个次生代谢通路 (表 3)。苯丙素的生物合成代谢通路 (ko00940) Unigenes 数量最多, 为 126 条; 萜类化合物骨架生物合成 (ko00900) Unigenes 数量次之, 为 73 条; 与黄酮类化合物生物合成 (ko00941) 有关的 Unigenes 有 58 条; 其他萜类化合物生物合成 (ko00130) 的 Unigenes 数量有 52 条; 类胡萝卜素生物合成 (ko00906) 相关的

表 3 八爪金龙转录组 Unigenes 次生代谢 KEGG 通路注释

Table 3 Biosynthetic pathway of secondary metabolites involved in major active substances of *A. crispa*

编号	通路	通路_id	Unigenes 数量
1	苯丙素生物合成 (phenylpropanoid biosynthesis)	ko00940	126
2	萜类化合物骨架生物合成 (terpenoid backbone biosynthesis)	ko00900	73
3	黄酮类化合物生物合成 (flavonoid biosynthesis)	ko00941	58
4	其它萜类化合物生物合成 (ubiquinone and other terpenoid-quinone biosynthesis)	ko00130	52
5	类胡萝卜素生物合成 (carotenoid biosynthesis)	ko00906	51
6	玉米素生物合成 (zeatin biosynthesis)	ko00908	30
7	莨菪烷类、哌啶、吡啶生物碱生物合成 (tropane, piperidine and pyridine alkaloid biosynthesis)	ko00960	30
8	异喹啉类生物碱生物合成 (isoquinoline alkaloid biosynthesis)	ko00950	28
9	倍半萜和三萜类化合物生物合成 (sesquiterpenoid and triterpenoid biosynthesis)	ko00909	20
10	核黄素代谢 (riboflavin metabolism)	ko00740	18
11	二萜类生物合成 (diterpenoid biosynthesis)	ko00904	14
12	油菜素内酯生物合成 (brassinosteroid biosynthesis)	ko00905	12
13	单萜类生物合成 (monoterpenoid biosynthesis)	ko00902	6
14	黄酮和黄酮醇生物合成 (flavone and flavonol biosynthesis)	ko00944	6
15	咖啡因代谢 (caffeine metabolism)	ko00232	5
16	柠檬烯和蒎烯降解 (limonene and pinene degradation)	ko00903	4
17	芥子油苷生物合成 (glucosinolate biosynthesis)	ko00966	2
18	花青素生物合成 (anthocyanin biosynthesis)	ko00942	1

Unigenes 数量有 51 条；分别有 30、30、28、20、18、14、12、6、6 条 Unigenes 参与玉米素生物合成 (ko00908)，莨菪烷类、哌啶、吡啶生物碱生物合成 (ko00960)，异喹啉类生物碱生物合成 (ko00950)，倍半萜和三萜类化合物生物合成 (ko00909)，核黄素代谢 (ko00740)，二萜类生物合成 (ko00904)，油菜素内酯生物合成 (ko00905)，单萜类生物合成 (ko00902) 以及黄酮和黄酮醇生物合成 (ko00944)。咖啡因代谢 (ko00232)、柠檬烯和蒎烯降解 (ko00903)、芥子油苷生物合成 (ko00966) 和花青素生物合成 (ko00942) 通路基因数量较少。

2.5.1 香豆素生物合成相关基因的挖掘 苗药八爪金龙中与香豆素生物合成相关的代谢通路为苯丙素的生物合成代谢通路 (图 4)。莽草酸通过苯丙氨酸和酪氨酸等芳香氨基酸，经脱氨、羟基化等一系列反应形成，其中涉及多种酶的参与，苯丙氨酸解氨酶 (phenylalanine ammonia-lyase, PAL)、肉桂-4-羟化酶 (cinnamate 4-hydroxylase, C4H) 和 4-香豆酸-辅酶-A (4-coumarate-CoA ligase, 4CL) 是这条途径中的关键酶。在八爪金龙转录组数据中，共注释得到 126 条 Unigenes 注释到苯丙素生物合成代谢通路。编码香豆素生物合成途径的 5 种关键酶，包括 5 条 PAL 序列 (最

高 FPKM=74.58)，2 条 C4H 序列 (最高 FPKM=42.90)，2 条 COMT 序列 (最高 FPKM=54.82)，1 条 C3H 序列 (最高 FPKM=25.32)，仅鉴定出来 1 条 4CL (FPKM=0.23) (表 4 和图 5)，因此可以通过增强 4CL 基因的表达，以增加整条代谢通路的通量。

2.5.2 黄酮类化合物合成相关基因的挖掘 根据 KEGG 途径分析结果，以及八爪金龙化学成分分析结果，并参考相关文献报道^[38-40]，对八爪金龙中主要黄酮类化合物的生物合成途径作出预测，见图 4。在黄酮类成分合成途径中，查耳酮合酶 (chalcone synthase, CHS)、查耳酮异构酶 (chalcone isomerase, CHI) 和黄酮合酶 (flavone synthase, FNS) 起着重要的作用。八爪金龙黄酮合成途径的上游阶段 CHS (最高 FPKM=841.87) 和 CHI (最高 FPKM=36.55) 基因的表达量相对较高。编码 CHS 的 Unigenes 数量为 9 条，其各条 Unigenes 的表达量不同，其 Unigene_186235 表达量达到 841.87 (图 5)。二氢黄酮 3-羟化酶 (flavonoid 3-hydroxylase, F3H, 最高 FPKM=107.65) 可以催化柚皮素/槲皮素生成二氢黄酮醇类，黄酮醇合酶 (flavonol synthase, FLS) 分别催化二氢黄酮醇类生成黄酮醇，FLS 作为合成各

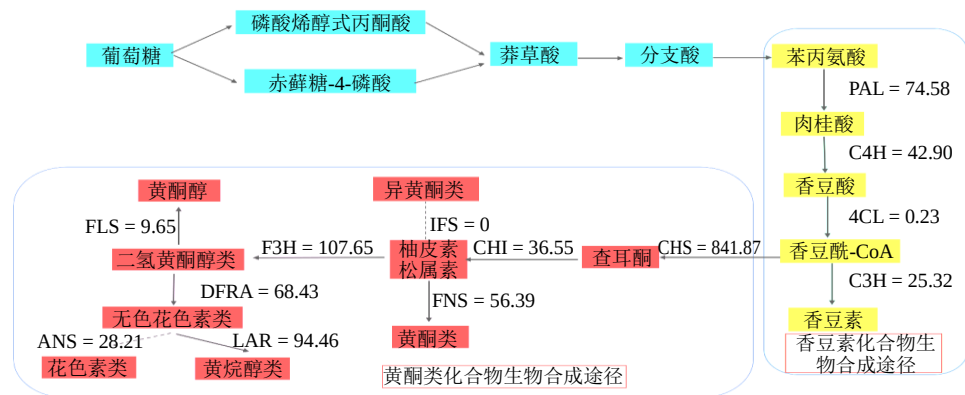


图 4 八爪金龙香豆素类和黄酮类物质生物合成途径

Fig. 4 Main coumarin and flavone biosynthesis pathway in *A. crispata*

表 4 编码香豆素、黄酮类和萜类生物合成相关酶的 Unigenes 数量

Table 4 Number of Unigenes encoding enzyme involved in coumarin, flavone, and terpenoid biosynthesis				
次生代谢生物合成途径	酶	Unigenes 数量	Unigenes ID	
香豆素生物合成	PAL	5	Unigene_025593、Unigene_102843、Unigene_102844、Unigene_108847、Unigene_177796	
	C4H	2	Unigene_013550、Unigene_181127	
	C3H	1	Unigene_019276	
	4CL	1	Unigene_171229	
	COMT	2	Unigene_096163、Unigene_177521	
黄酮类生物合成	ANS	1	Unigene_031279	
	ANR	1	Unigene_174753	
	AOMT	1	Unigene_173696	
	BIS1	1	Unigene_176702	
	CHI	3	Unigene_008799、Unigene_026449、Unigene_173931	
	CHS	9	Unigene_001289、Unigene_030909、Unigene_029241、Unigene_031790、Unigene_108709、Unigene_171144、Unigene_154152、Unigene_186236、Unigene_186235	
	CHS1	2	Unigene_008856、Unigene_182412	
	CHS2	2	Unigene_003950、Unigene_185402	
	CHSC	1	Unigene_168887	
	CYP450	1	Unigene_177152	
	CYP73A5	1	Unigene_162219	
	DFR	3	Unigene_026576、Unigene_106630、Unigene_173665	
	DFRA	5	Unigene_030972、Unigene_092658、Unigene_175249、Unigene_180983、Unigene_181720	
	F3'5'H	4	Unigene_094346、Unigene_094347、Unigene_170206、Unigene_170207	
	F3PH	1	Unigene_026466	
	FL3H	1	Unigene_010771	
	FLS	3	Unigene_013848、Unigene_014504、Unigene_015774	
	GhHCT	1	Unigene_011521	
	LAR	2	Unigene_008392、Unigene_174994	
	SHT	1	Unigene_102522	
萜类化合物	DXS	1	Unigene_026309、Unigene_004259	
	DXR	1	Unigene_020536	
	MCT	2	Unigene_027137、Unigene_179334	
	HMGR	3	Unigene_177002、Unigene_108513、Unigene_180184	
	HMGR1	1	Unigene_028276	
	HMGR2	1	Unigene_092745	
	HMGS	1	Unigene_010349	
	CMK	1	Unigene_011334	

续表 4

次生代谢生物合成途径	酶	Unigenes 数量	Unigenes ID
萜类化合物	HDS	2	Unigene_024431、Unigene_185236
	HDR	2	Unigene_020166、Unigene_172060
	AACT	5	Unigene_185261、Unigene_106598、Unigene_092108、Unigene_0850896、Unigene_176341
	β -AS	1	Unigene_026496、Unigene_099571
	MVD2	2	Unigene_027171、Unigene_031819
	FPS	3	Unigene_031229、Unigene_179181、Unigene_017588
	FPS1	1	Unigene_104061
	GGPPR	1	Unigene_011597
	GGPPS	4	Unigene_097676、Unigene_184574、Unigene_175076、Unigene_012804
	GGPPS1	2	Unigene_092983、Unigene_183619
	IDI	3	Unigene_092274、Unigene_099881、Unigene_012752
	MDD	1	Unigene_178872
	MVK	2	Unigene_082953、Unigene_011076
	SE	2	Unigene_010894、Unigene_173132
	SM	2	Unigene_179555、Unigene_030005
	SQS	2	Unigene_002401、Unigene_017821
	TS	1	Unigene_185916
	PMK	2	Unigene_106923、Unigene_178726
	MDS	1	Unigene_012674

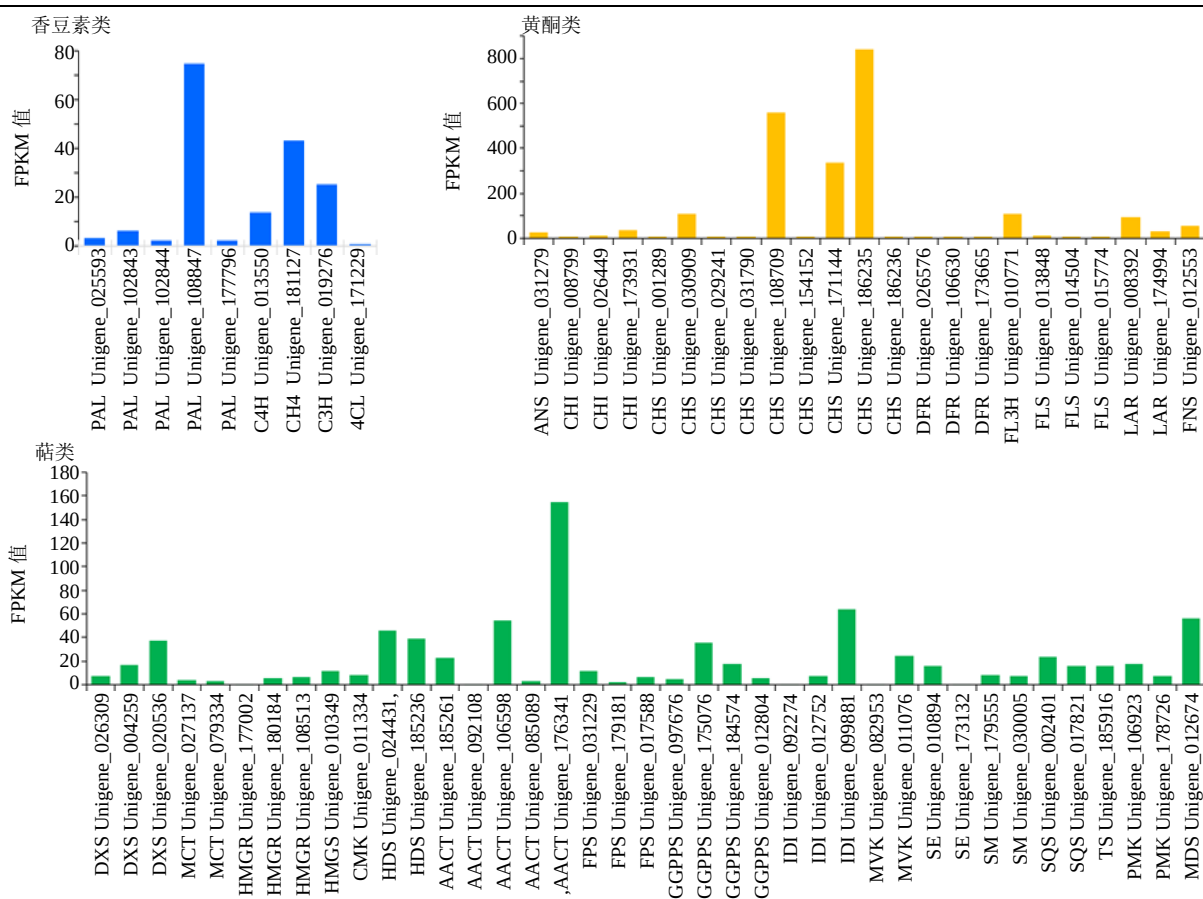


图 5 生物合成关键基因的表达式

Fig. 5 Expression of key genes in biosynthesis

种黄酮醇类的关键酶基因其表达量（最高 FPKM=9.65）很低（图 5），直接限制了黄酮醇类的合成；柚皮素/槲皮素可以在 FNS 生成黄酮类。同时，二氢黄酮醇 4-还原酶（dihydroflavonol-4-reductase, DFRA）（最高 FPKM=68.43）的高表达致使下游花青素合成支路通量大（图 5），间接降低黄酮醇类合成量，为了提高八爪金龙黄酮醇类的含量，也可通过降低抑制二氢黄酮还原酶（dihydroflavonol reductase, DFR）的表达来实现。无色花色素类可以在无色花色素还原酶（leucoanthocyanidin reductase, LAR）作用下可以生成黄烷醇类。根据 Unigenes 的 KEGG 途径，筛选到可能编码黄酮合成途径酶的一些 Unigenes，作为进一步分析的候选基因。

2.5.3 萜类化合物合成相关基因的挖掘 根据 Unigenes 的 KEGG 途径聚类结果，其中与萜类合成相关的代谢通路共有 3 条，包括萜类化合物骨架生物合成（terpenoid backbone biosynthesis），其他萜类化合物生物合成（ubiquinone and other terpenoid-quinone biosynthesis）和倍半萜和三萜类化合物生物合成（sesquiterpenoid and triterpenoid

biosynthesis），共计 143 条 Unigenes。萜类化合物均来源于 C5 单元构建异戊二磷酸（isopentenyl diphosphate, IPP）及其同分异构体二甲基烯丙基二磷酸（dimethylallyl diphosphate, DMAPP），由甲戊酸（mevanolate, MAV）和甲基赤藓糖醇磷酸（methyl-D-erythritphosphate, MEP）2 个途径合成（图 6）。在八爪金龙转录组数据中，共发现 14 个 Unigenes 可能编码 MVA 途径中的关键酶乙酰辅酶 A 酰基转移酶（acetyl-CoA acetyltransferase, AACT）、羟甲基戊二酰辅酶 A 合酶（3-hydroxy-3-methylglutaryl-CoA synthase, HMGS）、3-羟基-3-甲基戊二酰辅酶 A 还原酶（3-hydroxy-3-methylglutaryl coenzyme A reductase, HMGR），甲羟戊酸激酶（mevalonate kinase, MVK）和甲羟戊酸二磷酸脱羧酶（mevalonate diphosphate decarboxylase, MDD）（表 4）。但其 HMGS 和 HMGR 的表达量均不高（图 5），HMGS 的最高表达量为 11.59，HMGR 的最高表达量为 6.48，因此可以增强 HMGS 和 HMGR 基因的表达，以增加 MAV 代谢通路的通量。8 个 Unigenes 可能编码 MEP 途径中的 1-脱氧-D-木酮糖-5-磷酸合酶（1-deoxyxylulose-5-phosphate synthase,

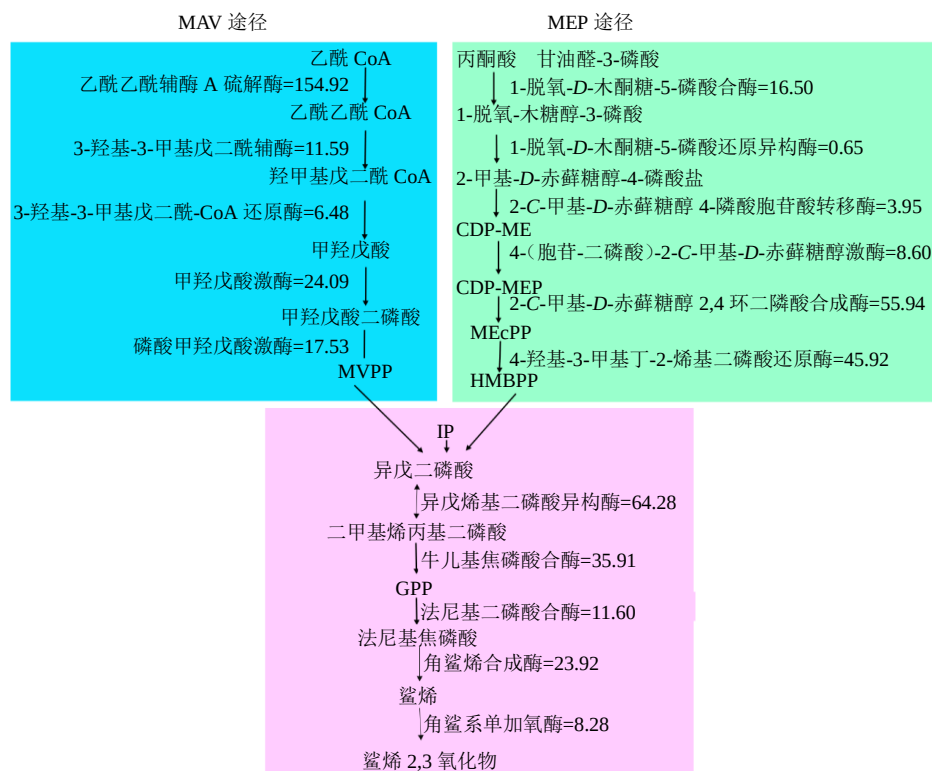


图 6 八爪金龙萜类物质生物合成途径

Fig. 6 Main terpenoid biosynthesis pathway in *A. crispa*

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-磷酸胞苷酸转移酶(2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, MCT)、4-(胞苷-二磷酸)-2-C-甲基-D-赤藓糖醇激酶(4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, CMK)、4-羟基-3-甲基丁-2-烯基二磷酸还原酶(4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, HDS)、4-羟基-3-甲基丁-2-烯基二磷酸还原酶(4-hydroxy-3-methylbut-2-en-1-yl diphosphatereductase, HDR)(表 4)。但其 DXR 和 MCT 的表达量较低(图 5), DXR 的最高表达量为 0.65, MCT 的最高表达量为 3.95, 较低的表达量可能限制了 MEP 途径的合成。萜类化合物均来源于 C5 单元构建 IPP 及其同分异构体 DMAPP, 同时催化中间体产生所需的酶, 在八爪金龙转录组数据中均找到编码相应酶的候选基因, 其中包括 6 个牛儿基焦磷酸合酶(geranylgeranyl diphosphate synthase, GGPPS)、4 个法尼基二磷酸合酶(farnesyl diphosphate synthase, FPS)、2 个鲨烯合酶(squalene synthase, SQS)、2 个鲨烯环氧化物(squalene epoxidase, SQE)、2 个 β -香树脂合成酶(beta-amyrin synthase AS, β -AS)等

(表 4)。

2.5.4 次生代谢后修饰酶 骨架形成后需要经过母核的氧化、糖基化等后修饰的反应才能形成结构各异的成分, 在萜类、黄酮、生物碱等次生代谢物的衍生修饰过程中, 细胞色素 P450 (cytochrome P450, CYP450) 和糖基化转移酶 (UDP-glycosyltransferase, UGT) 主要起催化氧化/羟基化和糖基的重要作用。通过搜索八爪金龙转录组 Swiss-Prot 数据库注释结果, 共找到 140 条 Unigenes 被注释为 CYP450, 隶属于 24 个 CYP450 家族, 其中 CYP71 家族的 Unigenes 最多 (22 条), 其次是 CYP94 和 CYP704, 分别为 15 条和 14 条。113 条 Unigenes 被注释为 UGT, 隶属于 24 个 UGT 亚家族, 包括 16 条 UGT85, 14 条 UTG73, 10 条 UTG80, 9 条 UTG89 和 8 条 UTG74 等。对各个样品中基因表达量进行聚类表达分析, 结果如图 7 所示, 其中 CYP450 家族 FPKM 值较高的有 CYP72A219 (FPKM = 206.71)、CYP704C1 (FPKM = 131.60) 和 CYP94A1 (FPKM = 131.18), UTG 家族 FPKM 值较高的有 UGT92A1 (FPKM = 278.69)、UGT73C3 (FPKM = 171.20) 和 UGT89B1 (FPKM = 125.18)。

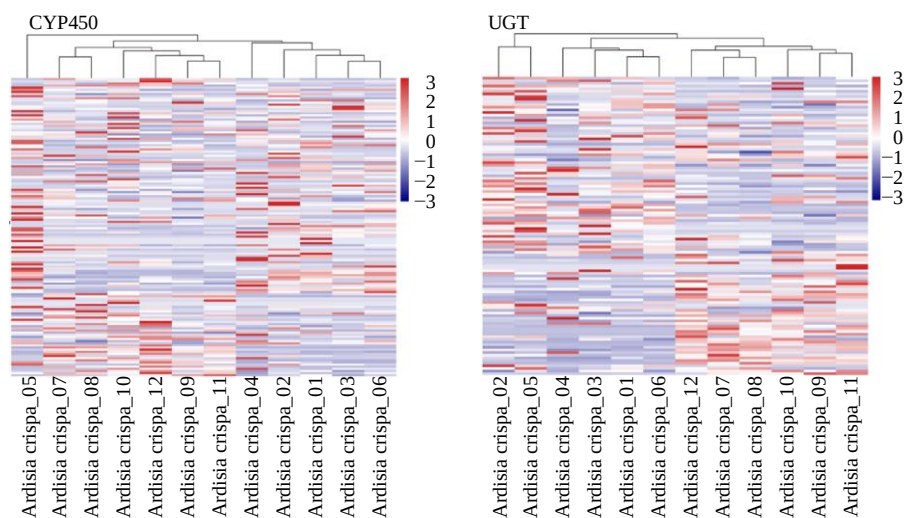


图 7 次生代谢后修饰酶 CYP450 和 UGT 表达模式聚类热图

Fig. 7 Heat map of differential gene expression of CYP450 and UGT

2.5.5 其他次生代谢产物生物合成通路 八爪金龙中还含有生物碱和内源性激素等生长调节物质, 根据 KEGG 代谢通路分析结果, 有 2 条代谢通路可能参与八爪金龙生物碱合成代谢途径, 包括 30 条 Unigenes 参与莨菪烷类、哌啶、吡啶生物碱生物合成, 28 条 Unigenes 参与异喹啉类生物碱生物合成。

分别有 51、30、1 条 Unigenes 参与类胡萝卜素、玉米素、花青素等的生物合成。

2.6 转录因子分析

转录因子能够激活或抑制植物次生代谢产物生物合成途径中功能基因的表达, 从而调控次生代谢产物合成积累。根据各种转录因子的隐马尔可夫模型文件, 利

用 HMMER 3.0 软件对八爪金龙转录组数据进行搜索。结果显示八爪金龙转录组数据预测共有 1265 个 Unigenes 被注释为转录因子,分属于 64 种转录因子类型。最多的转录因子类型是 MYB 类 (v-myb avian myeloblastosis viral oncogene homolog) (125 个), AP2/ERF-ERF 类 (APETALA2/ethylene-responsive factor) (105 个), 锌指蛋白 C2H2 类 (Cys2His2) 81 个, bHLH 类 (basic helix-loop-helix) 78 个, NAC 类 (NAM/ATAF/CUC) 66 个, WRKY 类 64 个, C3H 类 (Cys3-His) 61 个, bZIP 类 (basic leucine zipper) 类 55 个, GRAS 类 [根据 GAI (gibberellic acid insensitive), RGA (repressor of GA1-3 mutant) 和 SCR (scarecrow) 的特征字母而来] 48 个。MYB 转录因子基因家族与类黄酮生物合成紧密相关。

2.7 转录组序列中的 SSR 分析

利用 MISA 软件对八爪金龙转录组 Unigenes 进行 SSRs 分析,共发现 17 400 个 SSR 位点,SSR 序列总长度为 308 795 bp,包括 1245 个复合型 SSR 和 16 155 个完美型 SSR,分布在 12 334 条 Unigenes 中,其发生频率与分布频率分别为 48.96% 和 69.07%,1 条 Unigene 中最多分布 7 个 SSR。单碱基重复 SSR 共 8628 条 (53.41%),双碱基重复 SSR 共 5937 条 (36.75%),三碱基重复 SSR 共 1403 条 (8.68%),四碱基、五碱基、六碱基重复 SSR 总共为 187 条 (1.16%)。八爪金龙转录组 SSR 重复单元的重复次数主要分布在 5~36 次,SSR 重复单元主要以 5~14 次的重复为主,共有 14 937 个 SSR 位点,占总重复单元的 92.46%;15~36 次的重复有 12 186 个 SSR 位点,占 SSR 总数的 8% 以下;20 次以上的重复次数最少,仅有 20 个 SSR 位点,并且以单碱基重复为主 (图 8)。

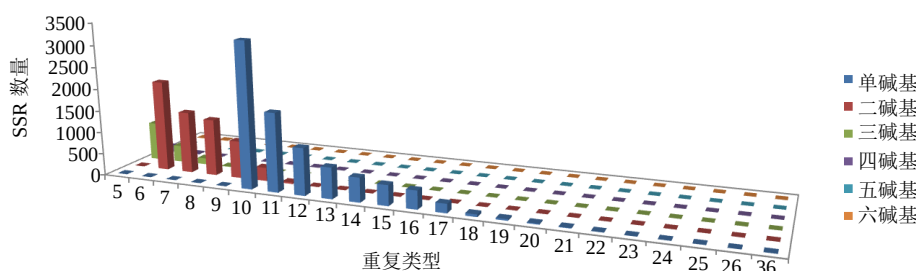


图 8 八爪金龙转录组 SSR 不同重复类型和重复次数的数量分布

Fig. 8 Quantitative distribution of different motif lengths and repeats in SSR of *A. crispata* transcriptome

3 讨论

本研究运用 Illumina HiSeq 4000 高通量测序平台对八爪金龙进行转录组测序,获得了 52 249 条 Unigenes,平均长度为 1440 nt, N₅₀ 为 2336 nt。基因功能注释结果显示 31 391 条 Unigenes 被成功注释到数据库中,占全部 Unigenes 的 60.07%。18 741 条 Unigenes 被注释到 25 种 KOG 类群中,其中注释到“次生代谢物的生物合成、运输和分解代谢”类群的 Unigenes 有 681 条;KEGG 路径注释结果显示八爪金龙中 Unigenes 主要参与萜类、生物碱、黄酮类等生物合成相关的 18 个次生代谢通路。近年来,大量药用植物的生长发育状态和次生代谢产物合成机理通过转录组分析得到阐明,八爪金龙注释信息的完成为基因组信息缺乏的八爪金龙次生代谢产物的生物合成的研究提供了新的方向。

PAL 在植物的次生代谢尤其是在苯丙烷类代

谢中有关键作用,为连接初级代谢和苯丙烷类代谢、催化苯丙烷类代谢途径第一步反应的酶,是苯丙烷类代谢的关键酶和限速酶^[41],该基因在植物中为多拷贝基因^[42-43]。Koukol 等^[44]1961 年首次从高等植物中成功分离并纯化,目前已在多种植物中测到 PAL 基因的序列。本研究在八爪金龙中共鉴定出 5 条 PAL 序列,其中主要表达的序列有 1 条 (FPKM=74.58),PAL 家族一般仅少量基因表达其余同源基因沉默^[45-46]。八爪金龙黄酮合成途径的上游阶段 4CL 基因的表达量相对较低,4CL 是苯丙氨酸途径中的关键性限速酶,香豆酸在 4CL 的作用下形成香豆酰-CoA^[47]。但实验也发现通过转录组测序鉴定出 6 条 4-香豆酰 CoA 连接酶类似物 Unigenes,可能增强 4CL 基因的表达,增加了整条代谢通路的通量。二氢黄酮类化合物能在 IFS 酶的催化下将芳香基团从 2 位向 3 位转移生成异黄酮类化合物^[48]。但在八爪金龙中未鉴定到 IFS 酶,目前

八爪金龙中也没有发现异黄酮类物质。八爪金龙中黄酮类代谢通路中, 二氢黄酮类化合物可能通过 FNS 合成黄酮类, 或者通过 F3H 合成二氢黄酮醇类物质。

八爪金龙三萜类物质是其次生代谢产物的重要组成部分, 而其含量和组分又主要取决于生物合成关键酶以及在细胞中的表达水平。植物中三萜类物质的生物合成的前体是由 MVA 途径和 MEP 途径共同合成的 IPP 或其异构体 DMAPP^[49]。在八爪金龙转录组的 KEGG 注释中, 共筛选到 143 个参与三萜合成上游部分基因, 分别属于 MVA 途径和 MEP 途径的各个环节。整个过程涉及多个酶, 因此关键酶的确定至关重要。

CYP450 是植物体内一类超基因家族编码的单加氧酶, 具有广泛的催化活性, 能够催化多种初级和次级代谢反应, 主要用于涉及萜类、生物碱类、甾醇类、黄酮类、异黄酮等的合成和代谢反应等^[50-51]。在八爪金龙的转录组中总共有 140 个 Unigenes 被注释为 CYP450 基因。UDP-糖基转移酶 (uridine diphosphate-glycosyl-transferases, UGTs) 参与次生代谢产物合成的最后阶段, 对于生物活性成分终端产物的多样性、稳定性和结构修饰具有重要意义^[52]。在本研究八爪金龙的转录组中总共有 113 条 Unigenes 被注释为 UGTs。八爪金龙药用部位是根, 药材采挖过程中地上茎、叶及花部则弃之不用。如何结合不同部位次生产物合成及积累特点进行新药药用部位的开发研究, 合理利用资源, 值得深入探讨。下一步将选择相关关键酶基因, 基于 qRT-PCR 进行系统的定量表达分析, 并开展不同部位八爪金龙黄酮类、三萜类物质合成和积累等相关基础性研究工作。

转录组测序技术可以从合成途径入手, 从整体水平上了解各个酶与产物的关系, 从而克服单个酶的表达与产物关系的局限性。本研究对八爪金龙进行转录组学研究, 构建八爪金龙转录组数据库, 极大地丰富了八爪金龙的基因资源, 为更深入地研究八爪金龙次生代谢合成及调控提供基础数据, 为阐明八爪金龙的次生代谢合成机制奠定了基础。同时 SSR 分子标记可以用来分析八爪金龙的遗传多样性、构建遗传图谱和分子标记辅助育种等。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 贵州省中医研究所. 贵州中草药名录 [M]. 贵阳: 贵州人民出版社, 1986: 411.

[2] 余成龙, 宋良科, 吴蜀星, 等. 朱砂根药用资源分布及研究进展 [J]. 安徽农业科学, 2012, 40(28): 13793-13794.

[3] 陈业文, 张灏, 甘亚平, 等. 岩白菜素对肝癌的抑制作用 [J]. 重庆医学, 2018, 47(26): 3365-3367.

[4] Hamid R A, Fong L M, Ting Y L. Anti-arthritis and gastroprotective activities of *Ardisia crispa* root partially mediated via its antioxidant effect [J]. *J Compl Integr Med*, 2017, 15(1): 225.

[5] 王聪, 朱晶晶, 郑秀静, 等. 岩白菜素衍生物 D-23 抗白血病作用及机制研究 [J]. 中国药理学杂志, 2018, 53(19): 1638-1644.

[6] 赵欧, 杜莹, 韦万丽. 开喉剑及组方药材山豆根、八爪金龙挥发油的 GC-MS 分析 [J]. 湖北农业科学, 2016, 55(6): 1548-1550.

[7] 张嫩玲, 胡江苗, 周俊, 等. 百两金的化学成分 [J]. 天然产物研究与开发, 2010, 22(4): 587-589.

[8] 范卫锋, 石勇新, 张文, 等. 广西朱砂根不同部位中岩白菜素含量的研究 [J]. 中国中医药信息杂志, 2009, 16(1): 53.

[9] 麻秀萍, 蒋朝晖, 丁宁, 等. 高效液相色谱法测定朱砂根中岩白菜素含量 [J]. 中国民族民间医药杂志, 2004, 13(6): 347-348.

[10] 黄伟, 谭桂山, 徐康平, 等. 百两金细胞毒活性成分研究 [J]. 天然产物研究与开发, 2010, 22(6): 949-951.

[11] Liu D L, Wang N L, Zhang X, et al. Two new triterpenoid saponins from *Ardisia crenata* [J]. *J Asian Nat Prod Res*, 2007, 9(2): 119-127.

[12] Liu D L, Zhang X, Zhao Y M, et al. Three new triterpenoid saponins from the roots of *Ardisia crenata* and their cytotoxic activities [J]. *Nat Prod Res*, 2016, 30(23): 2694-2703.

[13] Zheng Z F, Xu J F, Feng Z M, et al. Cytotoxic triterpenoid saponins from the roots of *Ardisia crenata* [J]. *J Asian Nat Prod Res*, 2008, 10(9/10): 833-839.

[14] 赵欧, 杜莹. 高效液相色谱法测定八爪金龙及组方制剂中岩白菜素的含量 [J]. 湖北农业科学, 2016, 55(1): 186-188.

[15] Hamsin D E, Hamid R A, Yazan L S, et al. The hexane fraction of *Ardisia crispa* Thunb. A. DC. roots inhibits inflammation-induced angiogenesis [J]. *BMC Complement Altern Med*, 2013, 13: 5.

[16] Hamid R A, Fong L M, Ting Y L. Anti-arthritis and gastroprotective activities of *Ardisia crispa* root partially mediated via its antioxidant effect [J]. *J Complementary Integr Med*, 2018, 15(1): 12.

[17] 陈新连, 周建国, 马双姣, 等. 基于 ITS2 序列的朱砂根及其混伪品分子鉴定 [J]. 中国现代中药, 2017, 19(7):

- 939-943.
- [18] Wang W, Wang Y J, Zhang Q, *et al.* Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing [J]. *BMC Genomics*, 2009, 10: 465.
 - [19] Sun C, Li Y, Wu Q, *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 Genom*, 2010, 11: 262.
 - [20] Jayakodi M, Lee S C, Lee Y S, *et al.* Comprehensive analysis of *Panax ginseng* root transcriptomes [J]. *BMC Plant Biol*, 2015, 15: 138.
 - [21] He L, Xu X L, Li Y, *et al.* Transcriptome analysis of buds and leaves using 454 pyrosequencing to discover genes associated with the biosynthesis of active ingredients in *Lonicera japonica* Thunb [J]. *PLoS One*, 2013, 8(4): e62922.
 - [22] 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.
 - [23] 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.
 - [24] 邓泱泱, 荔建琦, 吴松峰, 等. NR 数据库分析及其本地化 [J]. 计算机工程, 2006, 32(5): 71-73.
 - [25] Apweiler R, Bairoch A, Wu C H, *et al.* UniProt: the universal protein knowledgebase [J]. *Nucleic Acids Res*, 2004, 32: D115-D119.
 - [26] 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.
 - [27] 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.
 - [28] 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): 1-28.
 - [29] Huerta-Cepas J, Szklarczyk D, Forslund K, *et al.* eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences [J]. *Nucleic Acids Res*, 2016, 44(D1): D286-D293.
 - [30] Kanehisa M, Goto S, Kawashima S, *et al.* The KEGG resource for deciphering the genome [J]. *Nucleic Acids Res*, 2004, 32: D277-D280.
 - [31] 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.
 - [32] Eddy S R. Profile hidden Markov models [J]. *Bioinformatics*, 1998, 14(9): 755-763.
 - [33] Finn R D, Bateman A, Clements J, *et al.* Pfam: the protein families database [J]. *Nucleic Acids Res*, 2014, 42(Database issue): D222-D230.
 - [34] Langmead B, Trapnell C, Pop M, *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome [J]. *Genome Biol*, 2009, 10(3): 1-10.
 - [35] Li B, Dewey C N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome [J]. *BMC Bioinform*, 2011, 12: 323.
 - [36] Trapnell C, Williams B A, Pertea G, *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation [J]. *Nat Biotechnol*, 2010, 28(5): 511-515.
 - [37] Dixon R A, Pasinetti G M. Flavonoids and isoflavonoids: From plant biology to agriculture and neuroscience [J]. *Plant Physiol*, 2010, 154(2): 453-457.
 - [38] Nagatomo Y, Usui S, Ito T, *et al.* Purification, molecular cloning and functional characterization of flavonoid C-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) *Cotyledon* [J]. *Plant J*, 2014, 80(3): 437-448.
 - [39] Kang J H, McRoberts J, Shi F, *et al.* The flavonoid biosynthetic enzyme *Chalcone* isomerase modulates terpenoid production in glandular trichomes of tomato [J]. *Plant Physiol*, 2014, 164(3): 1161-1174.
 - [40] Achnine L, Blancaflor E B, Rasmussen S, *et al.* Colocalization of *L*-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis [J]. *Plant Cell*, 2004, 16(11): 3098-3109.
 - [41] Thomas V. Phenylpropanoid biosynthesis [J]. *Mol Plant*, 2010, 3(1): 2-20.
 - [42] Achnine L, Blancaflor E B, Rasmussen S, *et al.* Colocalization of *L*-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis[J]. *Plant Cell*, 2004, 16(11): 3098-3109.
 - [43] Olsen K M, Lea U S, Slimestad R, *et al.* Differential expression of four *Arabidopsis* PAL genes; PAL1 and PAL2 have functional specialization in abiotic environmental-triggered flavonoid synthesis [J]. *J Plant*

- Physiol*, 2008, 165(14): 1491-1499.
- [44] Koukol J, Conn E E. The metabolism of aromatic compounds in higher plants. IV. Purification and properties of the phenylalanine deaminase of *Hordeum vulgare* [J]. *J Biol Chem*, 1961, 236: 2692-2698.
- [45] Chang A, Lim M H, Lee S W, *et al.* Tomato phenylalanine ammonia-lyase gene family, highly redundant but strongly underutilized [J]. *J Biol Chem*, 2008, 283(48): 33591-33601.
- [46] 袁灿, 彭芳, 钟文娟, 等. 赶黄草的转录组测序及分析 [J]. 中草药, 2017, 48(21): 4507-4514.
- [47] Gao S, Yu H N, Xu R X, *et al.* Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum* [J]. *Phytochemistry*, 2015, 111: 48-58.
- [48] Gebhardt Y, Witte S, Forkmann G, *et al.* Molecular evolution of flavonoid dioxygenases in the family Apiaceae [J]. *Phytochemistry*, 2005, 66(11): 1273-1284.
- [49] Cheng A X, Lou Y G, Mao Y B, *et al.* Plant terpenoids: Biosynthesis and ecological functions [J]. *J Integr Plant Biol*, 2007, 49(2): 179-186.
- [50] Chapple C. Molecular-genetic analysis of plant cytochrome p450-dependent monooxygenases [J]. *Annu Rev Plant Physiol Plant Mol Biol*, 1998, 49: 311-343.
- [51] Schuler M A, Werck-Reichhart D. Functional genomics of P450s [J]. *Annu Rev Plant Biol*, 2003, 54: 629-667.
- [52] Jung S C, Kim W, Park S C, *et al.* Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg₃ and rd [J]. *Plant Cell Physiol*, 2014, 55(12): 2177-2188.

[责任编辑 时圣明]